nature portfolio

Peer Review File

Ultrasensitive Dim-Light Neuromorphic Vision Sensing via Momentum-Conserved Reconfigurable van der Waals Heterostructure

REVIEWER COMMENTS

Reviewer #1 (Remarks to the Author):

The manuscript by He et al. described an ultrasensitive and bipolar phototransistor based on $MAPbI_3/Bi_2O_2Se$ van der Waals heterojunction, which enables the implement of multiple neuromorphic processing functions for visual information. The extremely high photoresponsivity and emulation of multifunctional visual receptive fields are impressive. The manuscript needs however major revision. While I am excited by the results of this manuscript, there are some flaws that must be fixed before I can recommend its publication. The following are the detailed comments:

1. In lines 21-23, the authors state “effective bipolar modulations usually require ultrathin channel materials, which suffer from poor optoelectronic performances due to the intrinsic weak absorption”. Therefore, what fraction of the light is expected to be absorbed in proposed device for the spectrum used? The absorption spectrum should be provided.
2. In line 166, how did the authors calculate the detectivity? The authors should measure the noise density spectrum and calculate the noise for real detectivity.
3. How about the environmental stability of the device as most perovskites are sensitive to water?
4. The properties shown in Fig. 3g-h and the related figures in supplementary materials should be supplied for all samples in 3×3 transistor array to demonstrate the uniformity of array.
5. The cross-talk between adjacent pixels should be characterized. A detailed study of photocurrent mapping by scanning photo-current microscopy may be helpful.
6. The “in-sensor letter classification” and “in-sensor convolution kernel” functions have been reported in previous article (For example, “Ultrafast machine vision with 2D material neural network image sensors”, Nature, 2020, 579, 62-66, or “Gate-tunable van der Waals heterostructure for reconfigurable neural network vision sensor”, Sci. Adv. 2020; 6: eaba6173.). Could the authors make a brief comparison with these reported devices?
7. To demonstrate the capabilities of key feature extraction and traffic light detection in dim-light environments, the intensity of the dim-light image should be specific as described in the letter classification task.
8. In Figure 5b, the preprocessed digital image appears to show no discernible alterations when compared to the raw digital photograph.
9. The speed is another important evaluation criterion for object detection. The reader would like to observe the impact of the preprocessing function on the inference speed of the neural network.
10. Can author estimate the energy consumption of device during one light pulse, or energy consumption of sensor array for preprocessing functions?
11.The authors should clearly explain the preprocessing procedure for color of traffic light using the sensor array in Method part or Supplementary information.
12. There are some mistakes of figure’s number. For example: Line 293, the figure’s number “Supplementary Fig. S12” is wrong.
13. Figure 2b is high-resolution TEM (HRTEM) image, please correct the caption.
14. The following reference might be considered for comprehensiveness, Nature Communications 2023, 14: 6736.

Reviewer #2 (Remarks to the Author):

The manuscript presents a significant advancement in neuromorphic vision sensors by introducing a momentum-conserved reconfigurable phototransistor based on a van der Waals heterojunction between methylammonium lead iodide perovskite and two-dimensional Bi_2O_2Se semiconductor. The developed sensor demonstrates impressive optoelectronic performances with record-high photoresponsivity and specific detectivity, alongside a broad dynamic range suitable for dim-light conditions reminiscent of natural moonlight. The device integrates in-sensor analog multiply-accumulate operations, efficiently enhancing feature extraction and improving detection capabilities under challenging lighting conditions.
The manuscript is well-structured, with a clear and logical progression from the introduction of the concept, through the detailed experimental and simulation work, to the presentation of results and their discussion. The topic is of high relevance to current trends in photodetector development and neuromorphic computing, making it a timely contribution to the field. The integration of the device with a convolutional neural network to demonstrate practical applications adds considerable value to the study. I would recommend the publication of this manuscript in Nature Communications.
While the manuscript is well-written and the study robust, a more detailed comparison with current leading technologies in terms of performance metrics such as speed, cost, and scalability could enhance the manuscript.
How scalable is the fabrication process of the $MAPbI_3/Bi_2O_2Se$ heterotransistors? Are there potential challenges in large-scale production?
Can the authors provide insights into the long-term stability and operational durability of the heterotransistors under varying environmental conditions?
Is there potential for integrating this sensor technology with other types of neural networks or electronic components to further enhance its capabilities?

Reviewer #3 (Remarks to the Author):

This work discusses $MAPbI_3/Bi_2O_2Se$ phototransistors with neuromorphic vision sensing capabilities. The device demonstrated high, gate-tunable photoresponsivity, which is pivotal in developing computing demonstrations. The discussion on the device mechanism is clearly presented, and the work showcases impressive neuromorphic computing applications that will appeal to the broad readership of Nature Communications. My comments below aim to enhance this paper from a neuromorphic perspective.
1. The linear dependency of photoresponses on gate voltages and Pin (Fig. 3c, 3g, 3h) facilitates the employment of device characteristics into neuromorphic computing applications. Can the authors discuss more about the device mechanism and how it leads to this linear dependency?
2. In Fig. 3a-3c, R monotonically decreases as Pin increases at Vg = 0V. This monotonical trend, however, does not hold in Fig. 3g (640nm light) and Fig. S5 (532nm and 780nm light), irrespective of whether Vg is negative, zero or positive. There is no linear relation between R and Pin from Fig. 3g and Fig. S5. Can the authors explain this discrepancy?
3. For easier comparison, I suggest that the authors plot all figures at one specific wavelength (such as 532 nm) throughout Fig. 3 and then present other similar series of plots at wavelengths of 400nm, 640nm, and 780nm in the Supporting Information.
4. The significance of how the momentum conservation between $MAPbI_3$ and Bi_2O_2Se contributes to this device/this work should be more clearly established. The Fig. 3f shows previous work using $MAPbI_3/MoS_2$ can perform dim-light sensing (similar light power) with

photoresponsivity above 10^5 A/W. Assuming it is the state-of-art performance, this work should compare how its higher photoresponsivity (10^7 A/W) impacts the CNN applications compared to a potoresponsivity of 10^5 A/W.

5. How consistent is the device-to-device variation in this type of device, especially since a crossbar array of this device is fabricated for neuromorphic computing application? How are the noise, hysteresis and other properties of the device accounted for in the simulation?

6. Please explain the algorithm application in more detail for the reader that is not specialized in neuromorphics, perhaps in SI.

7. What does the multilayer structure represent in Fig. 4b?

Responses to Review Comments

We express our sincere thanks to all the Reviewers for your invaluable and constructive comments on this manuscript. After thorough considerations of each comment, we have made required revisions to improve both the quality and the comprehensibility of our work. Below, we present our point-by-point responses and the corresponding modifications made in accordance with your valuable suggestions.

Comments from Reviewer #1

Overall Remarks:

The manuscript by He et al. described an ultrasensitive and bipolar phototransistor based on $MAPbI_3/Bi_2O_2Se$ van der Waals heterojunction, which enables the implement of multiple neuromorphic processing functions for visual information. The extremely high photoresponsivity and emulation of multifunctional visual receptive fields are impressive. The manuscript needs however major revision. While I am excited by the results of this manuscript, there are some flaws that must be fixed before I can recommend its publication.

Response:

We are grateful for your acknowledgment of the significant advancements our work, and we deeply appreciate your comprehensive and constructive feedback. Following your recommendations, we have broadened the scope of our experiments and thoroughly revised our manuscript to enhance both its robustness and its appropriateness for publication.

==

Comment 1-1: In lines 21-23, the authors state “effective bipolar modulations usually require ultrathin channel materials, which suffer from poor optoelectronic performances due to the intrinsic weak absorption”. Therefore, what fraction of the light is expected to be absorbed in proposed device for the spectrum used? The absorption spectrum should be provided.

Response:

In the context of existing research, the attainment of bipolar photoresponse primarily hinges on modulating the band alignment or defect status, which generally necessitate materials characterized by extensive specific surface areas and minimal thickness. Nevertheless, such configurations frequently suffer from suboptimal light absorption, thereby decreasing the photocarrier generation. This juxtaposition poses challenges in simultaneously realizing high photoresponsivity alongside bipolar photoresponse attributes.

To this end, we introduce a novel synergy between Bi_2O_2Se and $MAPbI_3$ perovskite, distinguished by the superior photocarrier transport in Bi_2O_2Se and the exceptional absorption coefficient of $MAPbI_3$. The $MAPbI_3/Bi_2O_2Se$ heterojunction ensures comprehensive absorbance and expedited collection of photocarriers across the visible-NIR spectrum of 300-1500 nm. In fact, we measured the absorption spectra of $MAPbI_3$ film (i.e., ~170 nm thickness), Bi_2O_2Se film (i.e., ~10 nm thickness), and $MAPbI_3/Bi_2O_2Se$ heterojunction film (i.e., ~170 nm/10 nm thickness). The absorption curves in **Fig. R1** reveals that the perovskite film exhibited typical absorption within the visible range of 300-800 nm, and the absorption of Bi_2O_2Se film extended to the NIR range upon 1500 nm, due to its relatively low band gap. However, both the photoresponsivity of $MAPbI_3$ and Bi_2O_2Se films were not good, attributed to the poor photocarrier transport of $MAPbI_3$ and the ultrathin thickness of Bi_2O_2Se nanosheet. We have demonstrated in **Fig. 3e** that the photoresponsivity of heterotransistors maintained at a high-level ranging from 10^3 A/W to 10^6 A/W for both positive and negative photoresponses across the visible spectrum of 400-800 nm, demonstrating robust sensitivities for detecting a broad range of colorful images. We want to note that, according to our very recent results, the heterostructure effectively combines the absorption characteristics of $MAPbI_3$ and Bi_2O_2Se can achieve unparalleled bipolar photoresponses all across the extensive visible-NIR range of 300-1350 nm (will report in another manuscript), establishing a new benchmark in the field.

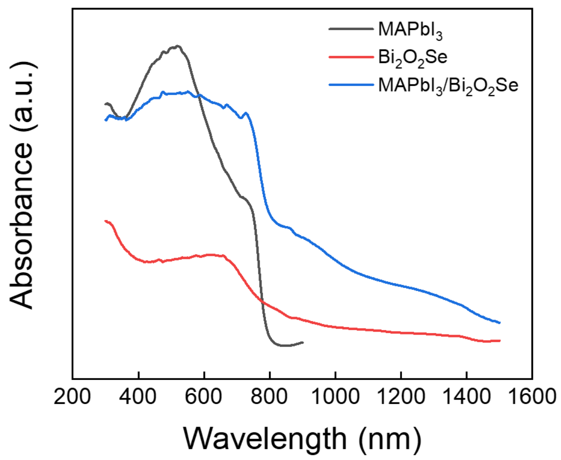
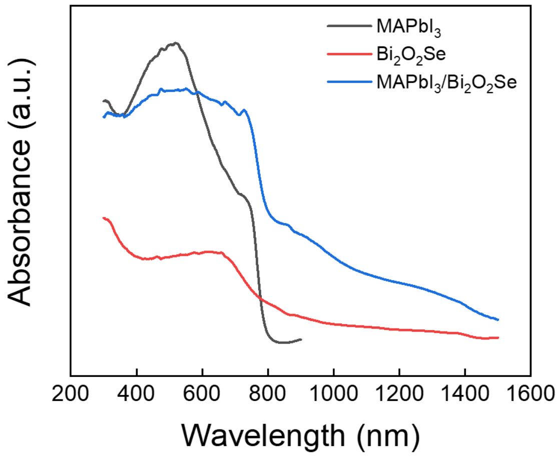


Figure R1. Absorption spectra of $MAPbI_3$, Bi_2O_2Se and $MAPbI_3/Bi_2O_2Se$ films.

Changes Made:

As suggested by the Reviewer, we have included the absorption spectra of the $MAPbI_3$, Bi_2O_2Se and $MAPbI_3/Bi_2O_2Se$ film in the **Supplementary Fig. S2** along with the following illustration in Page 5 of the main text (highlighted in red): "$MAPbI_3/Bi_2O_2Se$ heterojunction film exhibited excellent light absorption in the broadband visible-NIR spectrum of 300-1500 nm."



Supplementary Figure S2 | Absorption spectra of $MAPbI_3$, Bi_2O_2Se and $MAPbI_3/Bi_2O_2Se$ films.

==

Comment 1-2: In line 166, how did the authors calculate the detectivity? The authors should measure the noise density spectrum and calculate the noise for real detectivity.

Response:

The specific detectivity D^* characterizes the photodetector sensitivity towards minimal light intensities, encapsulating the interplay between photoresponse and inherent noise. The D^* can be formulated by **Equation R1**:

$$D^* = \frac{\sqrt{A\Delta f}}{NEP} \tag{R1}$$

where A is the effective detection area of photodetector，Δf is the operational bandwidth, and *NEP* is the noise equivalent power. The *NEP* parameter is pivotal, defined as the incident optical power of incident light that yields a signal-to-noise ratio (*SNR*) of unity across a 1 Hz bandwidth, fundamentally reflecting the lowest optical power discernible by the photodetector above the noise threshold. It is calculable as the quotient of the noise spectral density and the photoresponsivity of photodetector. Thus, a reformulated expression for the D^* is provided as **Equation R2**:

$$D^* = \frac{R\sqrt{A\Delta f}}{i_n} \tag{R2}$$

where R is the photoresponsivity, and i_n is the noise current.

Note that photodetector noise typically comprises shot noise, thermal noise, flicker ($1/f$) noise, and generation-recombination noise. [1-3] Predominantly, when the device's operation is significantly influenced by the shot noise attributed to the dark current, the noise current is quantified by **Equation R3**:

$$i_n = \sqrt{2qI_{dark}\Delta f} \tag{R3}$$

where q is the charge of an electron, and I_{dark} is the dark current. Integrating **Equations R2** and **R3** yields a streamlined formulation for the D^* as the following:

$$D^* = \frac{R}{\sqrt{2eI_{dark}/A}} \tag{4}$$

This equation reveals that in certain instances, the estimation of D^* value in low-dimensional photodetectors utilizes the dark current as a proxy for the noise current. [4-8] Initially, our analysis adopted this approach for calculating D^*. Now following the Reviewer's suggestion, we have carefully reevaluated the measurement accuracy of D^* within 2D photodetectors. Notably, for 2D photodetectors operating on photoconductive or photogating principles, the generation-recombination noise significantly correlates with the photoconductive gain. [9-10] Overlooking this facet of noise leads to a pronounced overestimation of D^*, a discrepancy that our initial methodology inadvertently fostered. Therefore, we embarked on a detailed examination of the noise spectrum characterizing our heterotransistor across varying gate voltages as shown in **Fig. R2**. Then we derived the values of D^* for various gate voltages, and compared these results with the D^* values obtained using the initial dark-current estimation method in **Table R1**.

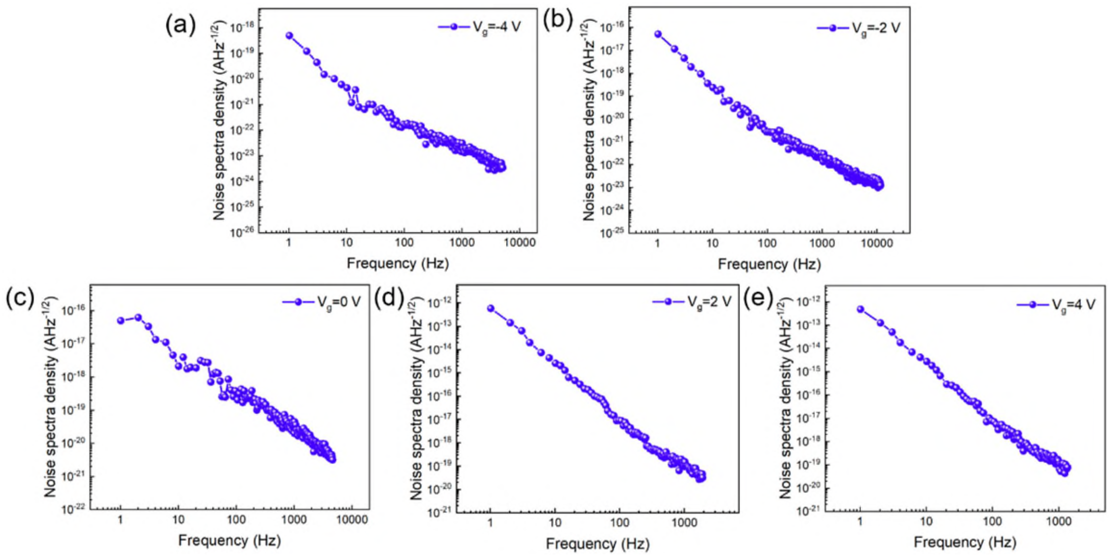


Figure R2. Noise spectral density of the $MAPbI_3/Bi_2O_2Se$ heterotransistor at the gate voltage of (a) V_g=-4 V, (b) V_g=-2 V, (c) V_g=0 V, (d) V_g=2 V, and (e) V_g=4 V.

Table R1. Comparison of the specific detectivity D^* calculated from the dark-current estimation and derived from the noise spectrum.

Gate Voltage (V)	Dark-Current Estimation (Jones)	Noise-Spectrum Measurement (Jones)
-4	1.3×10^{12}	2.3×10^{7}
-2	2.5×10^{15}	1.5×10^{11}
0	9.6×10^{14}	5.2×10^{11}
2	9.1×10^{14}	5.8×10^{9}
4	4.9×10^{14}	4.2×10^{9}

Corresponding Reference:

[1] García de Arquer, F., Armin, A., Meredith, P. & Sargent, E. Solution-processed semiconductors for next-generation photodetectors. *Nat. Rev. Mater.* **2**, 16100 (2017).

[2] Li, G., Wang, Y., Huang, L., & Sun, W. Research progress of high-sensitivity perovskite photodetectors: a review of photodetectors: noise, structure, and materials. *ACS Appl. Electron. Mater.* **4**, 1485-1505 (2022).

[3] Xu, Y., & Lin, Q. Photodetectors based on solution-processable semiconductors: Recent advances and perspectives. *Appl. Phys. Rev*. **7**, 011315 (2020).

[4] Ma, C., Shi, Y., Hu, W., Chiu, M. H., Liu, Z., Bera, A., ... & Wu, T. Heterostructured $WS_2/CH_3NH_3PbI_3$ photoconductors with suppressed dark current and enhanced photodetectivity. *Adv. Mater.* **28**, 3683-3689 (2016).

[5] Wang, Y., Fullon, R., Acerce, M., Petoukhoff, C. E., Yang, J., Chen, C., ... & Chhowalla, M. Solution-processed MoS_2/organolead trihalide perovskite photodetectors. *Adv. Mater.* **29**, 1603995 (2017).

[6] Lu, J., Carvalho, A., Liu, H., Lim, S. X., Castro Neto, A. H., & Sow, C. H. Hybrid bilayer WSe_2-$CH_3NH_3PbI_3$ organolead halide perovskite as a high-performance photodetector. *Angew. Chem.* **128**, 12124-12128 (2016).

[7] Peng, Z. Y., Xu, J. L., Zhang, J. Y., Gao, X., & Wang, S. D. Solution-processed high-performance hybrid photodetectors enhanced by perovskite/MoS_2 bulk heterojunction. *Adv. Mater. Interfaces* **5**, 1800505 (2018).

[8] Zhang, D. D., & Yu, R. M. Perovskite-WS_2 nanosheet composite optical absorbers on graphene as high-performance phototransistors. *Front. Chem.* **7**, 257 (2019).

[9] Rogalski, A. Detectivities of WS_2/HfS_2 heterojunctions. *Nat. Nanotechnol.* **17**, 217-219 (2022).

[10] Wang, F., Zhang, T., Xie, R., Wang, Z., & Hu, W. How to characterize figures of merit of two-dimensional photodetectors. *Nat. Commun.* **14**, 2224 (2023).

Changes Made:

We made the following revisions (highlighted in red):

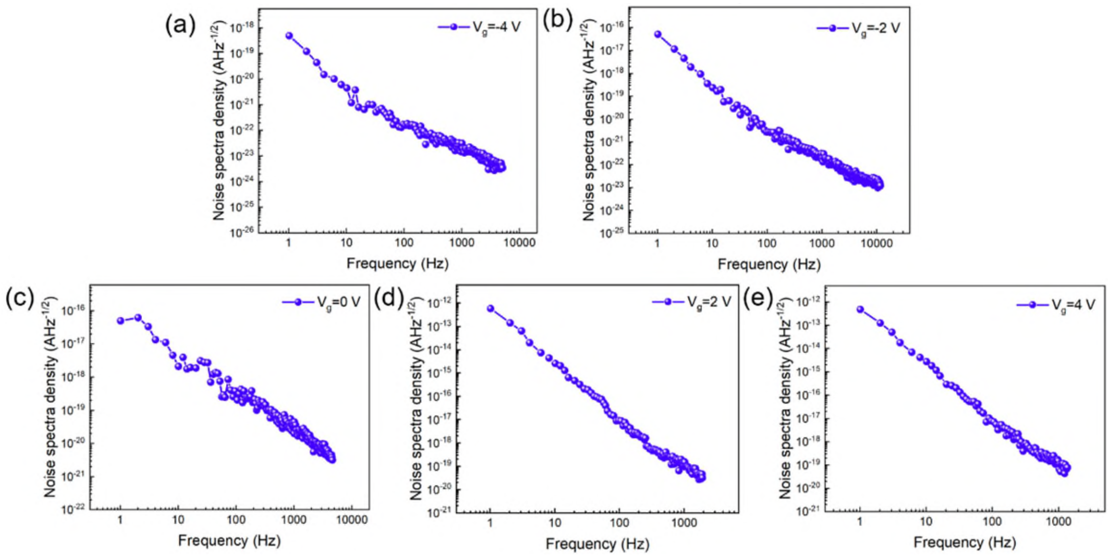
In the Abstract (Page 2), we have updated the specific detectivity value to "5.2×10^{11} Jones."

In the Introduction (Paragraph 2, Page 4), we have revised the text as "accompanied by a specific detectivity of 5.2×10^{11} Jones."

In the main text (Paragraph 2, Page 7), we have included the context: " The specific detectivity (D^*) is determined by $D^* = \frac{R\sqrt{A\Delta f}}{i_n}$,where A is effective detection area, Δf is the operational bandwidth, and i_n is the noise current. The noise spectrum of the $MAPbI_3/Bi_2O_2Se$ heterotransistor at varied gate voltages was characterized (**Supplementary Fig. S4**). An unprecedented R of 6×10^7 A/W was achieved accompanied with a D^* of 5.2×10^{11} Jones and an external quantum efficiency (EQE) of 1.2×10^8 %".

We have added the noise spectrum of the $MAPbI_3/Bi_2O_2Se$ heterotransistor at varied gate voltages to **Supplementary Fig. S4**.

In the Conclusion (Page 13), we have updated as "the specific detectivity of 10^{11} Jones".



Supplementary Figure S4 | Noise spectral density of the $MAPbI_3/Bi_2O_2Se$ heterotransistor at the gate voltage of (a) V_g=-4 V, (b) V_g=-2 V, (c) V_g=0 V, (d) V_g=2 V, and (e) V_g=4 V.

===

Comment 1-3: How about the environmental stability of the device as most perovskites are sensitive to water?

Response：

The primary concern within halide perovskites has centered on improve the environmental stability of these materials, a challenge that often eclipses efficiency considerations. This instability of halide perovskite is attributed to the intrinsic organic components,[1] ion migrations,[2] and phase segregations under illumination.[3] In our device preparation process, we spin coat a layer of polymethyl methacrylate (PMMA) on the $MAPbI_3$ film surface to protect the device from atmospheric moisture. As suggested by the Reviewer, we have tested the stability of $MAPbI_3/Bi_2O_2Se$ heterotransistor array as follows. We fabricated a 3×3 $MAPbI_3/Bi_2O_2Se$ heterotransistor array and subjected them to the stability testing. At the very beginning, all the nine devices in the array exhibited clear bipolar photoresponse and demonstrated good uniformity (**Fig. R3**). Subsequently, the array was stored in an ambient environment with a temperature of 27±1 ℃ and a relatively humidity of about 55±5 %. After about 168 hours (i.e., one week), we tested the array again. As shown in **Fig. R4**, all devices exhibited a positive photoresponse under negative gate voltage and a negative response when gate voltage was positive, maintaining 100 % bipolar behavior.

Further analysis was directed towards extracting the photoresponsivity as a function of gate voltage. Initial assessments highlighted that the photoresponsivity of all devices was linearly modulated by gate voltage (**Fig. R5**). Upon reexamination after 168 hours in ambient conditions, slight deviations in the symmetry of the bipolar photoresponse was observed (**Fig. R6**). This emerging asymmetry is expected to stem from band drift, potentially induced by alterations in the compositional integrity of perovskites. Moreover, this asymmetry became progressively pronounced after approximately 300 hours (i.e., nearly two weeks) in ambient air, culminating in a complete degradation of performance within a timeframe of 3-4 weeks, thus underscoring the critical need for enhanced protective strategies in the development of stable perovskite-based devices.

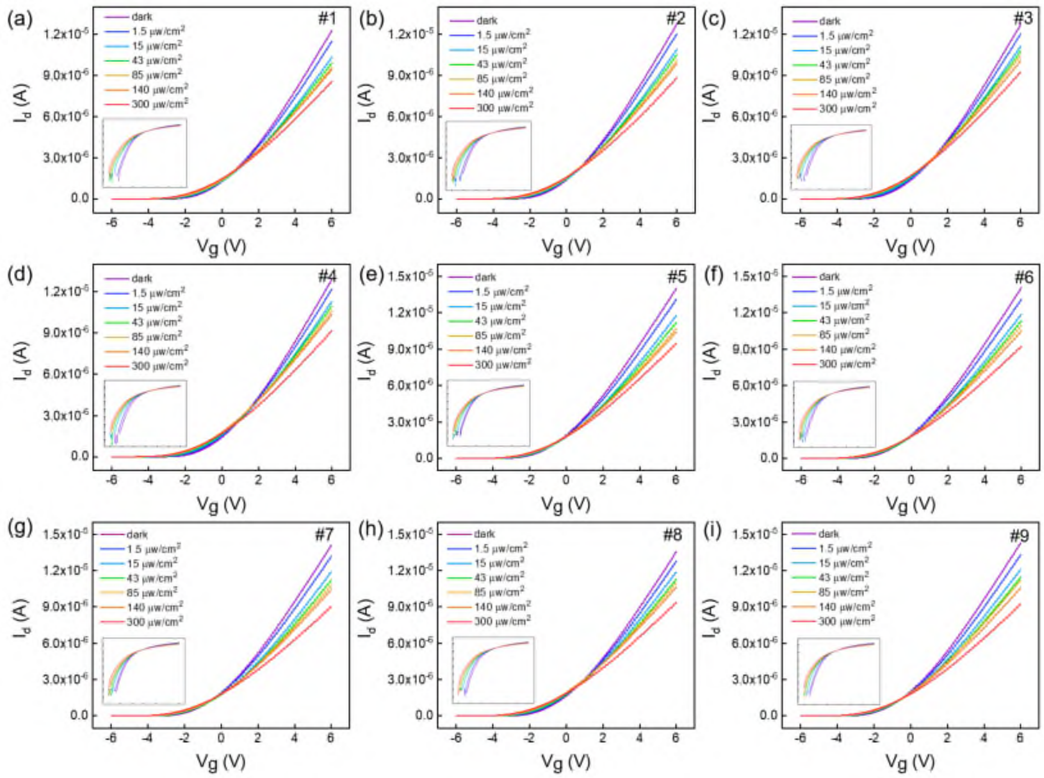


Figure R3. Transfer curves of $MAPbI_3/Bi_2O_2Se$ heterotransistor array under the light illumination of 640 nm with varied light intensities tested after manufacturing.

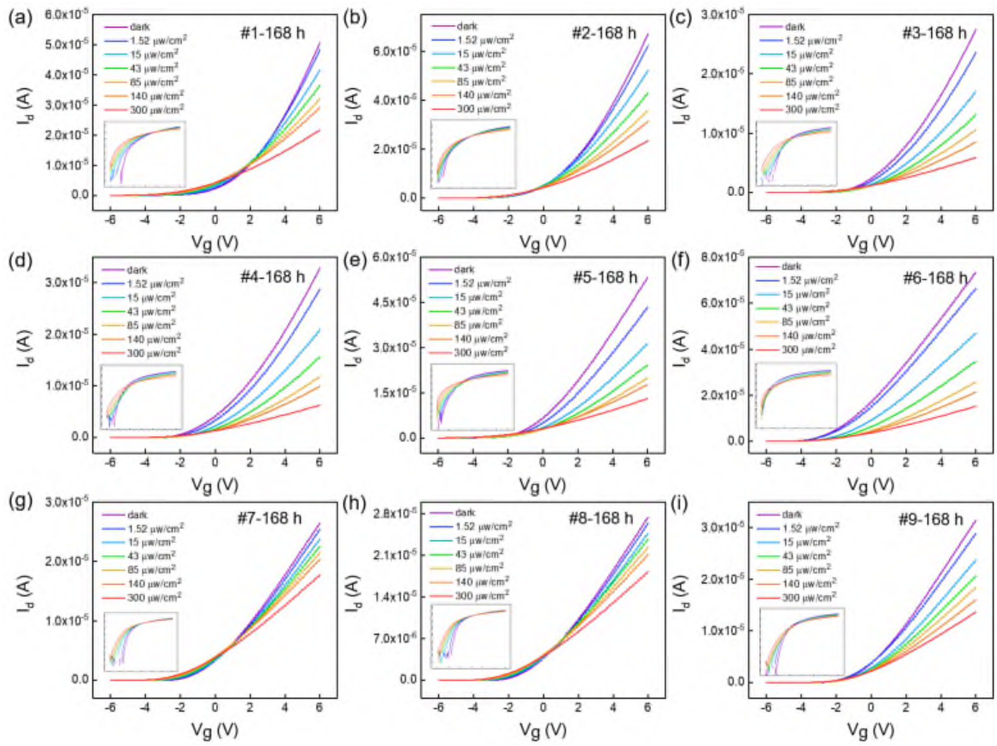


Figure R4. Transfer curves of $MAPbI_3/Bi_2O_2Se$ heterotransistor array under the light illumination of 640 nm with varied light intensities tested after one week at atmosphere.

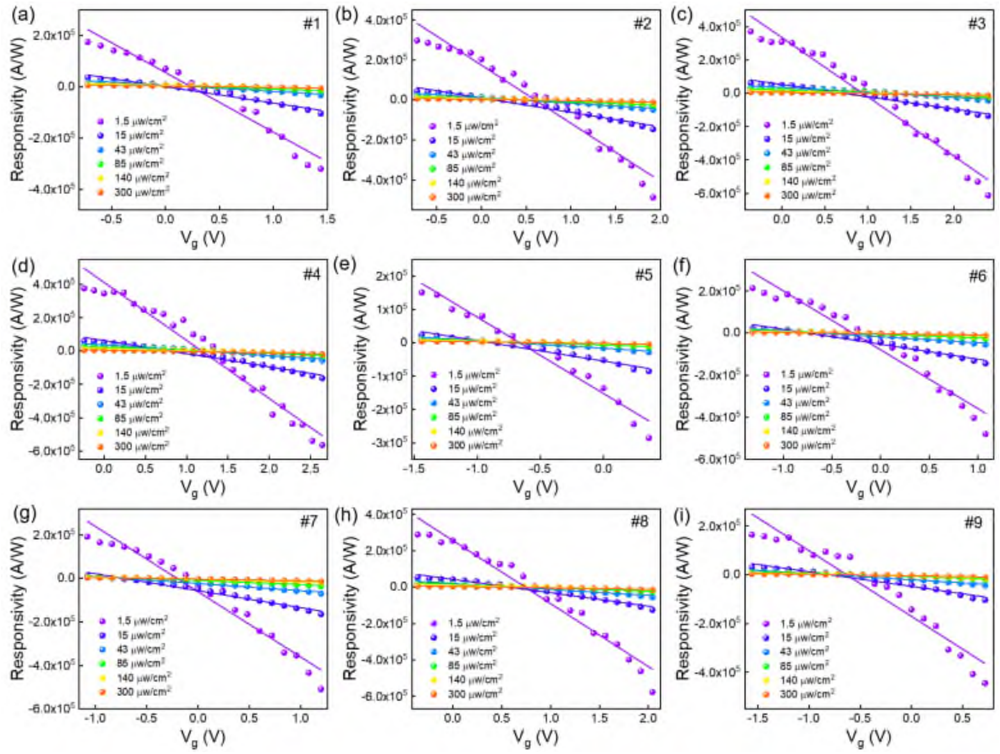


Figure R5. Linear dependence of the bipolar photoresponsivity on the gate voltage under the light illumination of 640 nm with varied light intensities tested after manufacturing.

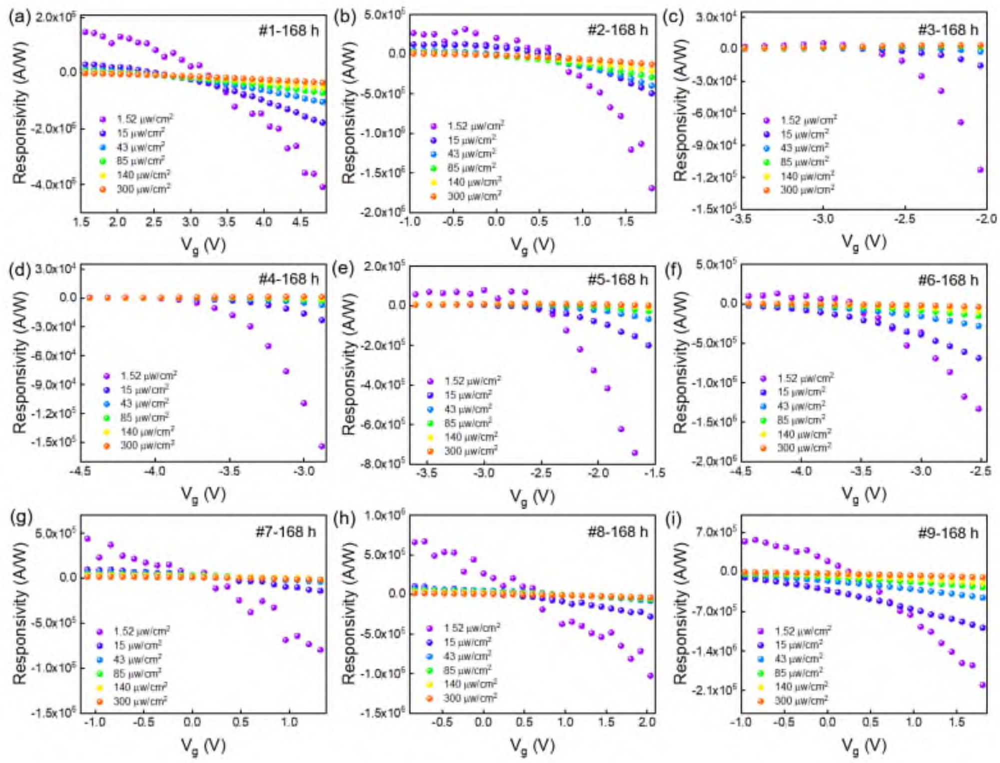


Figure R6. Linear dependence of the bipolar photoresponsivity on the gate voltage under the light illumination of 640 nm with varied light intensities tested after one week at atmosphere.

Corresponding Reference:

[1] Macpherson, S., Doherty, T.A.S., Winchester, A.J.et al. Local nanoscale phase impurities are degradation sites in halide perovskites. *Nature* **607**, 294–300 (2022).

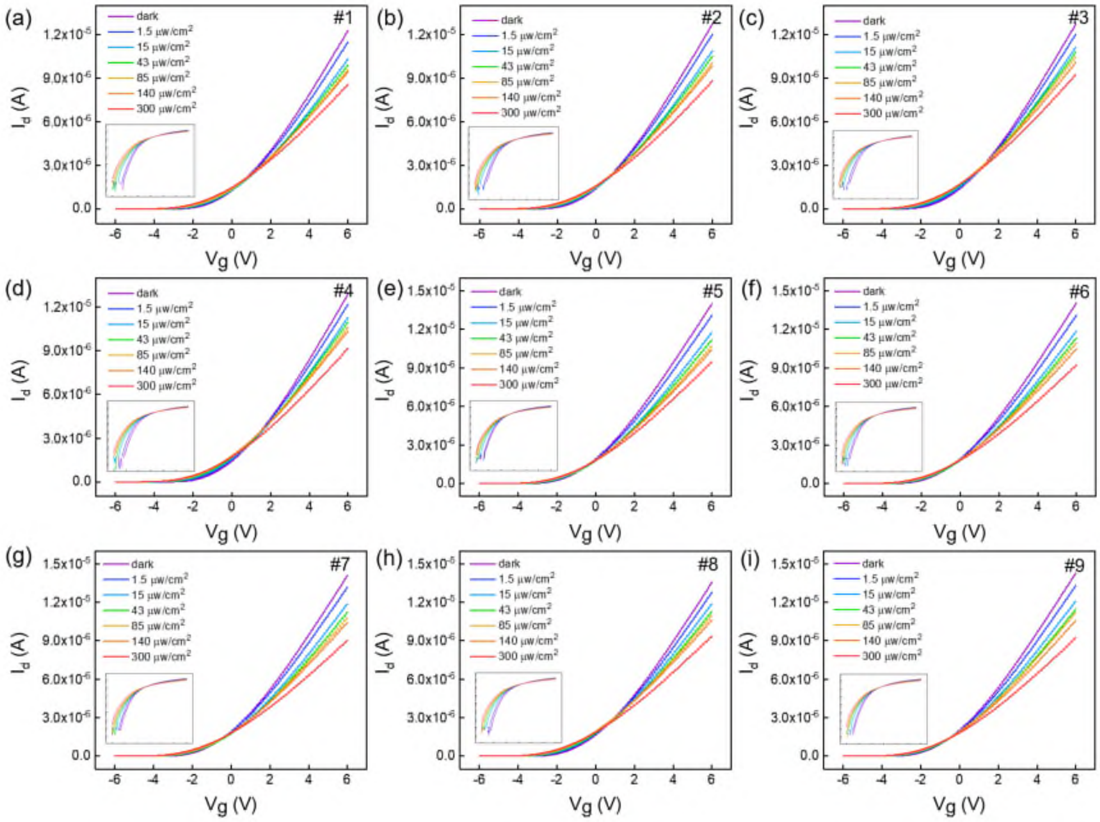
[2] Bertoluzzi L, Boyd C C, Rolston N, et al. Mobile ion concentration measurement and open-access band diagram simulation platform for halide perovskite solar cells. *Joule* **4**, 109-127, (2020).

[3] Lehmann F, Franz A, Többens D M, et al. The phase diagram of a mixed halide (Br, I) hybrid perovskite obtained by synchrotron X-ray diffraction. *RSC Adv.* **9**, 11151-11159, (2019).

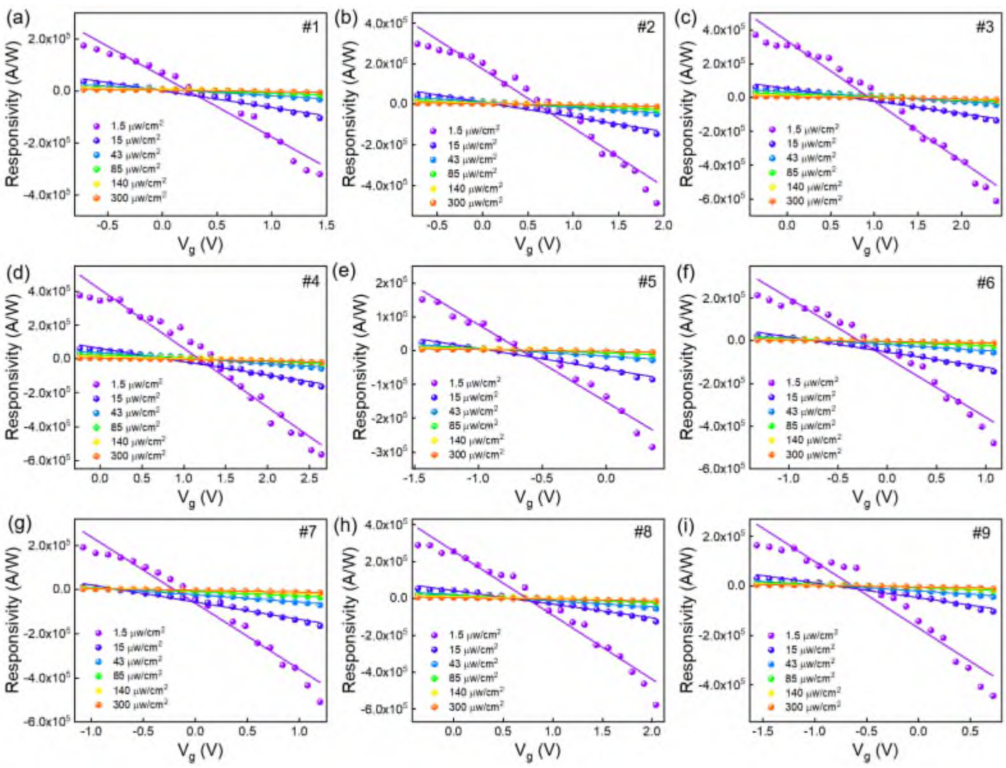
Changes Made:

In the main text (Paragraph 2, Page 8), we have deleted the sentence "The heterotransistors exhibited good stability and reproducibility in modulating bipolar photoresponses as shown in **Fig. 3i** and **Supplementary Fig. S7**." and updated as following (highlighted in red): "We also subjected them to the stability testing. At the very beginning, all the nine devices in the array exhibited clear bipolar photoresponse and demonstrated good uniformity (**Supplementary Fig. S12-S13**). After storing in an ambient environment (i.e., 27±1 °C and 55±5 % humidity) for 168 hours (i.e., one week), all devices maintained mostly the bipolar behavior (**Supplementary Fig. S14-15**). In addition, the device photoresponse to continuous light pulses was tested under different gate voltages, confirming their excellent reproducibility (**Fig. 3i and Supplementary Fig. S16**)".

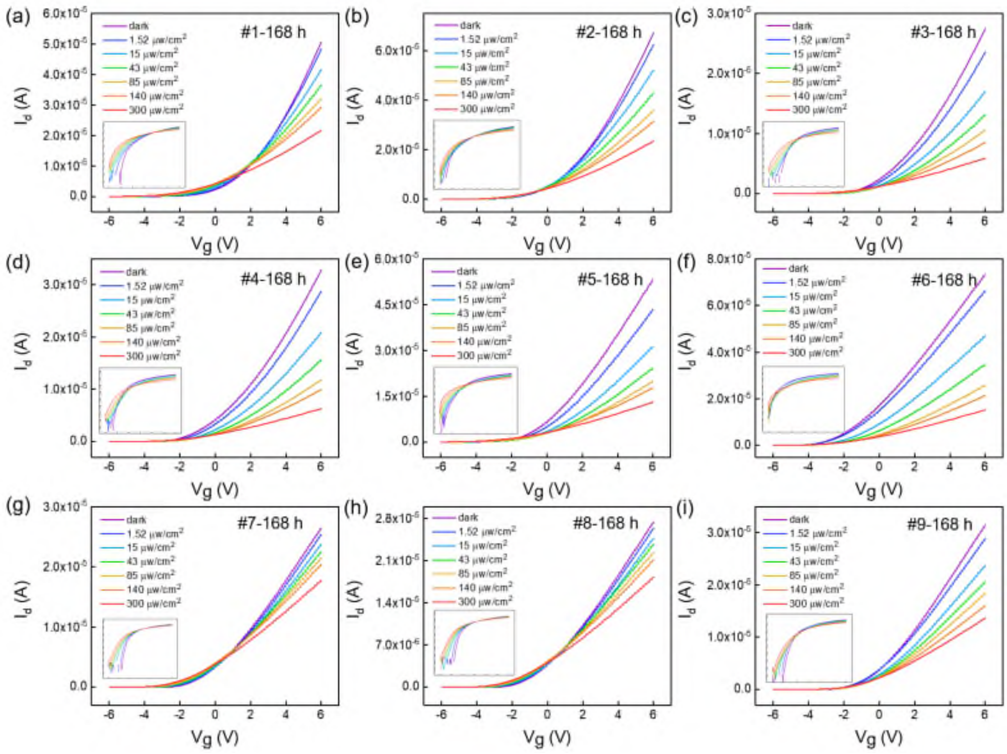
We have included the stability results in **Supplementary Fig. S12-S15**.



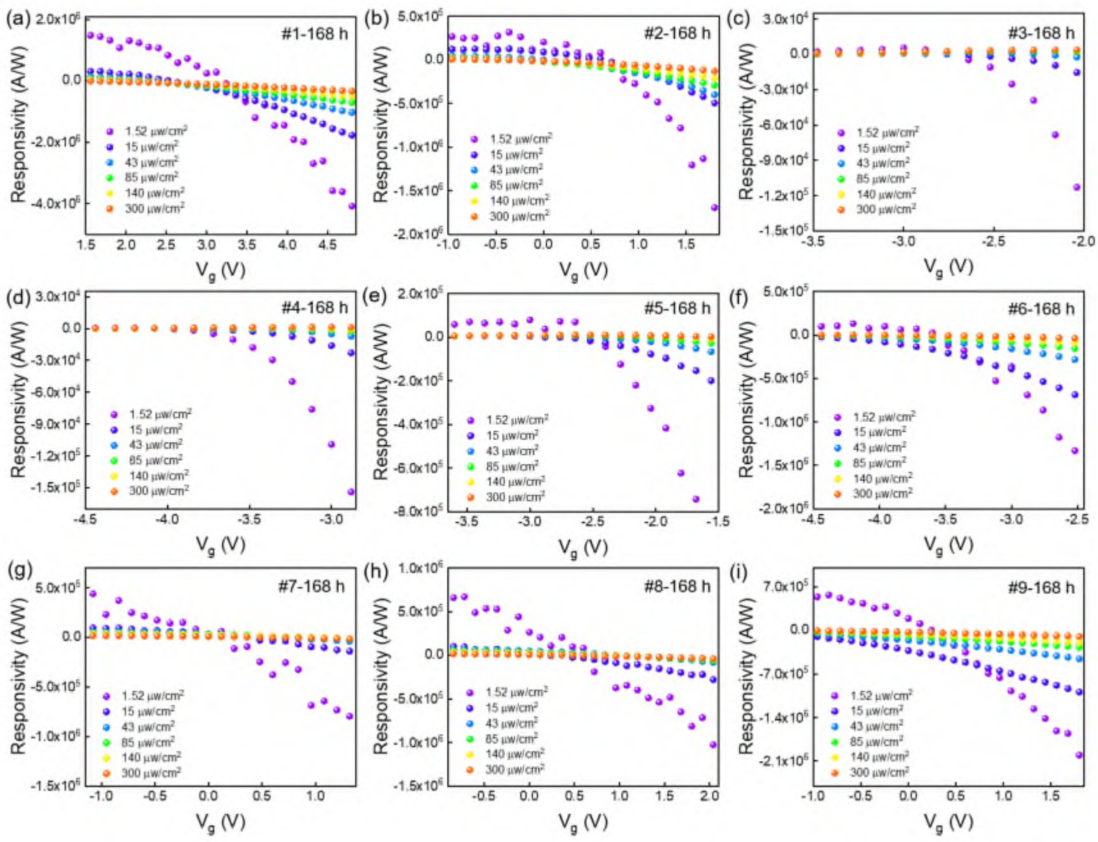
Supplementary Figure S12 | Transfer curves of $MAPbI_3/Bi_2O_2Se$ heterotransistor array under the light illumination of 640 nm with varied light intensities tested after manufacturing.



Supplementary Figure S13 | Linear dependence of the bipolar photoresponsivity on the gate voltage under the light illumination of 640 nm with varied light intensities tested after manufacturing.



Supplementary Figure S14 | Transfer curves of $MAPbI_3/Bi_2O_2Se$ heterotransistor array under the light illumination of 640 nm with varied light intensities tested after one week at atmosphere.



Supplementary Figure S15 | Linear dependence of the bipolar photoresponsivity on the gate voltage under the light illumination of 640 nm with varied light intensities tested after one week at atmosphere.

===

Comment 1-4: The properties shown in Fig. 3g-h and the related figures in supplementary materials should be supplied for all samples in 3´3 transistor array to demonstrate the uniformity of array.

Response:

We appreciate your valuable suggestion. To demonstrate the uniformity of our heterotransistor array, we characterized both the photoresponsivity and photocurrent of all nine devices within the 3×3 $MAPbI_3/Bi_2O_2Se$ heterotransistor array. **Fig. R7** shows the photoresponsivity of these devices as a function of gate voltage under the illumination of 640 nm, and **Fig. R8** presents the variation of the photocurrent with the light intensity. To delve deeper, we extracted the maximum positive and negative photoresponsivity values from the nine devices (**Fig. R9a**). The data revealed the maximum positive photoresponsivity of $(2.4\pm0.9)\times10^5$ A/W and a maximum negative photoresponsivity of (-4.8±1.1) $\times 10^5$ A/W, underscoring a commendable degree of uniformity (**Fig. R9b**).

Moreover, we investigated the photocurrent distribution across different gate voltages while maintaining a constant light intensity of 1.52 μW/cm^2 (**Fig. R9c**), revealing an excellent uniformity at varied gate voltages. When we increased the light intensity to 300 μW/cm^2, the variation in photocurrent remained consistent with the patterns observed at the lower light density (**Fig. R9d**). This consistency across different testing conditions attests to the remarkable uniformity of the $MAPbI_3/Bi_2O_2Se$ heterotransistor array, reinforcing the validity and reliability of our results.

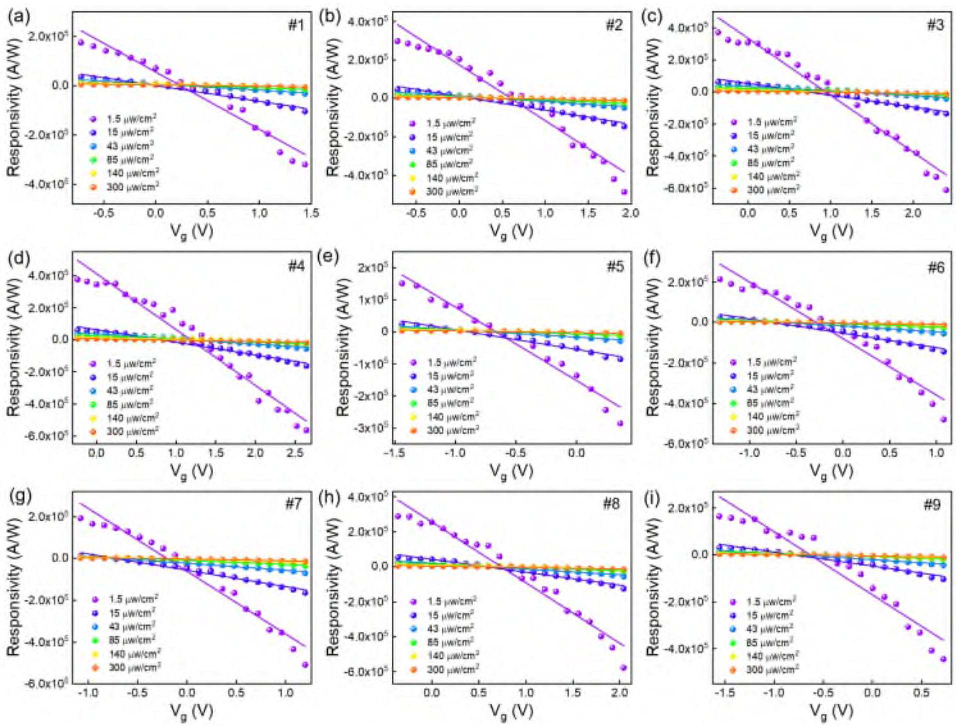


Figure R7. Linear photoresponsivity in response to light intensity across various gate voltages of the 3×3 $MAPbI_3/Bi_2O_2Se$ phototransistor array under the light illumination of 640 nm.

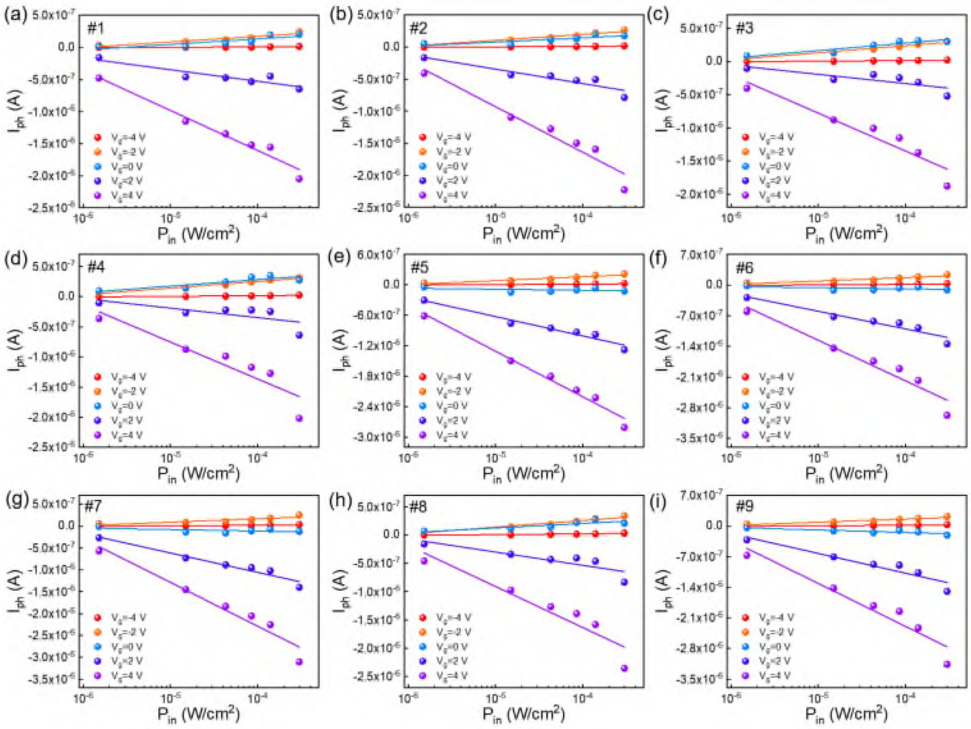


Figure R8. Photocurrent in response to light intensity across various gate voltages of the 3×3 $MAPbI_3/Bi_2O_2Se$ phototransistor array under illumination at 640 nm.

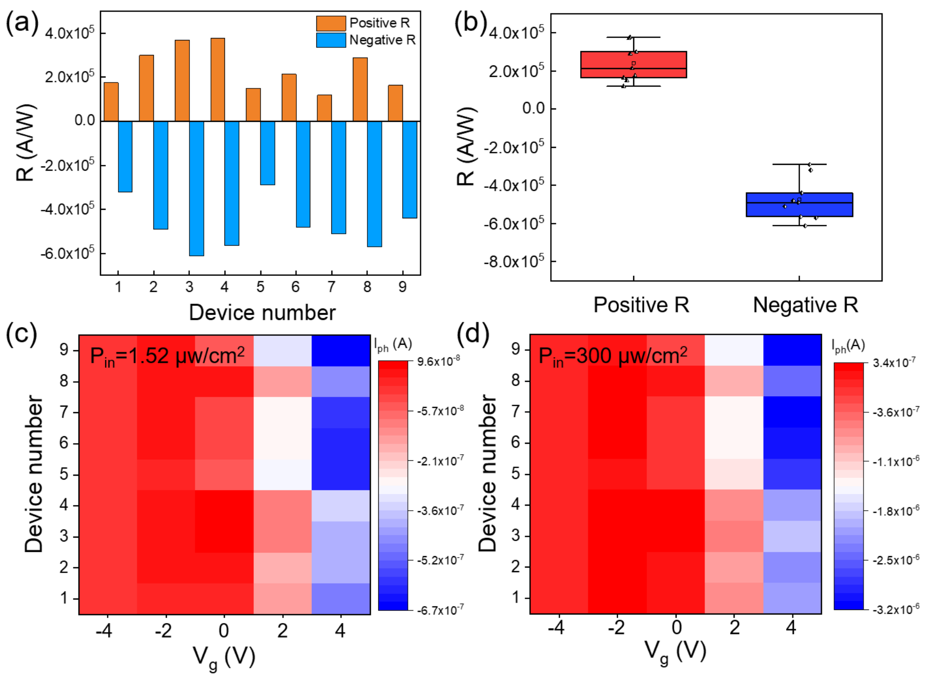
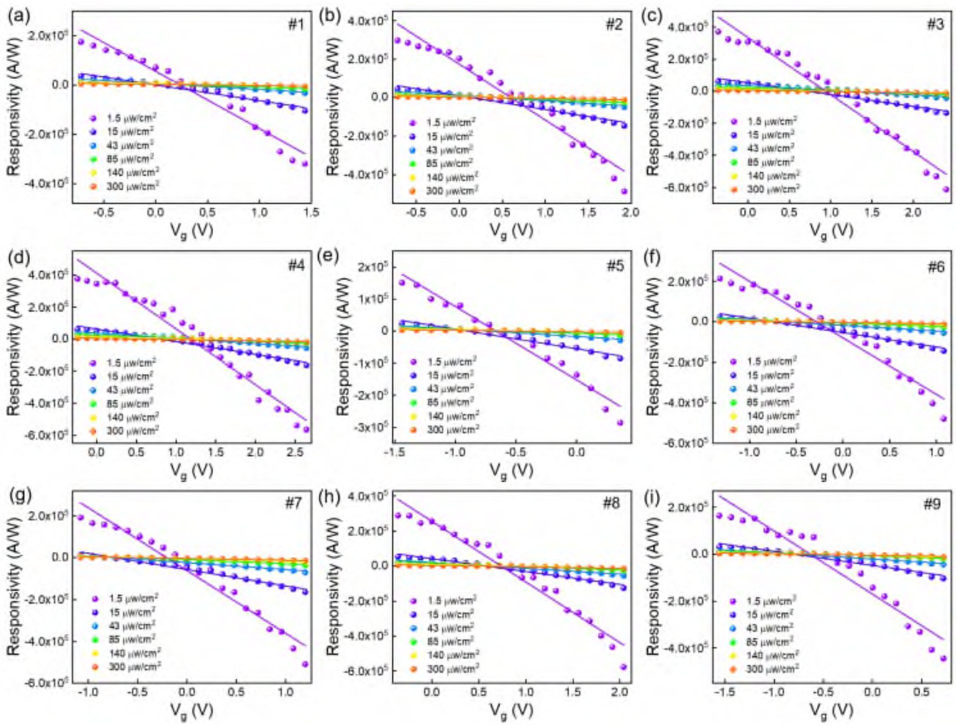


Figure R9. (a) Maximum positive and negative photoresponsivity of the 3×3 $MAPbI_3/Bi_2O_2Se$ phototransistor array. (b) Statistical distribution of maximum positive and negative photoresponsivity across the nine devices. (c) Photocurrent distribution across the nine devices as a function of gate voltage at the light intensity of 1.52 μW/cm². (d) Photocurrent distribution across the nine devices as a function of gate voltage at the light intensity of 300 μW/cm².

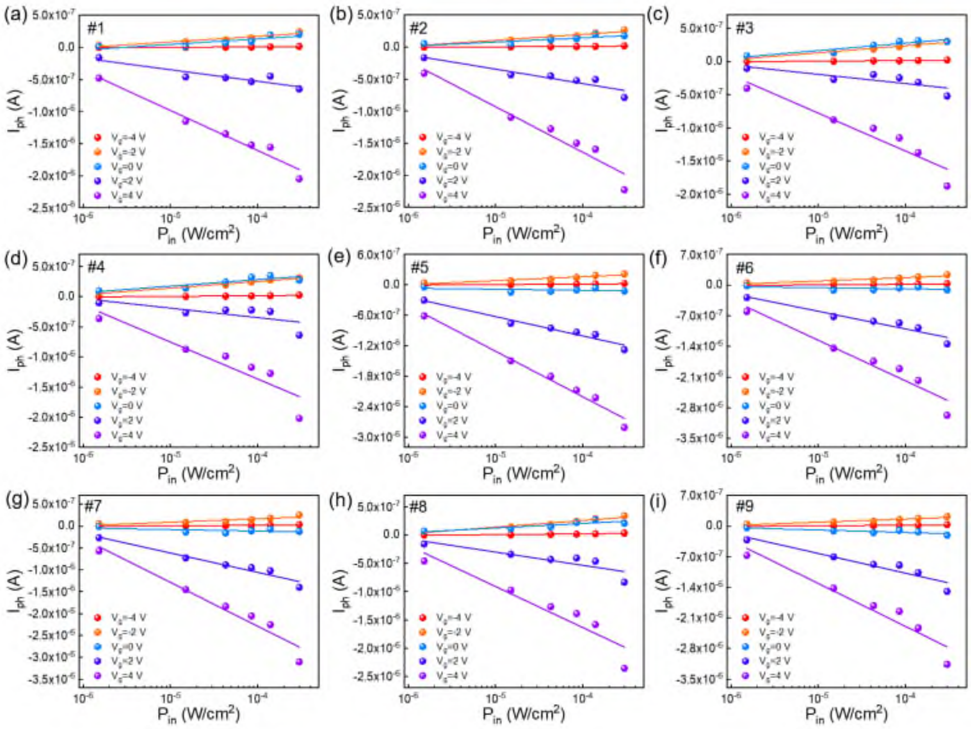
Changes Made:

In the main text (Paragraph 2, Page 8), we have included the following context: "Moreover, we comprehensively analyzed both the photoresponsivity and photocurrent of all nine devices within the 3×3 $MAPbI_3/Bi_2O_2Se$ heterotransistor array under 640 nm illumination (**Supplementary Fig. S9-S10**), revealing that the maximum positive photoresponsivity was consistent in the range of (2.4±0.9) $\times 10^5$ A/W and the maximum negative one maintained in the range of (-4.8±1.1) $\times 10^5$ A/W (**Supplementary Fig. S11a-b**). In addition, all the nine devices within the heterotransistor array maintained uniform photocurrents at varied gate voltages under both low and high light intensities (**Supplementary Fig. S11c-d**)".

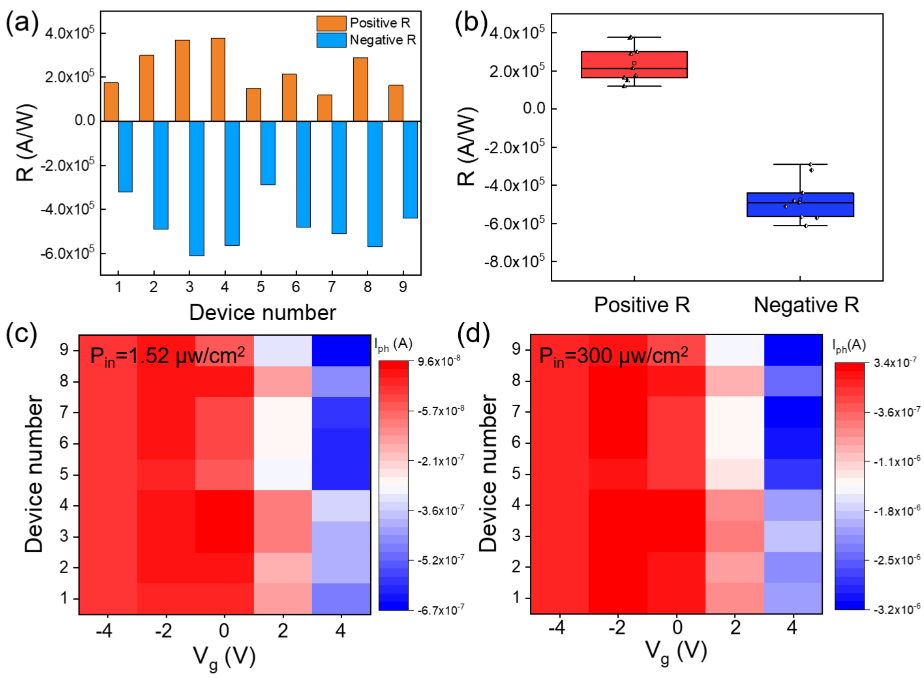
We have included the corresponding **Fig. R7-R9** into the supplementary information as **Supplementary Fig. S9-S11**.



Supplementary Figure S9 | Linear photoresponsivity in response to light intensity across various gate voltages of the 3×3 $MAPbI_3/Bi_2O_2Se$ phototransistor array under the light illumination of 640 nm.



Supplementary Figure S10 | Photocurrent in response to light intensity across various gate voltages of the 3×3 $MAPbI_3/Bi_2O_2Se$ phototransistor array under illumination at 640 nm.



Supplementary Figure S11 | (a) Maximum positive and negative photoresponsivity of the 3×3 $MAPbI_3/Bi_2O_2Se$ phototransistor array. (b) Statistical distribution of maximum positive and negative photoresponsivity across the nine devices. (c) Photocurrent distribution across the nine devices as a function of gate voltage at the light intensity of 1.52 μW/cm². (d) Photocurrent distribution across the nine devices as a function of gate voltage at the light intensity of 300 μW/cm².

===

Comment 1-5: The cross-talk between adjacent pixels should be characterized. A detailed study of photocurrent mapping by scanning photo-current microscopy may be helpful.

Response:

We acknowledge the insightful suggestion. Within our heterotransistor array framework, a critical concern of the crosstalk between pixels is primarily attributed to photogenerated current leakage. In fact, the intrinsic low-mobility characteristic of perovskite layer is expected to increase electrical resistance and subsequently reduce the photocurrent leakages. The optoelectronic comparation of the $MAPbI_3/Bi_2O_2Se$ heterotransistor and $MAPbI_3$-based photodetector under varied gate voltages is shown in **Fig. R10a-c**. As observed in our experiments, the electric current and photocurrent in the heterotransistor were at least three orders of magnitude higher than those in the single perovskite device. This indicates that the predominant photocurrent conduction occurs through the $MAPbI_3/Bi_2O_2Se$ domain, whereby the large electrical resistance and minimal photoresponse of the intervening perovskite layer between pixels act as effective barriers against photocurrent leakages.

To further elucidate the photocurrent distribution across the pixel surroundings, photocurrent mapping was conducted as suggested by the Reviewer. This approach delineates the photocurrent pathways across varying device sectors (**Fig. R10d**), highlighting the significantly higher photocurrent intensity within the $MAPbI_3/Bi_2O_2Se$ heterojunction zone in comparison to adjacent regions. Hence, it can be inferred that the marked electrical resistance and reduced photocurrent of the perovskite layer interspersed between pixels effectively block photocurrent leakages, thereby ensuring the integrity and reliability of sensor-computing functionalities.

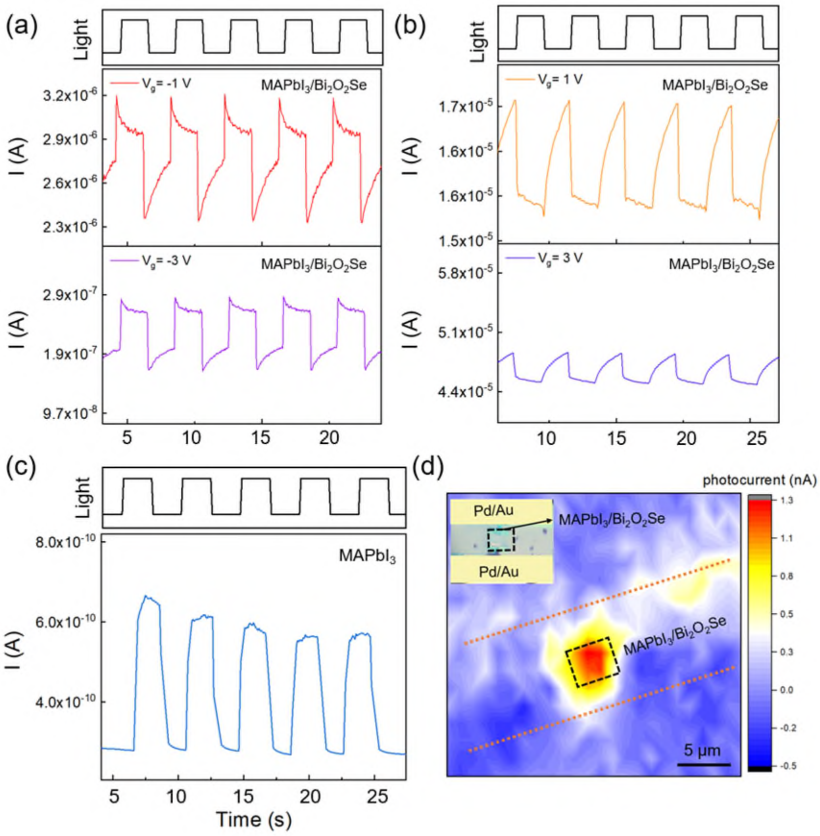
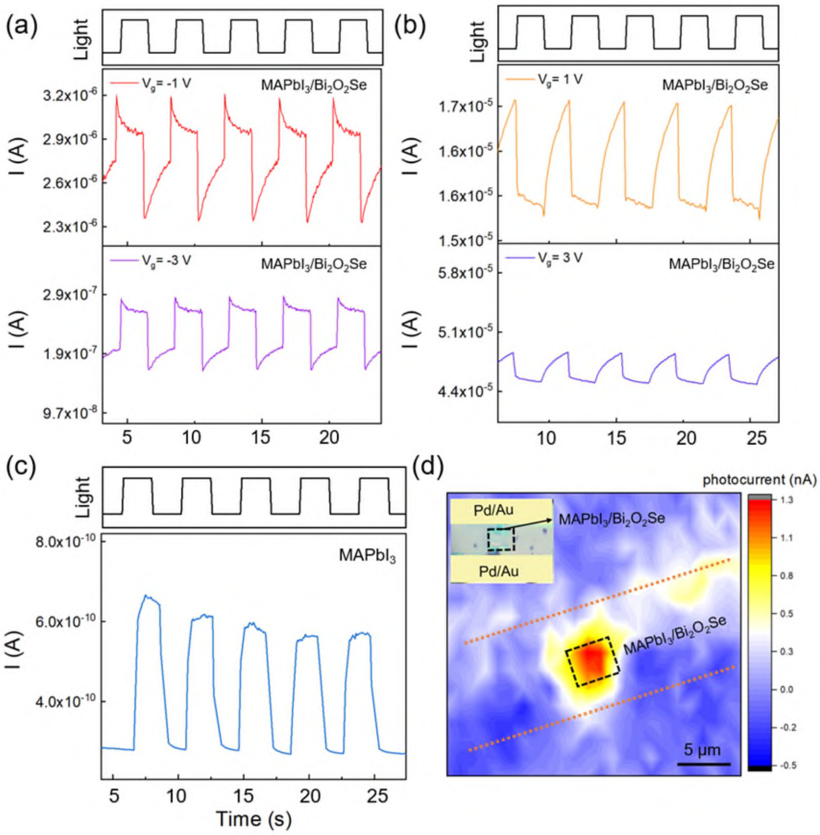


Figure R10. (a) Positive photoresponses of $MAPbI_3/Bi_2O_2Se$ heterotransistor under the gate voltage of V_g=-1 V and V_g=-3 V. (b) Negative photoresponses of $MAPbI_3/Bi_2O_2Se$ heterotransistor under the gate voltage of V_g=1 V and V_g=3 V. (c) Photoresponses of single $MAPbI_3$ photosensor. (d) Photocurrent mapping of the $MAPbI_3/Bi_2O_2Se$ heterotransistor and its surrounding regions. All optoelectronic measurements were conducted using a 532 nm laser.

Changes Made:

In the main text (Paragraph 3, Page 9), we have added the following results and discuss (highlighted in red): "The photocurrents through the $MAPbI_3/Bi_2O_2Se$ heterotransistor surpassed those through the singular perovskite device by at least three orders of magnitude (**Supplementary Fig. S20a-c**). Photocurrent mapping further confirms the significantly high photocurrent intensity within the $MAPbI_3/Bi_2O_2Se$ heterojunction zone in comparison to adjacent regions (**Supplementary Fig. S20d**). This indicates that the predominant photocurrent conduction occurs through the $MAPbI_3/Bi_2O_2Se$ domain, whereby the large electrical resistance and minimal photoresponse of the intervening perovskite layer between pixels act as effective barriers against photocurrent leakages".

We have included the characterizations of device photocurrents and photocurrent mapping of **Fig. R10** in the supplementary information as **Supplementary Fig. S20**.



Supplementary Figure S20 | (a) Positive photoresponses of $MAPbI_3/Bi_2O_2Se$ heterotransistor under the gate voltage of V_g=-1 V and V_g=-3 V. (b) Negative photoresponses of $MAPbI_3/Bi_2O_2Se$ heterotransistor under the gate voltage of V_g=1 V and V_g=3 V. (c) Photoresponses of single $MAPbI_3$ photosensor. (d) Photocurrent mapping of the $MAPbI_3/Bi_2O_2Se$ heterotransistor and its surrounding regions. All optoelectronic measurements were conducted using a 532 nm laser.

==

Comment 1-6: The “in-sensor letter classification” and “in-sensor convolution kernel” functions have been reported in previous article (For example, “Ultrafast machine vision with 2D material neural network image sensors”, Nature, 2020, 579, 62-66, or “Gate-tunable van der Waals heterostructure for reconfigurable neural network vision sensor”, Sci. Adv. 2020; 6: eaba6173.). Could the authors make a brief comparison with these reported devices?

Response:

In comparation of our work with these two previous papers, our discourse analysis is detailed as follows:

Fundamental Principle: Across the board, the operational principle of in-sensor computing adopted by both our device array (i, j) and those in prior studies, adheres to the formula $I_{out} = \sum_{ij} R_{ij} P_{ij}$, where I_{out} is the output photocurrent, R is the photoresponsivity, and P is the incident light intensity. Upon illumination, the heterotransistor array's output current emerges as the multiply accumulation of light intensity and photoresponsivity.

Distinguishing Aspect: The major divergence between our device and those previously reported lies in the principle of photoresponsivity adjustment. Our approach harnesses the synergy between the excellent absorption of perovskite and the superior photocarrier transport of Bi_2O_2Se, significantly promoting the photoresponsivity spectrum to an unprecedented 10^6 A/W. This enhancement not only capacitates our device towards responding to dim light intensities, but also fortifies its resistance against noises. In stark contrast, existing literature (e.g., *Nature*, **2020**, *579*, 62-66) caps the photoresponsivity adjustment at 0.05 A/W, and the absence of detailed photoresponsivity adjustment curves in the *Sci. Adv.* paper (*Sci. Adv.*, **2020**, *6*, eaba6173).

Application in Letter Classification: When it comes to the realm of letter recognition, the study in the *Nature* paper (*Nature*, **2020**, *579*, 62-66) opts for a light wavelength of 640 nm at a maximal light intensity threshold of 0.1 W/cm². On the other hand, the *Sci. Adv.* paper (*Sci. Adv.*, **2020**, *6*, eaba6173) abstains from specifying the wavelength and light intensity employed. Our $MAPbI_3/Bi_2O_2Se$ heterotransistor distinguishes itself by facilitating letter classifications at the remarkably low light intensity of 0.1 μW/cm². In fact, we extended these previous works by scrutinizing the impacts of photoresponsivity on letter-recognizing accuracy, yet unexplored in extant literature. As illustrated in **Fig. 4d-e**, specific letters that represented through varying light intensities were imprinted onto the heterotransistor array via laser irradiation, and the photosensitivity weights were modulated by adjusting gate voltages. **Fig. 4f-g** illustrates that lower photosensitivity engendered obvious fluctuations and impeded convergences, thereby complicating letter recognitions. In contrast, high photoresponsivity in the range 10^4 A/W and 10^6 A/W, led to a swift decrease in the loss function, culminating in 100% recognition accuracy with a reduced number of training epochs.

Dim-Light In-Sensor Convolution Kernel: convolution-kernel processing can perform preprocessing functions for images, such as edge detection, denoising, and formatting. Device performance significantly affects the quality of image preprocessing. Our high photoresponsivity of $MAPbI_3/Bi_2O_2Se$ heterotransistor enables dim-light in-sensor convolutions, which was not demonstrated in these two papers. **Fig. R11** shows the images captured at Peking University under bright and dim scenes, respectively. The pixel information of images was translated to optical signals, and projected onto the reconfigurable photosensors set as the Laplacian operator for edge detection: (1) V_g was set to -2 V/-0.2 V as the kernel weights correlated with the high bipolar photoresponsivities of 2.4×10^5 A/W and -0.3×10^5 A/W, respectively; and (2) V_g was set to -0.6 V/-0.5 V as the kernel weights correlated with the low bipolar photoresponsivities of 4.6×10^4 A/W and -5.8×10^3 A/W, respectively. The high-responsive sensor detected more detailed edges of the image. Notably, the in-sensor convolution preprocessing of dim-light images has not been reported in other papers.

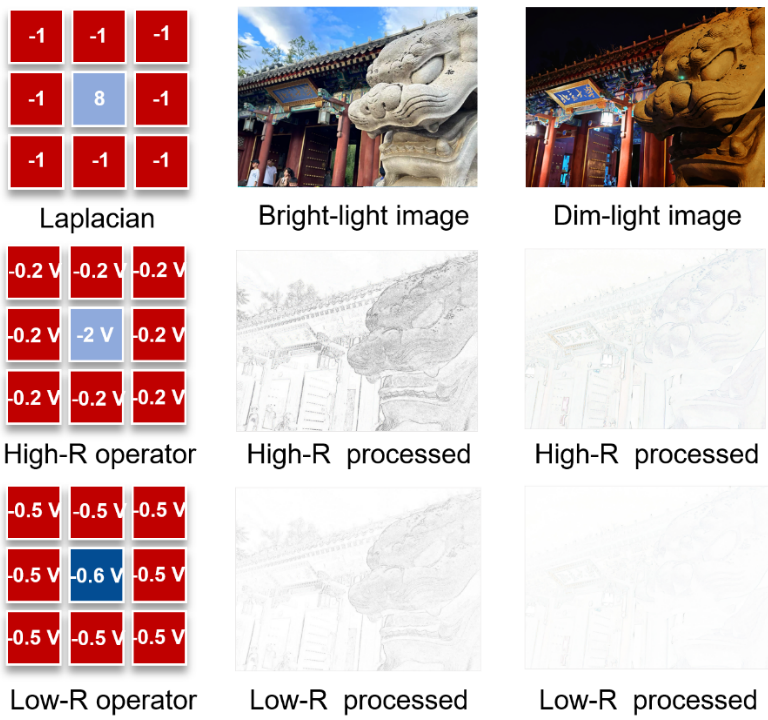
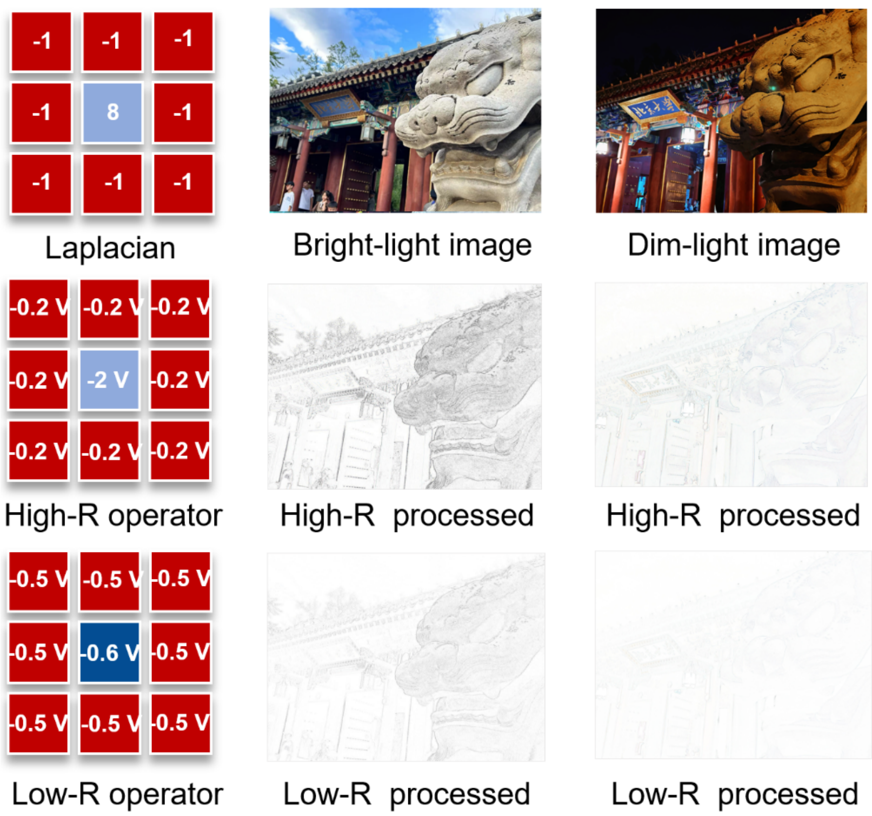


Figure R11. Implementation of dim-light in-sensor edge detection. Edge detection of images under bright and dim lights by the reconfigurable photosensors with high photoresponsivities and low photoresponsivities, respectively.

Changes Made:

In the main text (Paragraph 2, Page 12), we have addressed distinguishing aspect of novelty of our heterotransistor for in-sensor data processing in the context (highlighted in red): "To evaluate the effect of varying photoresponsivity levels on image processing, we configured the array into Laplacian convolution kernels by adjusting the gate voltages. As shown in **Supplementary Fig. S23**, the high-responsivity configurations successfully detected edges in both bright and dim images, whereas the low-responsivity configurations led to significant edge information loss in dim images. The ultrasensitive reconfigurable heterotransistor array holds promises for improving dim-light visual perception in CNNs. [2,48]" We have added **Figure R11** in the supplementary information as **Supplementary Fig. S23**, and cited these two papers as **Reference 2** and **48**.



Supplementary Figure S23 | Edge detection results for bright and dim images using $MAPbI_3/Bi_2O_2Se$ reconfigurable sensor array with high photoresponsivities and low photoresponsivities, respectively.

==

Comment 1-7: To demonstrate the capabilities of key feature extraction and traffic light detection in dim-light environments, the intensity of the dim-light image should be specific as described in the letter classification task.

Response:

We are sorry for not clarifying the light intensity of the dim-light images. To rectify this, we note that the dataset employed spans a diverse array of lighting conditions, from the dusky hues of twilight to the diffuse light of overcast skies. The illuminance levels across these scenarios fluctuate between approximately 10 lux to 10^3 lux. Translating this into the metric of optical power density, given that 1 lux corresponds to 0.146 μW/cm^2, the optical power density pertinent to our dataset oscillates from 1.46 μW/cm^2 to 146 μW/cm^2.

Changes Made:

We have enriched **Supplementary Note 3** with additional information pertinent to the preprocessing methodology employed for traffic light detection under dimly lit conditions (highlighted in red). In particular, we have described the range of incident light intensities ranging from 1.5 to 150 μW/cm^2. This specified range is integral to the execution of key feature extraction and the accurate detection of traffic lights within low-light environments.

==

Comment 1-8: In Figure 5b, the preprocessed digital image appears to show no discernible alterations when compared to the raw digital photograph.

Response:

We acknowledge that the distinctions between the preprocessed images and their raw counterparts might not be immediately apparent to the naked eye. However, a comprehensive analysis of pixel distribution offers insightful revelations. The images subjected to noise exhibit characteristics such as blurring and a notable reduction in detail, accompanied by a disorderly appearance of colors attributable to the introduction of random noise (**Fig. R12a**). A closer examination of the histogram in **Fig. R12b** illustrates that the superimposed Gaussian noise manifests as a random overlay that obscures the discrete features of the original image, adhering to a normal distribution pattern. This results in a histogram characterized by a smooth curve, signifying that the local variations and intricate details within the image are effectively camouflaged and diminished by the Gaussian noise.

Upon the denoising process, a significant restoration of image detail is observed, culminating in visuals that are markedly clearer and a more coherent color scheme (**Fig. R12c**). The histogram evolution from a singular-peaked distribution representative of the noise-afflicted image to a multi-peaked distribution post-denoising, coupled with a contraction in the width of the Gaussian distribution (**Fig. R12d**), is indicative of a partial reinstatement of the information structure and minute details. This analytical discourse underscores the pivotal role of the preprocessing steps, notably noise reduction and contrast enhancement, in mitigating noise interference within the images. These measures are instrumental in bolstering the precision of key feature extraction and enhancing the efficacy of traffic light detection in environments plagued by low-light conditions.

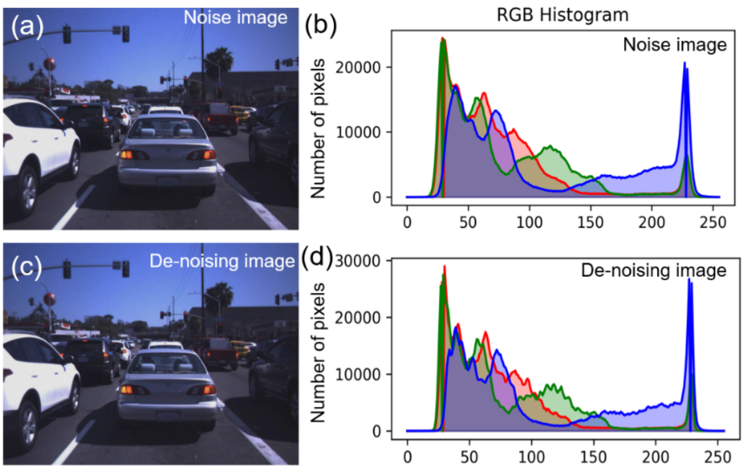
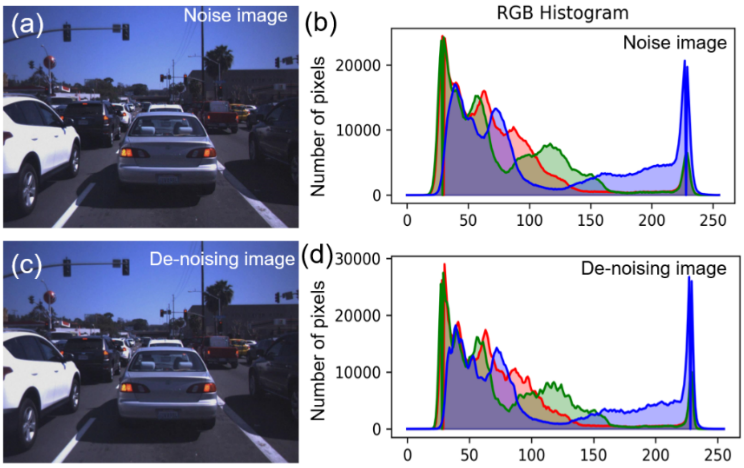


Figure R12. Visual comparison of images before and after noise reduction. (a) Original image with Gaussian noises. (b) Histogram of the noised image, showing a smooth and single-peaked distribution. (c) Image after denoising, with improved detail and clarity. (d) Histogram of the denoised image, displaying a multi-peaked distribution that reflects the restoration of image details.

Changes Made:

In the main text (Paragraph 1, Page 13), we have added the following description to illustrate the differences before and after denoising: "As illustrated in **Supplementary Fig. S24**, the presence of noise within RGB imagery precipitates a blurring of details and a distortion of color fidelity. Conversely, the application of denoising process facilitates a restoration of visual clarity and intricacy, a transformation underscored by the histogram evolution from a homogenously smooth distribution to one characterized by multiple peaks. The integration of denoising into the preprocessing regimen significantly elevates the quality of the image, simultaneously expediting the training convergence of neural network".

We have also included **Fig. R12** in the supplementary information as **Supplementary Figure S24**.



Supplementary Figure S24 | Visual comparison of images before and after noise reduction. (a) Original image with added Gaussian noise. (b) Histogram of the noised image, showing a smooth, single-peaked distribution. (c) Image after denoising, with improved detail and clarity. (d) Histogram of the denoised image, displaying a multi-peaked distribution that reflects the restoration of image details.

==

Comment 1-9: The speed is another important evaluation criterion for object detection. The reader would like to observe the impact of the preprocessing function on the inference speed of the neural network.

Response:

We appreciate the insightful comments from the Reviewer. In our methodology, we employ Gaussian convolutional kernels for denoising input images as a preliminary step. Our preprocessing notably improves the image quality, and thus enhances the precision of object detection in the dim-light conditions. However, this step does not directly contribute to an increase in the inference speed of neural networks, primarily due to the lack of image compression during the denoising process. Yet, it has been observed that neural networks trained on datasets comprising denoised images demonstrate quicker convergence than those trained on original images.[1] This is attributed to the inherent nature of Gaussian noise, which introduces random fluctuations into the data, compelling the network to undertake additional training iterations to effectively ignore the noise and distill useful features, often resulting in extended convergence times. [2] Conversely, when neural networks are trained on noise-free datasets, the data is inherently cleaner and the features more distinguishable, facilitating a more expedient feature extraction process and thus a swifter convergence. [3]

To accelerate object detections, several strategies can be adopted. One approach involves the adoption of lightweight models or the compression of existing models to reduce the number of model parameters and computational demands, thereby increasing processing speed.[4] For instance, Knowledge Distillation techniques allow for the transference of knowledge from a larger, more complex teacher network to a smaller, more efficient student network, which can then be utilized for inference processes, thus optimizing efficiency.[5] Additionally, employing hardware accelerators such as GPUs for the deployment of networks facilitates the parallel processing of extensive computations, significantly boosting algorithm execution speeds.[6] Furthermore, the application of Quantization techniques, which entail the quantization of network parameters and activation functions, can further diminish computational burdens and enhance inference speeds.[7-8]

Corresponding Reference:

[1]. Rodríguez-Rodríguez, J.A., López-Rubio, E., Ángel-Ruiz, J.A. & Molina-Cabello, M.A. The Impact of Noise and Brightness on Object Detection Methods. *Sensors* **24**, 821 (2024).

[2]. Chuah, J.H., Khaw, H.Y., Soon, F.C. & Chow, C.O. I, *TENCON 2017 - 2017 IEEE Region 10 Conference*, 2447-2450 (2017).

[3]. Liang, J., Chen, P. & Wu, M. Research on an Image Denoising Algorithm based on Deep Network Learning. *J. Phy: Conf. Ser.* **1802**, 032112 (2021).

[4] Guo, J.-M., Yang, J.-S., Seshathiri, S. & Wu, H.-W. A Light-Weight CNN for Object Detection with Sparse Model and Knowledge Distillation. *Electronics* **11**, 575 (2022).

[5]. Xing, Z., Chen, X. & Pang, F. DD-YOLO: An object detection method combining

knowledge distillation and Differentiable Architecture Search. *IET Comput. Vis.* **16**, 418-430 (2022).

[6]. Fukagai, T., Maeda, K., Tanabe, S., Shirahata, K., Tomita, Y., Ike, A., & Nakagawa, A. (2018, October). Speed-up of object detection neural network with GPU. *25th IEEE International conference on image processing (ICIP)*, pp. 301-305 (2018).

[7]. Nguyen, D.T., Nguyen, T.N., Kim, H. & Lee, H.-J. A high-throughput and power-efficient FPGA implementation of YOLO CNN for object detection. *IEEE Transactions on Very LargeScale Integration (VLSI) Systems* **27**, 1861-1873 (2019).

[8]. Li, R., Wang, Y., Liang, F., Qin, H., Yan, J., & Fan, R. Fully quantized network for object detection. *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR)*, pp. 2810-2819 (2019).

Changes Made:

In the main text (Paragraph 1, Page 13), we have noted the following discussion (highlighted in red)："This efficiency gain is attributed to the diminished necessity for the network to adaptively learn the exclusion of noise artifacts, thereby streamlining the computational process and enhancing the overall performance of the neural network in machine-learning applications" .

==

Comment 1-10: Can author estimate the energy consumption of device during one light pulse, or energy consumption of sensor array for preprocessing functions?

Response：

At present, there is usually an absence of comprehensive reporting on the energy consumption metrics for bipolar photoresponse devices in the references. These devices characteristically necessitate the applications of high voltages and substantial power demands. This situation underscores a gap in the research pertaining to the energy consumption of bipolar photoresponse devices, with a particular emphasis needed on the analysis of energy utilization by peripheral circuits, which often constitute a predominant portion of the overall energy expenditure. In response to the Reviewer's inquiry, we suggest a simply bifurcated model to conceptualize energy consumption in such scenarios. Initially, the device necessitates biasing to achieve the desired photoresponsivity, a process inherently linked to energy expenditure. This aspect of energy consumption can be quantitatively expressed as $E_1 = V_g \times I_g \times T_{duration}$, where V_g is the gate voltage essential for sustaining photoresponsivity, I_g is the gate current, and $T_{duration}$ is the duration of a single light pulse operation. The secondary term of energy consumption pertains to the in-sensor computing phase, which can be encapsulated as $E_2 = I_{light} \times V_{ds} \times T_{duration}$, where I_{light} is the photocurrent, and V_{ds} is the source-drain voltage. Employing a $T_{duration}$ of 100 ms, we have estimated the energy consumption of the device that ranges between 10^{-3} nJ to 1 nJ (**Table R2**).

Table R2. Energy consumption of the $MAPbI_3/Bi_2O_2Se$ heterotransistor under different gate voltages for one light pulse operation.

V_G (V)	E_1 (J)	E_2 (J)	E_{total} (J)
-4	3.4×10^{-12}	2.0×10^{-12}	5.4×10^{-12}
-2	2.2×10^{-12}	5.0×10^{-10}	5.0×10^{-10}
0	0	5.6×10^{-10}	5.6×10^{-10}
2	8.0×10^{-13}	1.7×10^{-9}	1.7×10^{-9}
4	2.0×10^{-13}	4.0×10^{-9}	4.0×10^{-9}

Changes Made:

In the main text (Paragraph 1, Page 10), we have included the following discussion (highlighted in red): "In quantifying the energy consumption of the heterotransistor device throughout a singular light pulse episode, we suppose the energy consumption into two principal components: (i) the energy E_1 requisite for sustaining photoresponsivity as $E_1 = V_g \times I_g \times T_{duration}$, where V_g is the gate voltage essential for sustaining photoresponsivity, I_g is the gate current, and $T_{duration}$ is the duration of a single light pulse operation; and (ii) the energy E_2 inherent to the in-sensor computing functions as $E_2 = I_{light} \times V_{ds} \times T_{duration}$, where I_{light} is the photocurrent, and V_{ds} is the source-drain voltage. Employing the $T_{duration}$ parameter set at 100 ms, we estimate the energy consumption spans from 10^{-3} nJ to 1 nJ dependent on the varied gate voltages”.

===

Comment 1-11: The authors should clearly explain the preprocessing procedure for color of traffic light using the sensor array in Method part or Supplementary information.

Response:

We are sorry for not clearly describing the preprocessing procedure. In detail, we employed a 3×3 sensor array to formulate a Gaussian filter operator, aimed at denoising images marred by noise. For the task of traffic-light detection, where the input images feature three RGB channels, we executed a convolution of each RGB channel individually with the Gaussian kernel. The images in the LISA Traffic Lights dataset possess a resolution of 1280×960 pixels, and to simulate realistic conditions, we introduced Gaussian noises to these images prior to their use in the traffic light recognition task. The Gaussian smoothing convolution operator in the size of 3×3 was designed based on the distribution patterns of a 2D Gaussian function, ensuring that the center pixel exerts the most significant influence on the outcome, with a progressively diminishing contribution from the peripheral pixels. This design facilitates a smoothing effect, with a typical discrete approximation of the weight distribution depicted in **Fig. 5a**, displaying the center: edge: corner weight ratio of 4:2:1 for the 3×3 matrix.

Given the tri-channel nature of the input images, we deployed three distinct 3×3 operators tailored for the red, green, and blue channels, respectively. For the red channel, a 780 nm wavelength was chosen, adjusting the biases for the central, edge, and corner pixels within the 3×3 sensor array to V_g=-1.3 V, -0.8 V, and -0.7 V, respectively. This resulted in varying photoresponsivity R of 1.0×10^{7} A/W, 5.0×10^{6} A/W, and 2.5×10^{6} A/W, respectively. Similarly, wavelengths of 532 nm and 400 nm were selected for the green and blue channels, with the sensor arrays configured accordingly to serve as convolution kernels in image processing. In subsequent simulations, the input images were dissected into their constituent RGB channels, with pixel values linearly transformed from the 0-255 range to correspond to an incident light intensity in the range of 1.5-150 μW/cm². A convolutional kernel was then applied by sliding it across the image, commencing with the initial 3×3 window, and employing a stride of 1. For each 3×3-pixel sub-window, the simulation computed the photocurrent $I_{ph}=\Sigma RP$, where P is the product of pixel value. The output value of each sliding window was then used as the pixel value for the corresponding position in the denoised output image.

Changes Made:

To address the Reviewer's suggestion, we have included a detailed explanation of the preprocessing procedure for traffic-light color detection.

In the main text (Paragraph 2, Page 12), we have added (highlighted in red): "In the preprocessing of traffic-light images, each of the RGB channels underwent individual processing facilitated by a 3×3 bipolar phototransistor array that serves as a Gaussian filter. This approach involved the application of distinct wavelength and bias parameters to each channel, meticulously configured to optimize the convolution and denoising phases of image preprocessing".

The detailed procedure has been clearly described in the **Supplementary Note 3** as following (highlighted in red):

"Supplementary Note 3: Preprocessing for Recognizing Traffic-Light Color

For the image processing and convolutional neural networks (CNNs), convolutional kernels play a pivotal role in tasks such as image blurring, sharpening, edge detection, and the extraction of salient features (e.g., edges, textures, and patterns). These kernels are instrumental in discerning specific attributes within images, thereby facilitating a nuanced understanding and manipulation of visual data. **Fig. 5a** illustrates the implementation of three quintessential convolutional kernels. The inaugural function, inversing, serves as a stylization technique, transforming the visual aesthetics of an image. Subsequently, the denoising function, characterized by its operator weights arranged in accordance with the 2D Gaussian distribution $G=\frac{1}{2\Pi\sigma^2}e^{-\frac{x^2+y^2}{2\sigma^2}}$, is colloquially termed Gaussian blur due to its efficacy in blurring and noise reduction. The terminal function, dedicated to edge detection, employs a discrete form of Laplacian operator $L(x,y)=\frac{\partial^2 I}{\partial x^2}+\frac{\partial^2 I}{\partial y^2}$, where $L(x,y)$ is the operator and I is the pixel intensity of images. The Laplacian operator, characterized by its 2D isotropic measure of the second-order spatial derivative, exhibits high sensitivity to noise, thereby necessitating a preliminary denoising phase to optimize the clarity and precision of edge detection.

We employed a 3×3 sensor array to formulate a Gaussian filter operator, aimed at denoising images marred by noise. For the task of traffic-light detection, where the input images feature three RGB channels, we executed a convolution of each RGB channel individually with the Gaussian kernel. The images in the LISA Traffic Lights dataset possess a resolution of 1280×960 pixels, and to simulate realistic conditions, we introduced Gaussian noises to these images prior to their use in the traffic light recognition task. The Gaussian smoothing convolution operator in the size of 3×3 was designed based on the distribution patterns of a 2D Gaussian function, ensuring that the center pixel exerts the most significant influence on the outcome, with a progressively diminishing contribution from the peripheral pixels. This design facilitates a smoothing effect, with a typical discrete approximation of the weight distribution depicted in **Fig. 5a**, displaying the center: edge: corner weight ratio of 4:2:1 for the 3×3 matrix.

Given the tri-channel nature of the input images, we deployed three distinct 3×3 operators tailored for the red, green, and blue channels, respectively. For the red channel, a 780 nm wavelength was chosen, adjusting the biases for the central, edge, and corner pixels within the 3×3 sensor array to V_g=-1.3 V, -0.8 V, and -0.7 V, respectively. This resulted in varying photoresponsivity R of 1.0×10^7 A/W, 5.0×10^6 A/W, and 2.5×10^6 A/W, respectively. Similarly, wavelengths of 532 nm and 400 nm were selected for the green and blue channels, with the sensor arrays configured accordingly to serve as convolution kernels in image processing. In subsequent simulations, the input images were dissected into their constituent RGB channels, with pixel values linearly transformed from the 0-255 range to correspond to an incident light intensity in the range of 1.5-150 μW/cm². A convolutional kernel was then applied by sliding it across the image, commencing with the initial 3×3 window, and employing a stride of 1. For each 3×3-pixel sub-window, the simulation computed the photocurrent $I_{ph}=\Sigma RP$, where P is the product of pixel value. The output value of each sliding window was then used as the pixel value for the corresponding position in the denoised output image".

==

Comment 1-12: There are some mistakes of figure's number. For example: Line 293, the figure's number "Supplementary Fig. S12" is wrong.

Response:

Thank you for your meticulous correction. In the revised version, we have corrected this error and checked all other figure numbers to ensure they are correct.

==

Comment 1-13: Figure 2b is high-resolution TEM (HRTEM) image, please correct the caption.

Response:

We appreciate your thorough review of our manuscript. We have revised the caption for **Figure 2b** to clearly indicate that it is a cross-sectional high-resolution transmission electron microscopy (HRTEM) image. Additionally, we have also updated the relevant descriptions in the main text to reflect this correction.

Changes Made:

We have revised the caption for **Figure 2b** as the following (highlighted in red): "(b) Cross-sectional high-resolution transmission electron microscopy (HRTEM) image of the $MAPbI_3/Bi_2O_2Se$ heterojunction". Also, we have updated the description in the main text (Paragraph 2, Page 5) as the following (highlighted in red): "The cross-sectional high-resolution transmission electron microscopy (HRTEM) in **Fig. 2b** displays a clear interface between $MAPbI_3$ and Bi_2O_2Se domains".

==

Comment 1-14: The following reference might be considered for comprehensiveness, Nature Communications 2023, 14: 6736.

Response:

We highly value your recommendation and acknowledge the significance of this paper (*Nature Communications* **2023**, *14*:6736) in relation to our research. The cited work offers crucial insights into the advancement and application of neuro-inspired optical sensor arrays. In particular, the exploration of NbS_2/MoS_2 hybrid films for improving photo-induced conductance plasticity and facilitating reduced electrical energy consumption aligns closely with the aims of our study, thereby enriching our discussion and reinforcing the contextual foundation of our work.

Changes Made:

We have cited this paper in the main text as Reference 4 on Paragraph 1, Page 3: “Bioinspired neuromorphic vision (NV) sensor offers a compelling opportunity to reduce data shuttling, latency, and energy consumption by directly assigning specific computing tasks inside the sensor.[1-4]”

Comments from Reviewer #2

Overall Remarks: The manuscript presents a significant advancement in neuromorphic vision sensors by introducing a momentum-conserved reconfigurable phototransistor based on a van der Waals heterojunction between methylammonium lead iodide perovskite and two-dimensional Bi_2O_2Se semiconductor. The developed sensor demonstrates impressive optoelectronic performances with record-high photoresponsivity and specific detectivity, alongside a broad dynamic range suitable for dim-light conditions reminiscent of natural moonlight. The device integrates in-sensor analog multiply-accumulate operations, efficiently enhancing feature extraction and improving detection capabilities under challenging lighting conditions.

The manuscript is well-structured, with a clear and logical progression from the introduction of the concept, through the detailed experimental and simulation work, to the presentation of results and their discussion. The topic is of high relevance to current trends in photodetector development and neuromorphic computing, making it a timely contribution to the field. The integration of the device with a convolutional neural network to demonstrate practical applications adds considerable value to the study. I would recommend the publication of this manuscript in Nature Communications.

Response:

We sincerely appreciate your thorough review and positive comments on our manuscript. We have carefully adopted your valuable suggestion to include more detailed comparison with current leading technologies, thus enhancing the comprehensive quality of this manuscript.

==

Comment 2-1: How scalable is the fabrication process of the $MAPbI_3/Bi_2O_2Se$ heterotransistors? Are there potential challenges in large-scale production?

Response:

To demonstrate the scalability of $MAPbI_3/Bi_2O_2Se$ heterotransistors, we prepared a 3×3 heterotransistor array and evaluated its uniformity. We characterized both the photoresponsivity and photocurrent of all nine devices within the 3×3 $MAPbI_3/Bi_2O_2Se$ heterotransistor array. **Fig. R13** shows the photoresponsivity of these devices as a function of gate voltage under the illumination of 640 nm, and **Fig. R14** presents the variation of the photocurrent with the light intensity. To delve deeper, we extracted the maximum positive and negative photoresponsivity values from the nine devices (**Fig. R15a**). The data revealed the maximum positive photoresponsivity of (2.4±0.9) $\times 10^5$ A/W and a maximum negative photoresponsivity of (-4.8±1.1) $\times 10^5$ A/W, underscoring a commendable degree of uniformity (**Fig. R15b**).

Moreover, we investigated the photocurrent distribution across different gate voltages while maintaining a constant light intensity of 1.52 μW/cm^2 (**Fig. R15c**), revealing an excellent uniformity at varied gate voltages. When we increased the light intensity to 300 μW/cm^2, the variation in photocurrent remained consistent with the patterns observed at the lower light density (**Fig. R15d**). This consistency across different testing conditions attests to the remarkable uniformity of the $MAPbI_3/Bi_2O_2Se$ heterotransistor array, reinforcing the validity and reliability of our results.

In fact, the fabrication methodology employed in the assembly of the $MAPbI_3/Bi_2$O2Se heterotransistor is indeed amenable to upscaling. The construction process embodies the synthesis of a back-gated Bi_2O_2Se FET via a transfer technique alongside the deposition of perovskite films through spin-coating. The inherent simplicity and cost-efficiency of perovskite film solution processing methods furnish them with a significant potential for scalability, especially facilitating economical production on a larger scale. [1-3] Moreover, recent advancements have catalyzed the exploration of synthesis techniques for large-area single-crystalline Bi_2O_2Se, including the employment of symmetrically matched substrates to achieve wafer-scale single crystals and the continuous growth methodology originating from a singular nucleus, [4-6] enabling the fabrication of millimeter-scale single-crystalline thin films.

Nonetheless, several challenges have to be well addressed to the large-scale manufacturing: (1) the environmental stability of halide perovskites requires considerable attention due to their propensity for degradation under various conditions; [7,8] (2) ensuring uniformity in the thickness and composition of perovskite films across extensive areas is imperative to maintain consistent heterostructure functionality; [9,10] and (3) the implementation of clean and non-detrimental transfer methods for Bi_2O_2Se films on a large scale is crucial, since the transfer process, if not carefully managed, can lead to the introduction of defects, surface irregularities, and wrinkles. [11-12]

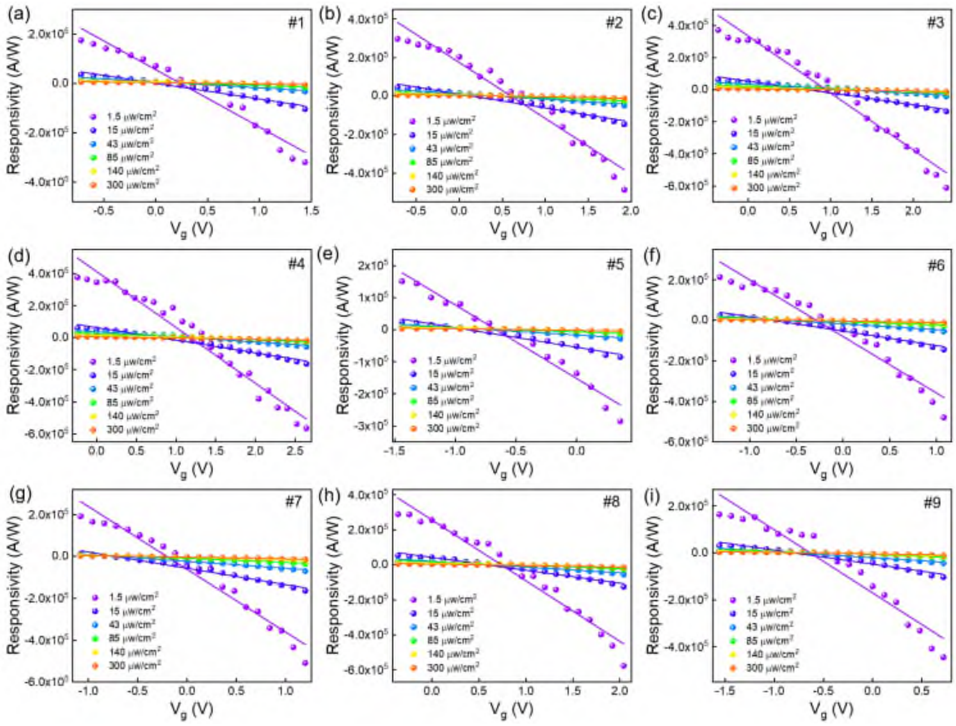


Figure R13. Linear photoresponsivity in response to light intensity across various gate voltages of the 3×3 $MAPbI_3/Bi_2O_2Se$ phototransistor array under the light illumination of 640 nm.

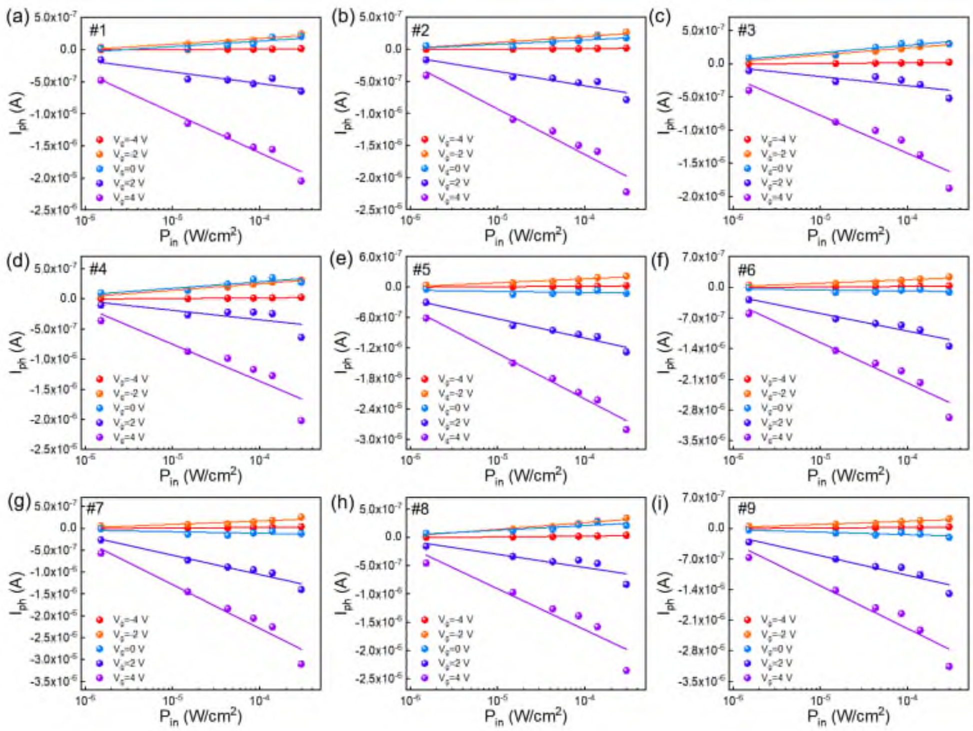


Figure R14. Photocurrent in response to light intensity across various gate voltages of the 3×3

MAPbI3/Bi2O2Se phototransistor array under illumination at 640 nm.

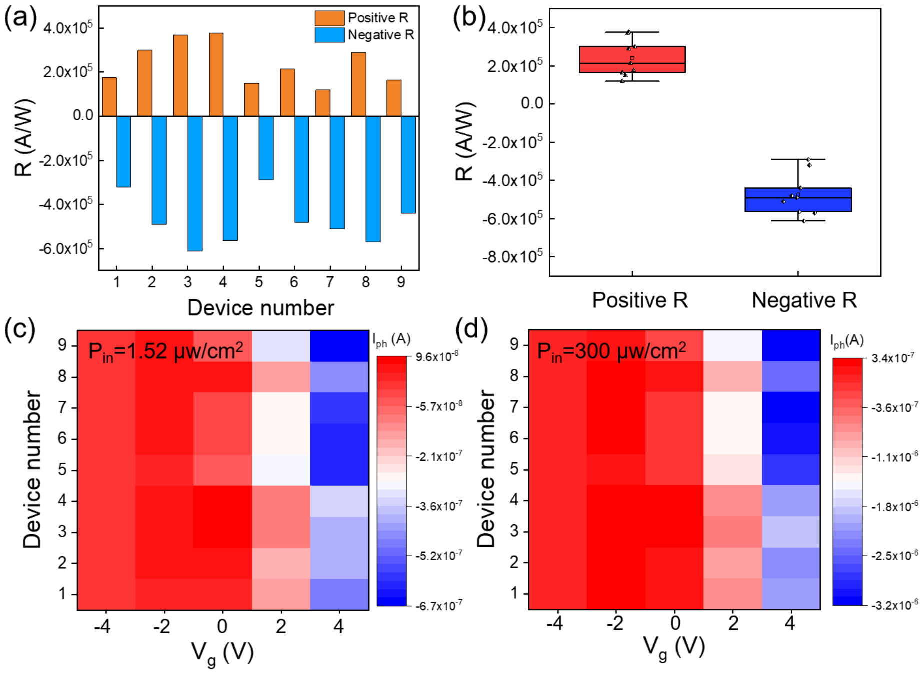


Figure R15. (a) Maximum positive and negative photoresponsivity of the 3×3 $MAPbI_3/Bi_2O_2Se$ phototransistor array. (b) Statistical distribution of maximum positive and negative photoresponsivity across the nine devices. (c) Photocurrent distribution across the nine devices as a function of gate voltage at the light intensity of 1.52 μW/cm². (d) Photocurrent distribution across the nine devices as a function of gate voltage at the light intensity of 300 μW/cm².

Corresponding Reference:

[1] Jana, S., Carlos, E., Panigrahi, S., Martins, R. & Fortunato, E. Toward stable solution-processed high-mobility p-type thin film transistors based on halide perovskites. *ACS Nano* **14**, 14790-14797 (2020).

[2] Younis, A., Lin, C. H., Guan, X., Shahrokhi, S., Huang, C. Y., Wang, Y., ... & Wu, T. Halide perovskites: a new era of solution‐processed electronics. *Adv. Mater*. **33**, 2005000 (2021).

[3]. Qiu, L., He, S., Ono, L.K., Liu, S. & Qi, Y. Scalable fabrication of metal halide perovskite solar cells and modules. *ACS Energy Lett.* **4**, 2147-2167 (2019).

[4] Tan, C., Tang, M., Wu, J., Liu, Y., Li, T., Liang, Y., ... & Peng, H. Wafer-Scale Growth

of Single-Crystal 2D Semiconductor on Perovskite Oxides for High-Performance Transistors. *Nano Lett.* **19**, 2148-2153 (2019).

[5] Yang, X., Zhang, Q., Song, Y., Fan, Y., He, Y., Zhu, Z., ... & Novoselov, K. High Mobility Two-Dimensional Bismuth Oxyselenide Single Crystals with Large Grain Size Grown by Reverse-Flow Chemical Vapor Deposition. *ACS Appl. Mater. & Interfaces* **13**, 49153-49162 (2021).

[6] Khan, U., Luo, Y., Tang, L., Teng, C., Liu, J., Liu, B., & Cheng, H. M. Controlled Vapor–Solid Deposition of Millimeter-Size Single Crystal 2D Bi_2O_2Se for High-Performance Phototransistors. *Adv. Funct. Mater.* **29**, 1807979 (2019).

[7] Kumar, V., Barbé, J., Schmidt, W. L., Tsevas, K., Ozkan, B., Handley, C. M., ... & Rodenburg, C. Stoichiometry-dependent local instability in $MAPbI_3$ perovskite materials and devices. *J. Mater. Chem. A* **6**, 23578-23586 (2018).

[8] Nagabhushana, G., Shivaramaiah, R. & Navrotsky, A. Direct calorimetric verification of thermodynamic instability of lead halide hybrid perovskites. *Proc. Natl. Acad. Sci.* **113**, 7717-7721 (2016).

[9] Xie, L., Hwang, H., Kim, M. & Kim, K. Ternary solvent for $CH_3NH_3PbI_3$ perovskite films with uniform domain size. *Phys. Chem. Chem. Phys.* **19**, 1143-1150 (2017).

[10] Rocks, C., Svrcek, V., Maguire, P. & Mariotti, D. Understanding surface chemistry during MAPbI 3 spray deposition and its effect on photovoltaic performance. *J. Mater. Chem. C* **5**, 902-916 (2017).

[11] Nipane, A., Choi, M. S., Sebastian, P. J., Yao, K., Borah, A., Deshmukh, P., ... & Teherani, J. T. Damage-free atomic layer etch of WSe_2: A platform for fabricating clean two-dimensional devices. *ACS Appl. Mater. & Interfaces* **13**, 1930-1942 (2021).

Changes Made:

We have addressed the challenges for large-scale manufacturing in the Paragraph 2, Page 8 of revised manuscript as follows (highlighted in red): “It is noteworthy that critical challenges of ensuring the long-term stability of $MAPbI_3$ perovskite, maintaining uniformity in film qualities, and preventing defects during the film transfer process must be addressed to achieve large-scale productions”.

==

Comment 2-2: Can the authors provide insights into the long-term stability and operational durability of the heterotransistors under varying environmental conditions?

Response:

Thank you for your important question regarding the long-term stability and operational durability of the heterotransistors under varying environmental conditions. In our study, the long-term stability of the heterotransistors is primarily influenced by the perovskite layer, since Bi_2O_2Se is an ambient stable oxide semiconductor. The stability of perovskite-based devices is often affected by environmental factors such as oxygen and moisture in the air.

We greatly appreciate your insightful concerns on the long-term stability and operational durability of the heterotransistors. In our devices, the stability is predominantly determined by the perovskite layer, since Bi_2O_2Se is as an ambient-stable oxide semiconductor. The instability of halide perovskite is attributed to the intrinsic organic components,[1] ion migrations,[2] and phase segregations under illumination.[3] In our device preparation process, we spin coat a layer of polymethyl methacrylate (PMMA) on the $MAPbI_3$ film surface to protect the device from atmospheric moisture. As suggested by the Reviewer, we have tested the stability of $MAPbI_3/Bi_2O_2Se$ heterotransistor array as follows. We fabricated a 3×3 $MAPbI_3/Bi_2O_2Se$ heterotransistor array and subjected them to the stability testing. At the very beginning, all the nine devices in the array exhibited clear bipolar photoresponse and demonstrated good uniformity (**Fig. R16**). Subsequently, the array was stored in an ambient environment with a temperature of 27±1 ℃ and a relatively humidity of about 55±5 %. After about 168 hours (i.e., one week), we tested the array again. As shown in **Fig. R17**, all devices exhibited a positive photoresponse under negative gate voltage and a negative response when gate voltage was positive, maintaining 100 % bipolar behavior.

Further analysis was directed towards extracting the photoresponsivity as a function of gate voltage. Initial assessments highlighted that the photoresponsivity of all devices was linearly modulated by gate voltage (**Fig. R18**). Upon reexamination after 168 hours in ambient conditions, slight deviations in the symmetry of the bipolar photoresponse was observed (**Fig. R19**). This emerging asymmetry is expected to stem from band drift, potentially induced by alterations in the compositional integrity of perovskites. Moreover, this asymmetry became progressively pronounced after approximately 300 hours (i.e., nearly two weeks) in ambient air, culminating in a complete degradation of performance within a timeframe of 3-4 weeks, thus underscoring the critical need for enhanced protective strategies in the development of stable perovskite-based devices.

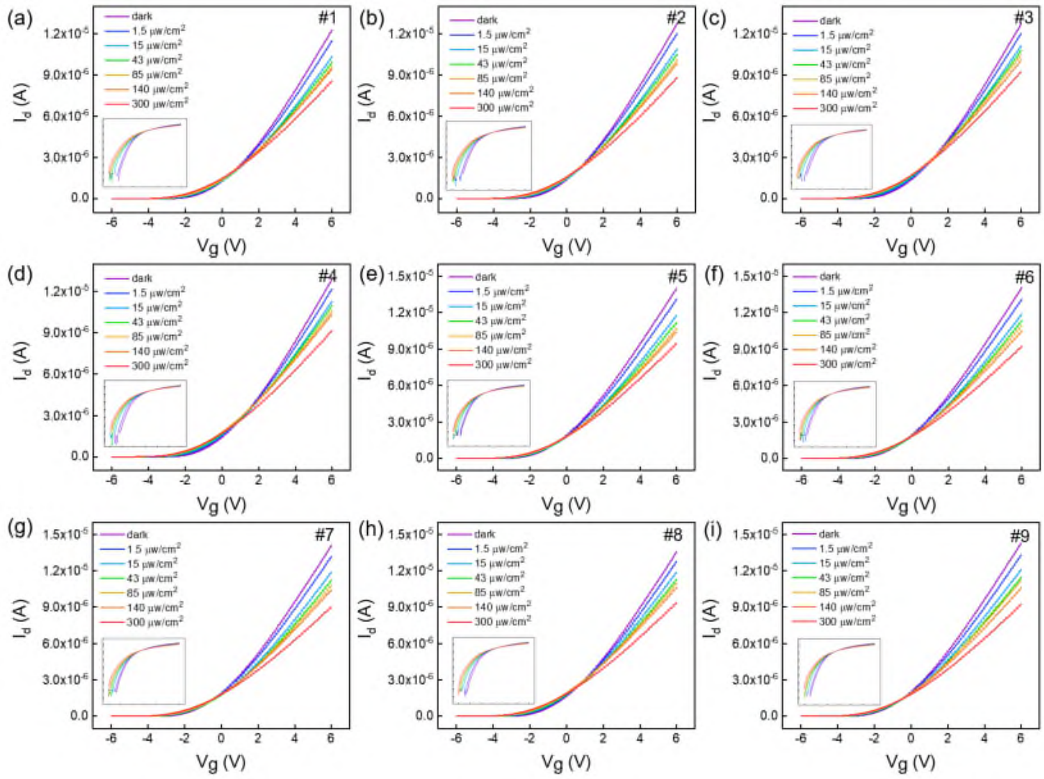


Figure R16. Transfer curves of $MAPbI_3/Bi_2O_2Se$ heterotransistor array under the light illumination of 640 nm with varied light intensities tested after manufacturing.

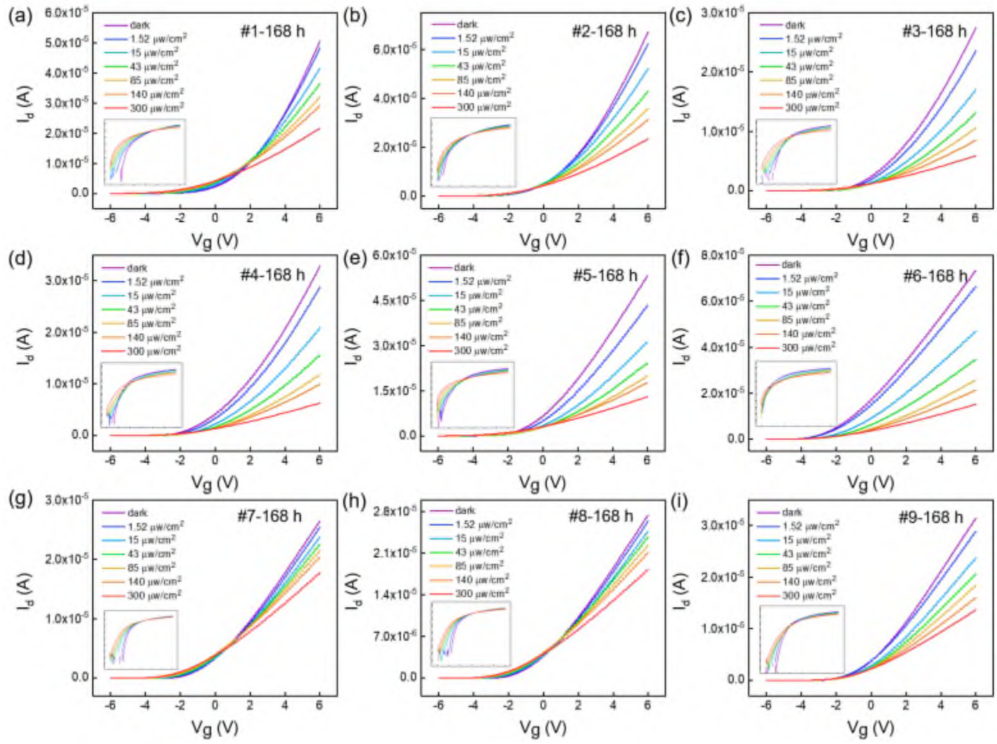


Figure R17. Transfer curves of $MAPbI_3/Bi_2O_2Se$ heterotransistor array under the light illumination of 640 nm with varied light intensities tested after one week at atmosphere.

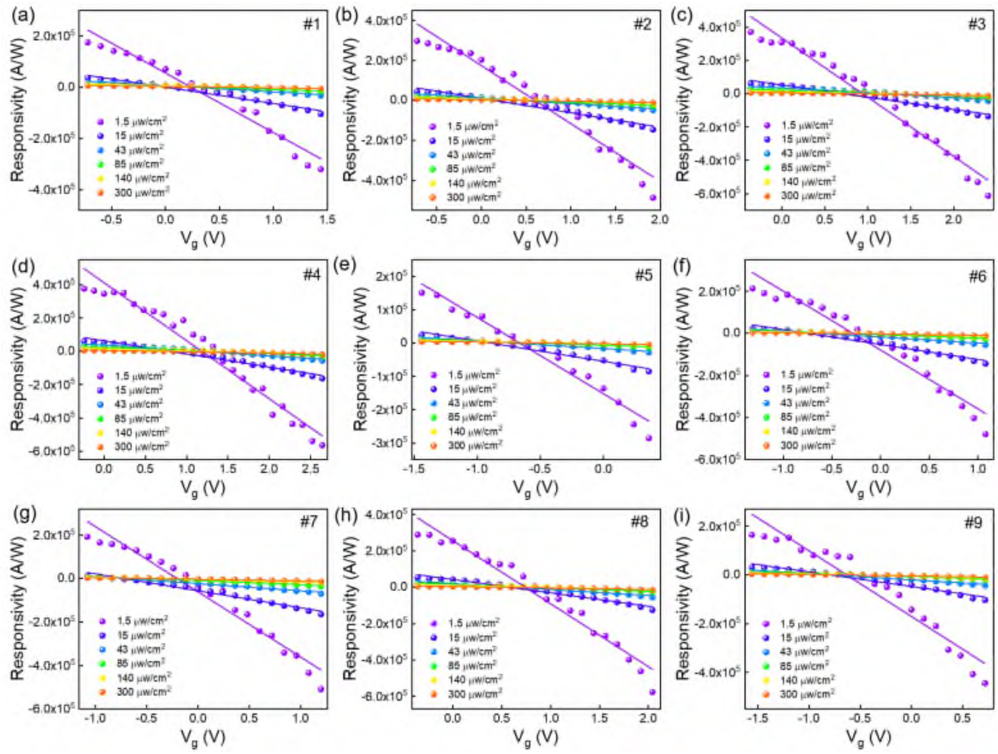


Figure R18. Linear dependence of the bipolar photoresponsivity on the gate voltage under the light illumination of 640 nm with varied light intensities tested after manufacturing.

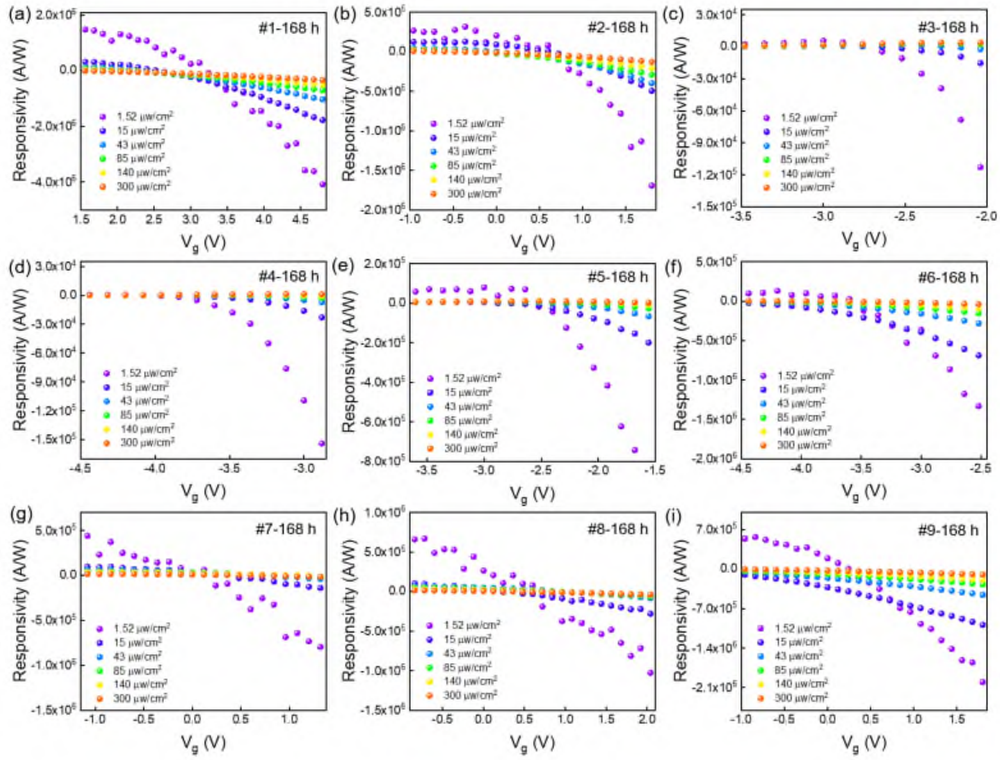


Figure R19. Linear dependence of the bipolar photoresponsivity on the gate voltage under the light illumination of 640 nm with varied light intensities tested after one week at atmosphere.

Corresponding Reference:

[1] Macpherson, S., Doherty, T.A.S., Winchester, A.J.et al. Local nanoscale phase impurities are degradation sites in halide perovskites. *Nature* **607**, 294–300 (2022).

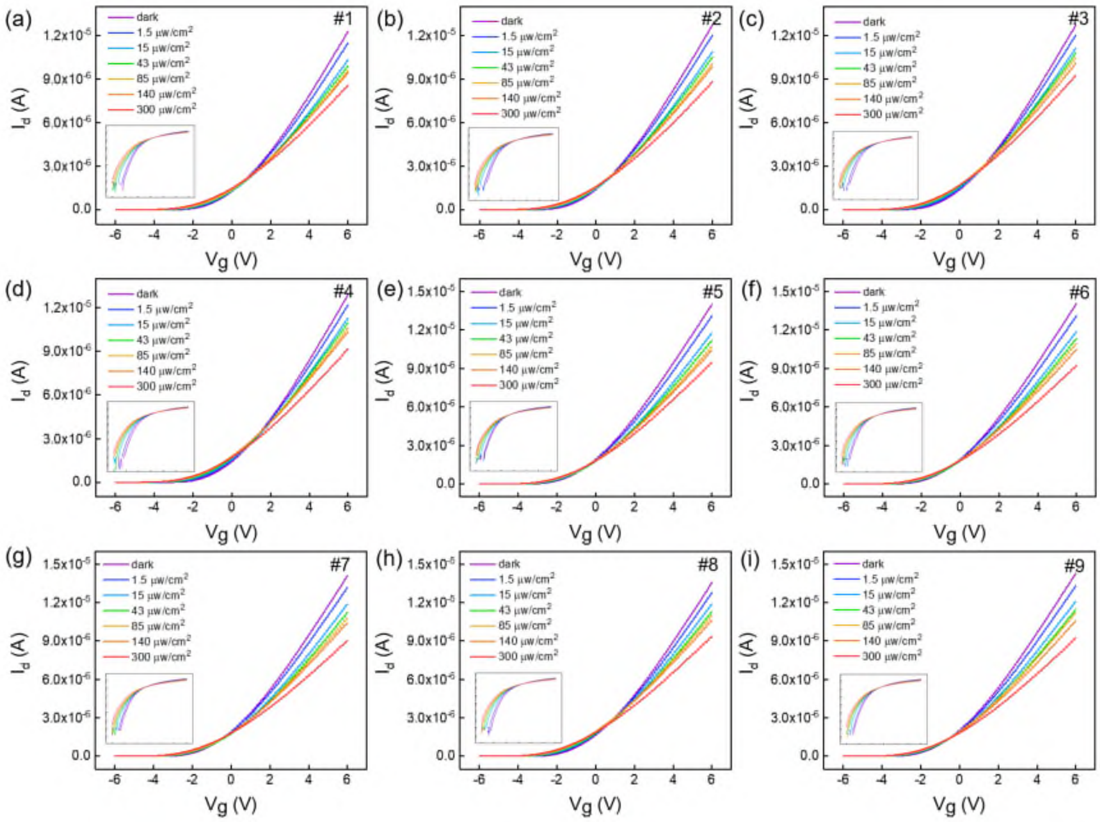
[2] Bertoluzzi L, Boyd C C, Rolston N, et al. Mobile ion concentration measurement and open-access band diagram simulation platform for halide perovskite solar cells. *Joule* **4**, 109-127, (2020).

[3] Lehmann F, Franz A, Többens D M, et al. The phase diagram of a mixed halide (Br, I) hybrid perovskite obtained by synchrotron X-ray diffraction. *RSC Adv.* **9**, 11151-11159, (2019).

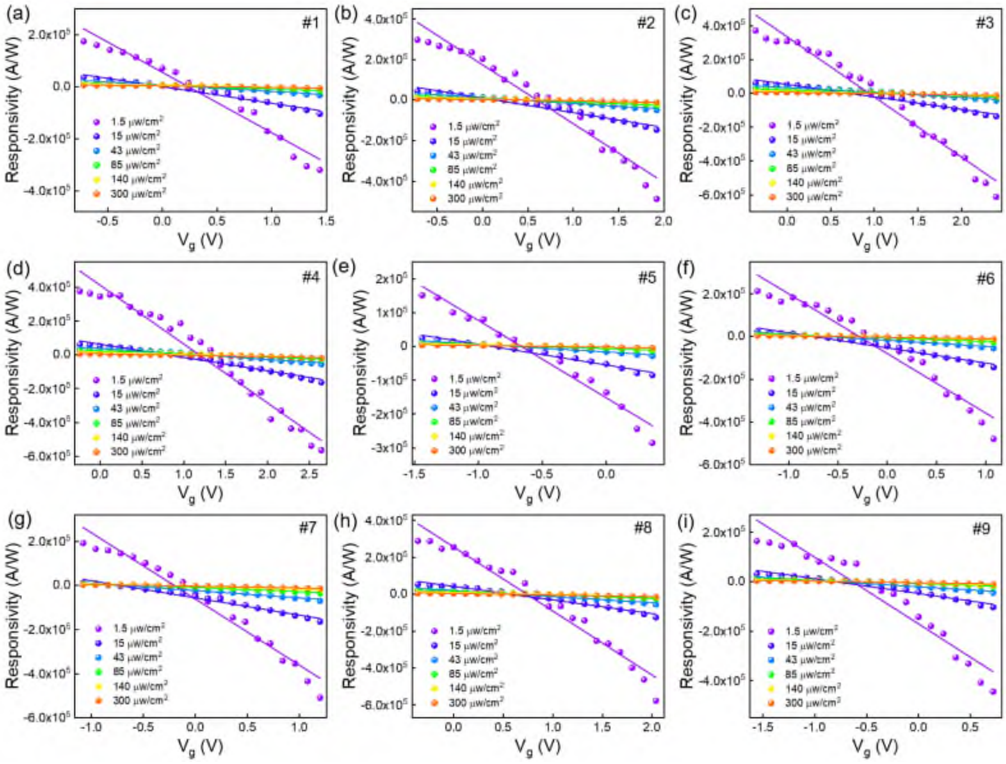
Changes Made:

In the main text (Paragraph 2, Page 8), we have deleted the sentence "The heterotransistors exhibited good stability and reproducibility in modulating bipolar photoresponses as shown in **Fig. 3i** and **Supplementary Fig. S7**." and updated as following (highlighted in red): "We also subjected them to the stability testing. At the very beginning, all the nine devices in the array exhibited clear bipolar photoresponse and demonstrated good uniformity (**Supplementary Fig. S12-S13**). After storing in an ambient environment (i.e., 27±1 °C and 55±5 % humidity) for 168 hours (i.e., one week), all devices maintained mostly the bipolar behavior (**Supplementary Fig. S14-S15**). In addition, the device photoresponse to continuous light pulses were tested under different gate voltages, confirming their excellent reproducibility (**Fig. 3i and Supplementary Fig. S16**)".

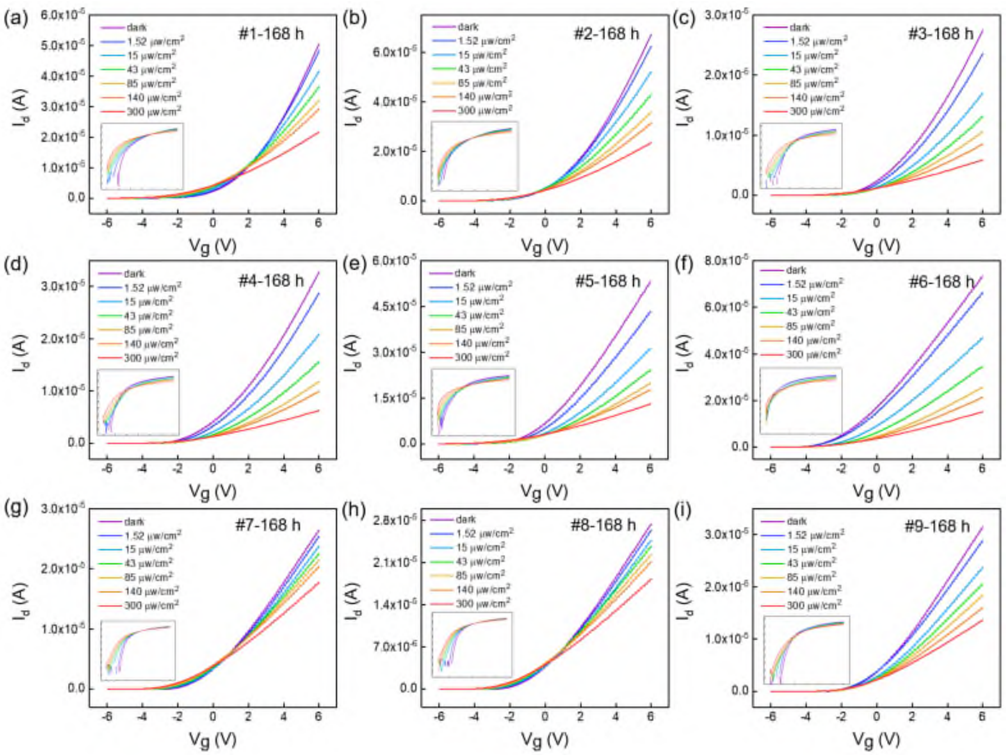
We have included the stability results in **Supplementary Fig. S12-S15**.



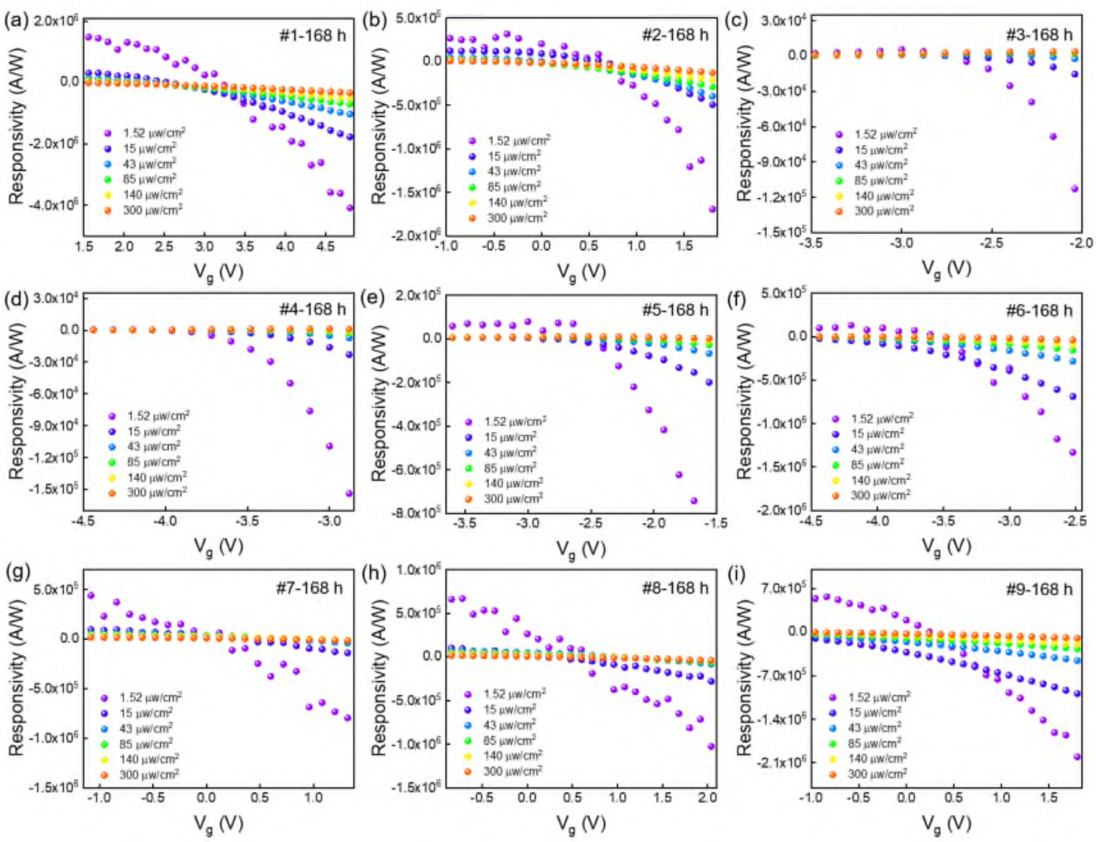
Supplementary Figure S12 | Transfer curves of $MAPbI_3/Bi_2O_2Se$ heterotransistor array under the light illumination of 640 nm with varied light intensities tested after manufacturing.



Supplementary Figure S13 | Linear dependence of the bipolar photoresponsivity on the gate voltage under the light illumination of 640 nm with varied light intensities tested after manufacturing.



Supplementary Figure S14 | Transfer curves of $MAPbI_3/Bi_2O_2Se$ heterotransistor array under the light illumination of 640 nm with varied light intensities tested after one week at atmosphere.



Supplementary Figure S15 | Linear dependence of the bipolar photoresponsivity on the gate voltage under the light illumination of 640 nm with varied light intensities tested after one week at atmosphere.

==

Comment 2-3: Is there potential for integrating this sensor technology with other types of neural networks or electronic components to further enhance its capabilities?

Response:

We sincerely thank the Reviewer for the valuable question regarding the potential extensions of our work. This inquiry has propelled us into a more profound exploration of the practical applications of our sensor technology. In light of your question, we are inspired to delve deeper into how our sensor technology can be seamlessly integrated with other neural networks and electronic components, an aspect we will elaborate on further.

Our heterotransistor sensor, distinguished by its high responsiveness and in-sensor computing capabilities, presents a promising avenue for creating systems that boast improved performance and functionality when integrated with neural networks or various electronic components. Regarding neural networks, by embedding the kernel weights from a convolutional neural network (CNN) initial convolutional layer, or the first-layer weights from a multilayer perceptron (MLP) into our sensor array, our system is poised to execute data collection and preprocessing directly at the edge. Such an integration streamlines neural network operations by lightening the computational burden and accelerating processing speeds. This is particularly beneficial for real-time processing applications, including autonomous vehicles, medical diagnostics, and industrial automation, where efficiency is paramount.

In the context of integration with electronic components, the compatibility of our device with other devices such as microcontrollers and field-programmable gate arrays (FPGAs) opens up avenues for creating highly versatile systems. Microcontrollers can be tasked with managing sensor operations, while FPGAs offer the flexibility to be programmed for executing complex algorithms. This synergistic combination facilitates the development of customized solutions that cater to specific needs, significantly enhancing computational efficiency and expanding functionality.

Changes Made:

We have discussed the potential for integrating this sensor technology with various types of neural networks and electronic components in the discussion section of our manuscript (Paragraph 2, Page 13) as follows (highlighted in red): "Distinguished by its high responsiveness and in-sensor computing capabilities, the $MAPbI_3/Bi_2O_2Se$ heterotransistor presents a promising avenue for creating systems that boast improved performance and functionality when integrated with neural networks or various electronic components".

===

Comments from Reviewer #3

Overall Remarks: This work discusses $MAPbI_3/Bi_2O_2Se$ phototransistors with neuromorphic vision sensing capabilities. The device demonstrated high, gate-tunable photoresponsivity, which is pivotal in developing computing demonstrations. The discussion on the device mechanism is clearly presented, and the work showcases impressive neuromorphic computing applications that will appeal to the broad readership of Nature Communications. My comments below aim to enhance this paper from a neuromorphic perspective.

Response:

We are deeply grateful for your affirmative evaluation and the constructive insights you have provided. Thank you for acknowledging the feature of ultrahigh and gate-tunable photoresponsivity in our devices, which are crucial for dim-light neuromorphic vision computing.

Your recommendations that aim to enrich our paper from a neuromorphic viewpoint are highly esteemed. In response to your invaluable comments, we commit to incorporate your suggestions into our revised manuscript. We believe that your help will markedly elevate the manuscript quality.

===

Comment 3-1: The linear dependency of photoresponses on gate voltages and P_{in} (Fig. 3c, 3g, 3h) facilitates the employment of device characteristics into neuromorphic computing applications. Can the authors discuss more about the device mechanism and how it leads to this linear dependency?

Response:

We appreciate your insightful concerns on the mechanisms of linear dependency. Precisely, the photoresponsivity exhibited a linear relation to the gate voltage (**Fig. 3g**), while the photoresponsivity and photocurrent exhibited the sublinear relation to light intensity when plotted on a logarithmic scale (**Fig. 3c**, **h**). This behavior is characteristic of transistor-type photodetectors, which are known for their considerable photoconductive gain, resulting in a sublinear photoresponsivity to varying incident light intensities.[1] Specifically, in the case of heterotransistors, the generation of photocurrent is predominantly governed by the photovoltaic effect. In the photovoltaic mode, particularly for *n*-channel devices, light illumination induces a shift in the threshold voltage (V_{th}) towards more negative values (**Fig. 3a**). The photocurrent can be described by the following equation:[2]

$$I_{ph} = g_m \Delta V_{th} = \frac{AKT}{q} \ln\left(1 + \frac{\eta q \lambda P_{in}}{I_{pd} hc}\right)$$

where g_m is the transconductance, ΔV_{th} is the threshold voltage shift, A is a proportionality parameter, η is the quantum efficiency associated with the absorption process, P_{in} is the incident light intensity, hc/λ is photon energy, and I_{pd} is dark current for holes. This logarithmic response to light intensity by our device is designed to prevent the saturation of photogenerated charges under high optical power densities, thereby providing a broader dynamic range akin to that of the human eye. The dynamic range of our phototransistor extends up to 110 dB, exhibiting its capability to effectively perceive both brightly lit and dimly illuminated environments. This feature significantly bolsters the potential of our device for use in neuromorphic sensing-computing applications, enabling it to mimic human visual perception with remarkable efficiency.

The second aspect is the linear dependence of photoresponsivity on gate voltage. Photoresponsivity serves as a weight for neural network trainings, making this linear adjustment crucial for enhancing the accuracy. The photoresponsivity R is calculated by $R = I_{ph}/P_{in}$, where I_{ph} is the photocurrent and P_{in} is the incident light intensity per area. The generation of photocurrent is intricately linked to several factors: the quantity of photocarriers, their separation efficiency, and their collection efficiency. The quantity of photocarriers is directly influenced by the incident light intensity and the light absorption characteristics of the perovskite layer. With a constant light intensity, the photocarrier intensity remains unchanged. The separation efficiency of these photocarriers is affected by the heterojunction interface barrier, whereas their collection efficiency is mediated by the mobility of Bi_2O_2Se and interface characteristics, both modulated by the applied gate voltage.

The gate voltage distinctly influences the separation and collection efficiencies of photocarriers. Across the entire gate voltage spectrum (-6 V to 6 V), the relationship between gate voltage and these efficiencies is nonlinear. However, When the gate voltage is near the threshold voltage of the positive and negative photoresponse transition, the electron and hole potential barriers at the interface are equivalent (**Fig. 3j**). In the vicinity of the transition voltage for positive and negative photoresponses, the gate voltage has a negligible effect on interface barrier and mobility, rendering the changes in photocarrier separation and collection efficiencies gradual. This nuanced behavior enables a linear modulation of photoresponsivity with gate voltage within this narrow range, facilitating precise control over device performance in neural network applications.

Corresponding Reference:

[1] Pierre, A., Gaikwad, A. & Arias, A. Charge-integrating organic heterojunction phototransistors for wide-dynamic-range image sensors. *Nat. Photonics* **11**, 193–199 (2017).

[2] Y. Takanashi, K. Takahata and Y. Muramoto. Characteristics of InAlAs/InGaAs high electron mobility transistors under 1.3-μm laser illumination. *IEEE Electron Device Lett.* **19**, 472-474 (1998).

Changes Made:

In the main text on Paragraph 2, Page 7, we have updated the description as (highlighted in red): "More impressively, the heterotransistors consistently delivered high *R* values over a broad light-intensity range from 0.1 μW/cm^2 to 30 mW/cm^2 following a sublinear relationship on the logarithmic scale".

In the main text on Paragraph 2, Page 8, we have removed the sentence: "For a constant gate voltage, the bipolar photocurrents also presented a linear correlation with the incident light intensity as revealed in **Fig. 3h** and **Supplementary Fig. S6**, enabling the feasibility of photocurrents correlated with pixel data to concisely represent captured image information.[39]" And we have updated this part as (highlighted in red): "The photocurrent of our transistor-type photodetectors exhibited sublinear dependences on the light intensity on the logarithmic scale with a dynamic range of up to 110 dB (**Fig. 3h** and **Supplementary Fig. S8**). This logarithmic sublinear photoresponse prevents the saturation under high light intensities, similar to the human eye, allowing the device to effectively perceive both bright and dark images.[39]"

Accordingly, we have replaced Reference 39 by "Pierre, A., Gaikwad, A. & Arias, A. C. Charge-integrating organic heterojunction phototransistors for wide-dynamic-range image sensors. *Nat. Photonics.* 11, 193-199, (2017)."

==

Comment 3-2: In Fig. 3a-3c, R monotonically decreases as P_{in} increases at V_g = 0V. This monotonical trend, however, does not hold in Fig. 3g (640 nm light) and Fig. S5 (532 nm and 780 nm light), irrespective of whether V_g is negative, zero or positive. There is no linear relation between R and P_{in} from Fig. 3g and Fig. S5. Can the authors explain this discrepancy?

Response:

We are sorry for not clearly addressing this**.** We would like to clarify that the photoresponsivity (R) indeed shows a monotonic decrease with increasing light intensity across **Fig. 3a-c**, **Fig. 3g**, and **Supplementary Fig. S7** (i.e., the serial number of Fig. S5 changes to Fig. S7 in the revision). In Fig. 3g and Fig. S5, the curves represent the bipolar photoresponsivity as a function of gate voltage under different light intensities of 640 nm, 532 nm, and 780 nm. Taking **Fig. 3g** as an example, the purple curve corresponds to an optical power density of 2.5 μW/cm², while the red curve corresponds to 2.6 mW/cm². It can be observed that for a fixed gate voltage, the responsivity decreases as light intensity increases, transitioning from purple to red. This trend is consistent for all tested gate voltages as displayed in **Fig. R20**.

Precisely, in our response to Comment 3-1, we elucidated that the photoresponsivity of our device demonstrated a linear relationship with the gate voltage (**Fig. 3g**), while the photoresponsivity and photocurrent exhibited the sublinear relation to light intensity when plotted on a logarithmic scale (**Fig. 3c**, **h**). This behavior is characteristic of transistor-type photodetectors, which are known for their considerable photoconductive gain, resulting in a sublinear photoresponsivity to varying incident light intensities. This sublinear relationship is significant because the logarithmic sensing of light intensity by our device prevents the saturation of photogenerated charges under high optical power densities and provides a wider dynamic range, as the dynamic range of the phototransistor can reach up to 110 dB, allowing it to effectively perceive both bright and dark images.

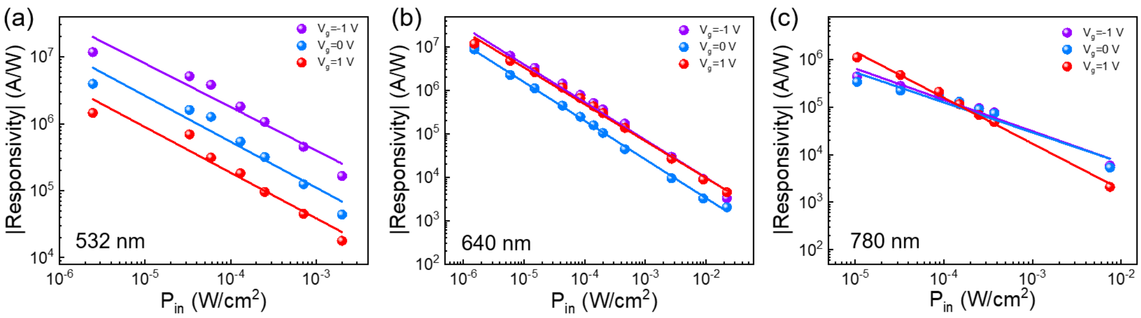
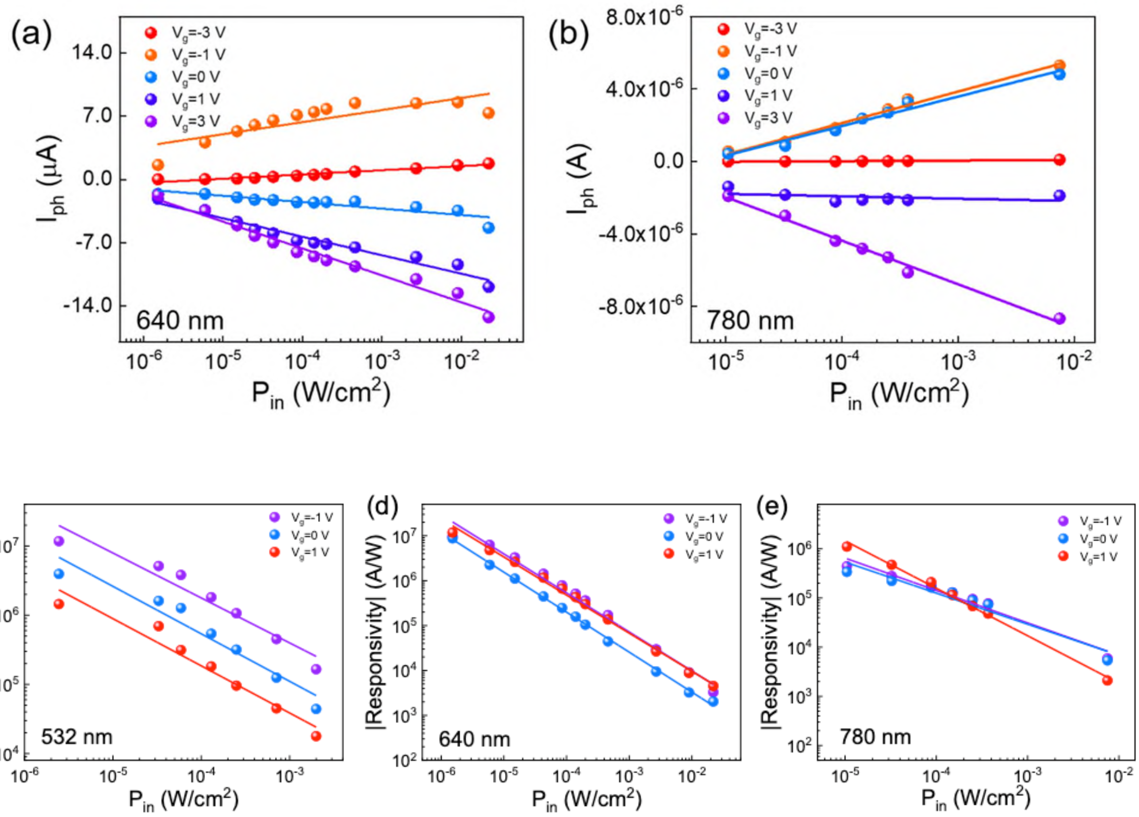


Figure R20. Dependence of the photoresponsivity on the light intensity at (a) 532 nm, (b) 640 nm, and (c) 780 nm, measured at gate voltage of -1 V, 0 V and 1 V.

Changes Made:

We have included **Fig. R20** to the supplementary information as **Supplementary Fig. S8c-e** to clearly display the sublinear relationships.



Supplementary Figure S8 | Dependence of the net photocurrent on the light intensity under the light illumination of **(a)** 640nm, and **(b)** 780 nm with the gate voltage varying from -3 V to 3 V. Dependence of the photoresponsivity on the light intensity at **(c)** 532 nm, **(d)** 640 nm, and **(e)** 780 nm, measured at gate voltage of -1 V, 0 V and 1 V.

==

Comment 3-3: For easier comparison, I suggest that the authors plot all figures at one specific wavelength (such as 532 nm) throughout Fig. 3 and then present other similar series of plots at wavelengths of 400nm, 640nm, and 780nm in the Supporting Information.

Response:

Thank you for your constructive suggestion. In response, we have standardized the representation of all figures at the specific wavelength of 532 nm in **Fig. 3** to facilitate a more straightforward comparison. Additionally, we have included a similar series of plots at wavelengths of 640 nm and 780 nm within the Supporting Information. We believe these modifications will significantly enhance the clarity and coherence of our data presentation.

==

Comment 3-4: The significance of how the momentum conservation between $MAPbI_3$ and Bi_2O_2Se contributes to this device/this work should be more clearly established. The Fig. 3f shows previous work using $MAPbI_3/MoS_2$ can perform dim-light sensing (similar light power) with photoresponsivity above 10^5 A/W. Assuming it is the state-of-art performance, this work should compare how its higher photoresponsivity (10^7 A/W) impacts the CNN applications compared to a potoresponsivity of 10^5 A/W.

Response:

Thank you for your insightful inquiries regarding the significance of momentum conservation between $MAPbI_3$ and Bi_2O_2Se, as well as the comparison with the state-of-the-art performance of $MAPbI_3/MoS_2$. First, we want to clarify that while the $MAPbI_3/MoS_2$ photodetector (*Adv. Mater.*, **2017**, *29*, 1603995) exhibited an impressive photoresponsivity of 10^5 A/W, it was not able for in-sensor computing, due to its lack of bipolar photoresponse feature. As we discussed in the Introduction part, both positive and negative photoelectrical responses have been manifested in 2D van der Waals heterostructure phototransistors, however, these 2D bipolar phototransistors often suffer from suboptimal optoelectronic performances in terms of narrow wavelength range, low photoresponsivity, and small signal-to-noise ratio. The significance of our work is the simultaneous achievement of high bipolar photoresponses, which not only parallels but surpasses the capabilities of traditional single-functional photodetectors.

Our $MAPbI_3/Bi_2O_2Se$ heterojunction exhibits tunable interfacial band alignments and energy-level momentum conservations, enabling not only enhanced photodetections but also bipolar photoresponses. Density Functional Theory (DFT) calculations of Bi_2O_2Se and $MAPbI_3$ reveal that the Valence Band Maximum (VBM) of $MAPbI_3$ and the Conduction Band Minimum (CBM) of both materials align at the Γ point (**Fig. 1b-c)**, indicating the presence of energy-band momentum matching. This theoretical alignment was further experimentally verified through the temperature-dependent steady-state photoluminescence (PL) spectra as shown in **Fig. R21**. At low temperatures, the suppression of non-radiative recombination processes accentuates the PL signals emanating from the heterojunction. We detected peaks indicative of the indirect bandgap of Bi_2O_2Se at 1270 nm, as well as the direct bandgap of $MAPbI_3$ at 765 nm. More important, the emerging peak at 965 nm was correlated with the interband transfer between the VBM of $MAPBI_3$ and the CBM of Bi_2O_2Se based on the ultraviolet photoelectron spectroscopy (UPS) analyses in **Supplementary Fig. S18** (i.e., ~1.28 eV), affirming the momentum-matched bandgap at the heterojunction interface.

This momentum-matched energy band at the $MAPbI_3/Bi_2O_2Se$ interface is beneficial for enabling direct photocarrier transition and transport. Our comprehensive investigations, encompassing absorption spectra, photoluminescence spectra, and femtosecond spectroscopy, confirm that the momentum-matching between Bi_2O_2Se and $MAPbI_3$ significantly bolsters optical carrier transition and transport. This is achieved by (1) facilitating direct excitation and transition of electrons from $MAPbI_3$ to the Bi_2O_2Se channel (**Fig. 2d**), and (2) promoting the extraction of hot carriers from $MAPbI_3$ to the Bi_2O_2Se channel (**Fig. 2e-h**). The preservation of momentum at the heterojunction markedly enhances the photocarrier transition and transport, thereby significantly improving the photon utilization and the overall photo-sensing performance of the device.

Convolution-kernel processing can perform preprocessing functions for images, such as edge detection, denoising, and formatting. The photoresponsivity significantly affects the quality of image preprocessing. Our high photoresponsivity of $MAPbI_3/Bi_2O_2Se$ heterotransistor enables dim-light in-sensor convolutions, which were not demonstrated in these two papers. **Fig. R22** shows the images captured at Peking University under bright and dim scenes, respectively. The pixel information of images was translated to optical signals, and projected onto the reconfigurable photosensors set as the Laplacian operator for edge detection: (1) V_g was set to -2 V/-0.2 V as the kernel weights correlated with the high bipolar photoresponsivities of 2.4×10^5 A/W and -0.3×10^5 A/W, respectively; and (2) V_g was set to -0.6 V/-0.5 V as the kernel weights correlated with the low bipolar photoresponsivities of 4.6×10^4 A/W and -5.8×10^3 A/W, respectively. The high-responsive sensor detected more detailed edges of the image. Notably, the in-sensor convolution preprocessing of dim-light images has not been reported in other papers.

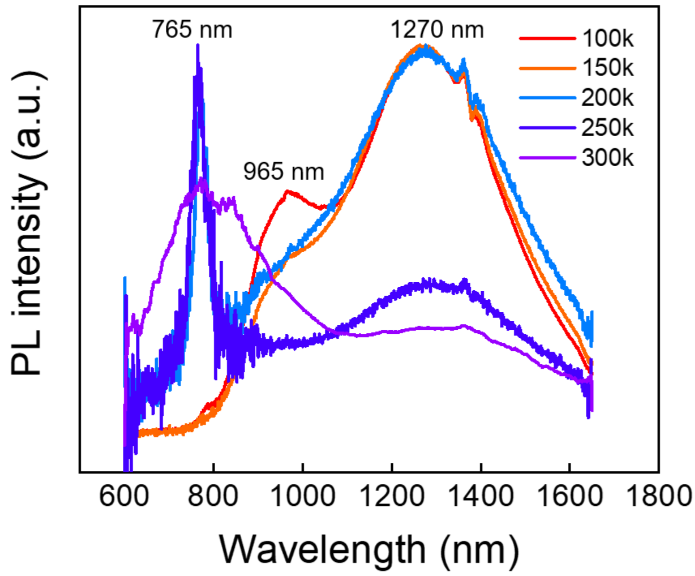


Figure R21. Temperature-dependent steady-state photoluminescence spectra of the $MAPbI_3/Bi_2O_2Se$ heterojunction.

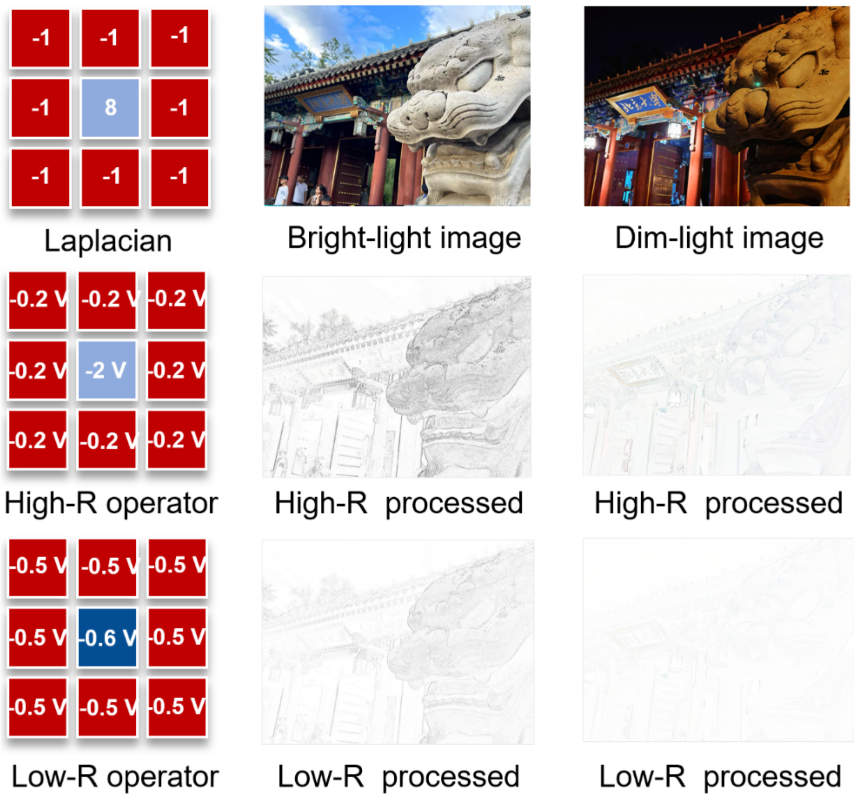


Figure R22. Implementation of dim-light in-sensor edge detection. Edge detection of images under bright and dim lights by the reconfigurable photosensors with high photoresponsivities and low photoresponsivities, respectively.

Changes Made:

In the main text (Paragraph 1, Page 5), we have updated the PL result as the following (highlighted in red): "This theoretical alignment was further experimentally verified through the temperature-dependent steady-state photoluminescence (PL) spectra (Fig. 1d), wherein the emerging peak at 965 nm was correlated with the interband transfer between the VBM of $MAPBI_3$ and the CBM of Bi_2O_2Se (i.e., ~1.28 eV), affirming the momentum-matched bandgap at the heterojunction interface". Correspondingly, we have reorganized the context of Fig. 1, and included **Fig. R21** as **Fig. 1d**.

In the main text (Paragraph 2, Page 12), we have addressed distinguishing aspect of novelty of our heterotransistor for in-sensor data processing in the context (highlighted in red): "To evaluate the effect of varying photoresponsivity levels on image processing, we configured the array into Laplacian convolution kernels by adjusting the gate voltages. As shown in

Supplementary Fig. S23**, the high-responsivity configurations successfully detected edges in both bright and dim images, whereas the low-responsivity configurations led to significant edge information loss in dim images. The ultrasensitive reconfigurable heterotransistor array holds promises for improving dim-light visual perception in CNNs. [2,48]" We have added **Figure R22** in the supplementary information as **Supplementary Fig. S23**, and cited these two papers as **Reference 2** and **48**.

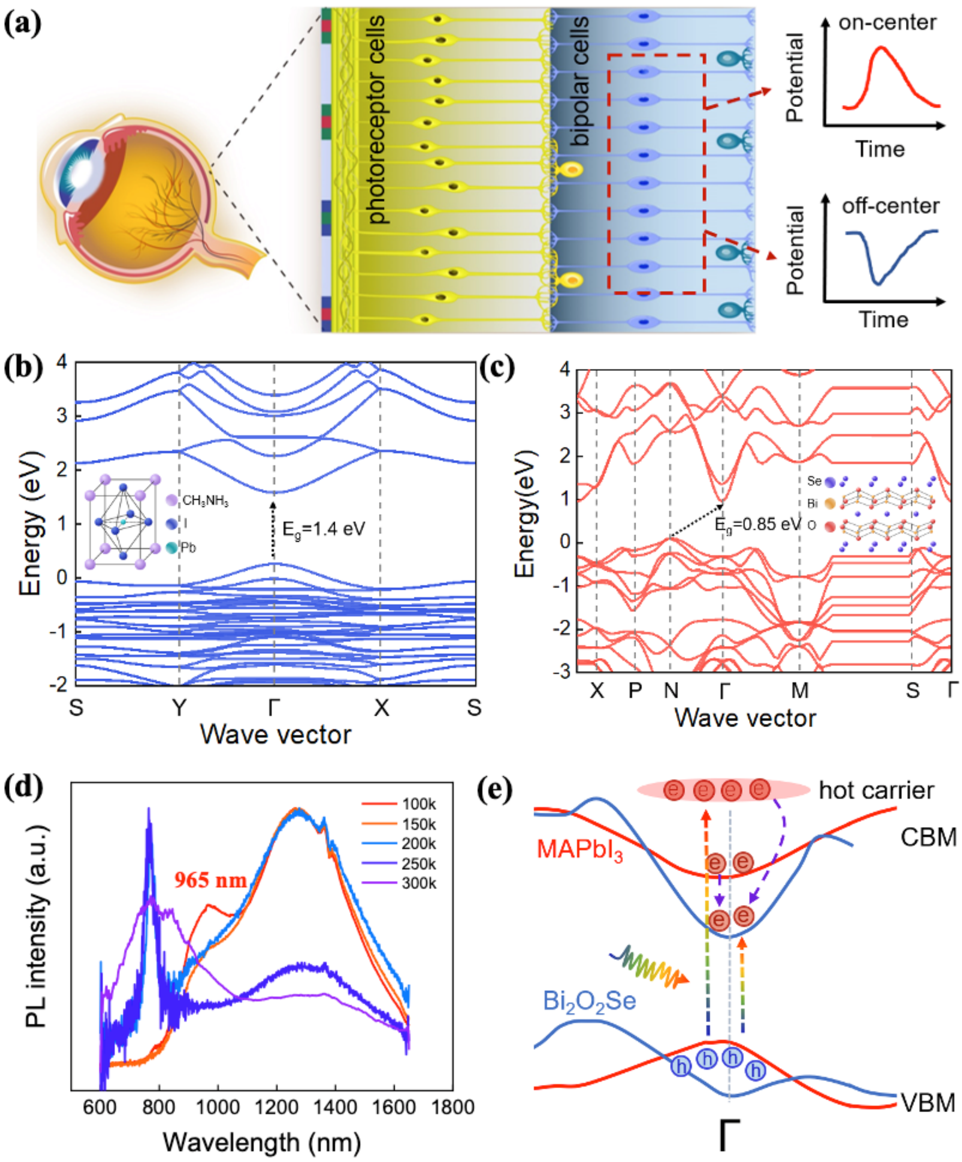
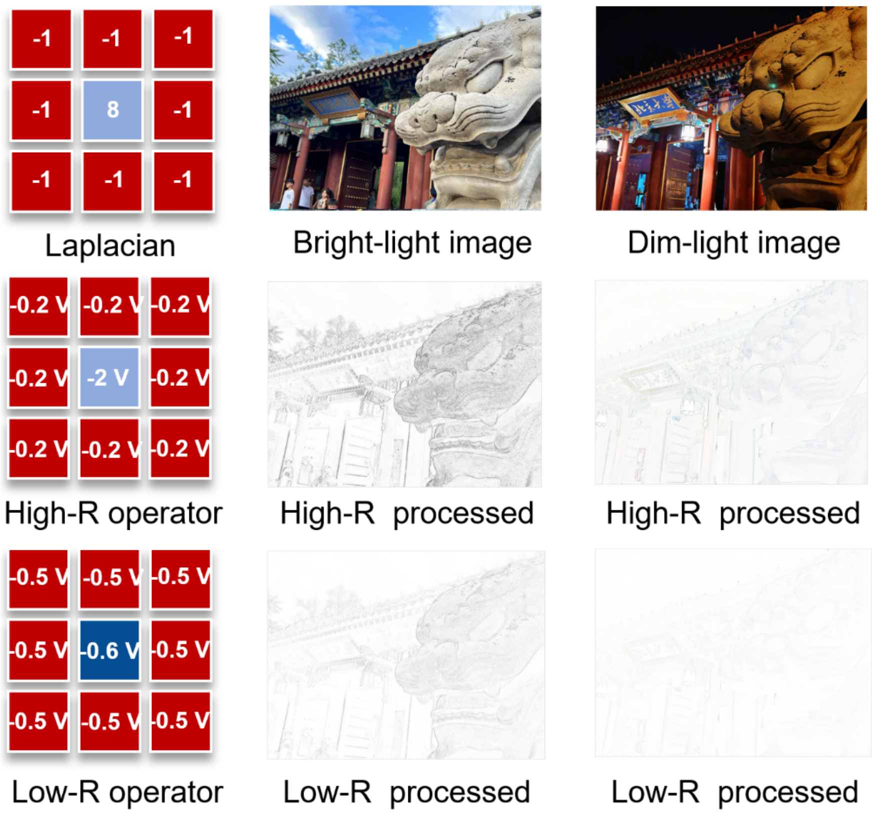


Figure 1. Reconfigurable neuromorphic vision sensing and momentum conservation in $MAPbI_3/Bi_2O_2Se$ heterojunction. (a) Schematic illustration of human retina consisting of photoreceptor and bipolar cells for vision information processing via the on/off features. Characteristic positive and negative photocurrents of bipolar photosensors manipulated by

electric fields. Energy band structure of **(b)** $MAPbI_3$, and **(c)** Bi_2O_2Se calculated by the density functional theory. **(d)** Temperature-dependent steady-state photoluminescence spectra of the $MAPbI_3/Bi_2O_2Se$ heterojunction. **(e)** Diagram of the interband photoexcitation and hot-carrier transition in the momentum-conserved $MAPbI_3/Bi_2O_2Se$ heterojunction.



Supplementary Figure S23 | Edge detection results for bright and dim images using $MAPbI_3/Bi_2O_2Se$ reconfigurable sensor array with high photoresponsivities and low photoresponsivities, respectively.

==

Comment 3-5: How consistent is the device-to-device variation in this type of device, especially since a crossbar array of this device is fabricated for neuromorphic computing application? How are the noise, hysteresis and other properties of the device accounted for in the simulation?

Response:

We appreciate the Reviewer's insightful comments. To demonstrate the device-to-device variation of our heterotransistor array, we characterized both the photoresponsivity and photocurrent of all nine devices within the 3×3 $MAPbI_3/Bi_2O_2Se$ heterotransistor array. **Fig. R24** shows the photoresponsivity of these devices as a function of gate voltage under the illumination of 640 nm, and **Fig. R25** presents the variation of the photocurrent with the light intensity. To delve deeper, we extracted the maximum positive and negative photoresponsivity values from the nine devices (**Fig. R26a**). The data revealed a maximum positive photoresponsivity of $(2.4\pm0.9)\times10^5$ A/W and a maximum negative photoresponsivity of $(-4.8\pm1.1)\times10^5$ A/W, underscoring a commendable degree of uniformity (**Fig. R26b**).

Moreover, we investigated the photocurrent distribution across different gate voltages while maintaining a constant light intensity of 1.52 μW/cm^2 (**Fig. R26c**), revealing an excellent uniformity at varied gate voltages. When we increased the light intensity to 300 μW/cm^2, the variation in photocurrent remained consistent with the patterns observed at the lower light density (**Fig. R26d**). This consistency across different testing conditions attests to the remarkable uniformity of the $MAPbI_3/Bi_2O_2Se$ heterotransistor array, reinforcing the validity and reliability of our results.

We completely agree with the Reviewer's insightful comments. In the course of the simulation process, it is necessary to account for noise, hysteresis, and various other pertinent factors. This comprehensive approach ensures that the simulation outcomes closely mirror the device performance under real-world operating conditions. By integrating these considerations into our simulations, we aim to enhance the reliability and predictive accuracy of our results, thereby providing more convincing representation of the device capabilities in practical applications.

The device-to-device variations represent significant non-ideal factors that must be considered. Particularly in the context of the "P", "K", "U" letter recognition task, we edited the photoresponse-gate voltage (R-V_g) data measured from the nine devices into a lookup table. This enables us to accurately simulate the conductance modulation process of these devices during the training phase, ensuring a precise representation of device behavior under specific conditions. For the tasks involving image convolution processing, this approach is equally beneficial. We utilized the R-V_g data to configure each device within the array to exhibit proportional responsiveness. By applying convolution kernels over the input image, we achieved accurate convolution results, effectively harnessing the unique properties of each device to fulfill the task requirements. This comprehensive methodology underscores our commitment to addressing the challenges posed by device non-idealities, ensuring the fidelity and reliability of our simulation-based studies.

Regarding noise considerations, thermal noise and shot noise stand out as prevalent sources. Thermal noise originates from the stochastic thermal movement of charge carriers, whereas shot noise results from the quantization of charges or photons. Both types of noise typically exhibit as Gaussian white noise, characterized by a mean value of zero. Their temporal fluctuations are quantifiable through their root mean square (*RMS*) values. The *RMS* value for thermal noise can be derived using the formula $I_{th} = \sqrt{4kTB/r}$, where k is the Boltzmann's constant, T is the absolute temperature, B is the noise power bandwidth, and r is the resistance. For example, setting T=300 K and a sampling interval t_s=2 ms yields $B = \frac{1}{2t_s} = 250Hz$. For the device resistance $r = \frac{V_{ds}}{I_{ds}} = \frac{0.1V}{3.2\times10^{-4}A}$ equating to 312.5 Ω, we calculate $I_{th} = 0.115\, nA$ Similarly, the *RMS* value of shot noise can be expressed as $I_{shot} = \sqrt{2qIB}$, where q is the elementary charge, and I is the average readout current. By setting B to 250 Hz and I to 320 μA, I_{shot} computes to 0.505 nA. To simulate the influence of ambient noise on the output current amidst strong interference and to elucidate the impact of different responsiveness on recognition outcomes, we incorporate a white noise component I_{am} with a mean of zero into our computations. The simulated total readout current signal is thus formulated as $I_{total} = I_{out} + I_{th} + I_{shot} + I_{am}$. For simplified representation, these noise contributions are summarized in the simulation as a random Gaussian noise signal with a variance of 0.003.

To address the issue of hysteresis, which typically extends the response time of devices and diminishes their accuracy, we have taken a strategic approach in our simulations. Recognizing that hysteresis leads to different response times across various devices, potentially affecting the accuracy of the summed current readout, we have implemented a specific mitigation strategy. By setting the reading time for the summed current to exceed the maximum observed device response time (e.g., 100 ms), we ensure the stability and reliability of the readout current results. This careful adjustment allows us to accommodate the inherent delays introduced by hysteresis, thereby enhancing the overall accuracy of our simulation outcomes.

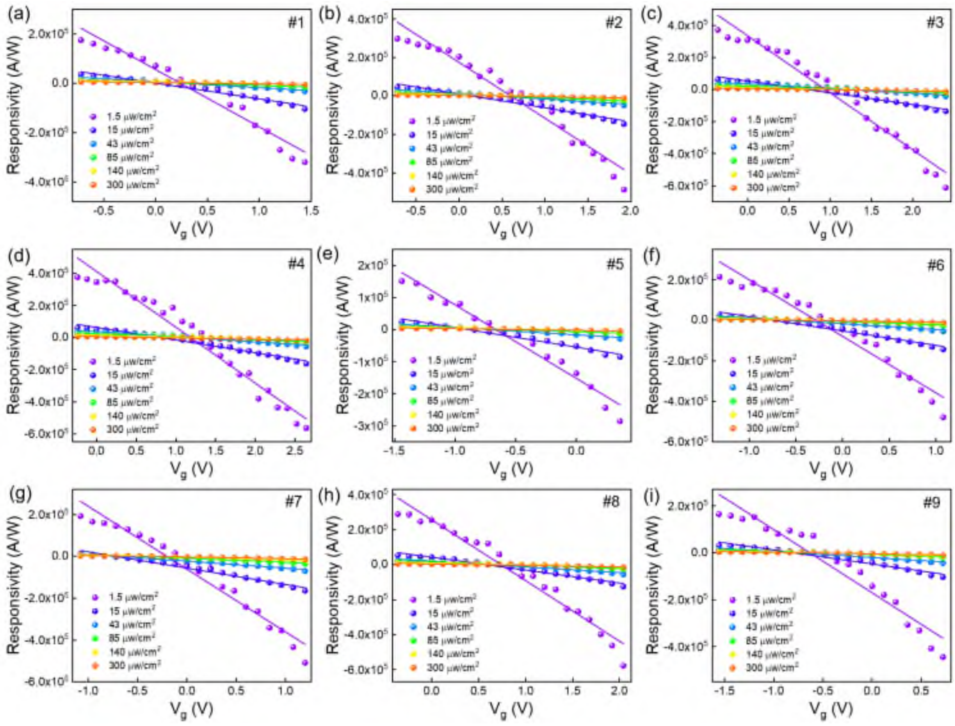


Figure R24. Linear photoresponsivity in response to light intensity across various gate voltages of the 3×3 $MAPbI_3/Bi_2O_2Se$ phototransistor array under the light illumination of 640 nm.

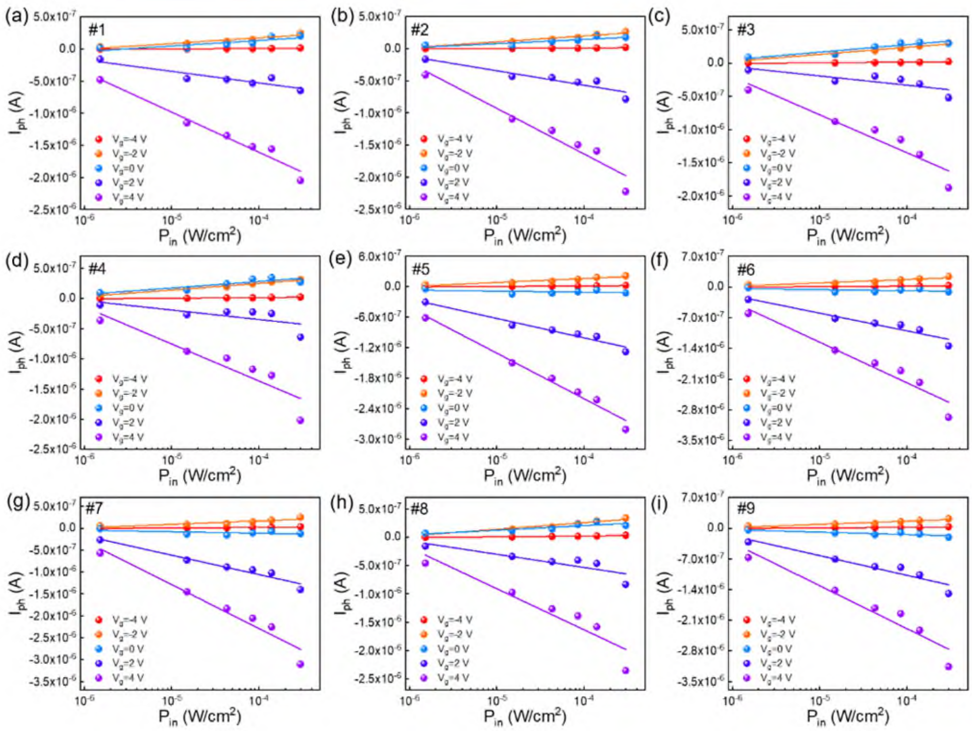


Figure R25. Photocurrent in response to light intensity across various gate voltages of the 3×3 $MAPbI_3/Bi_2O_2Se$ phototransistor array under illumination at 640 nm.

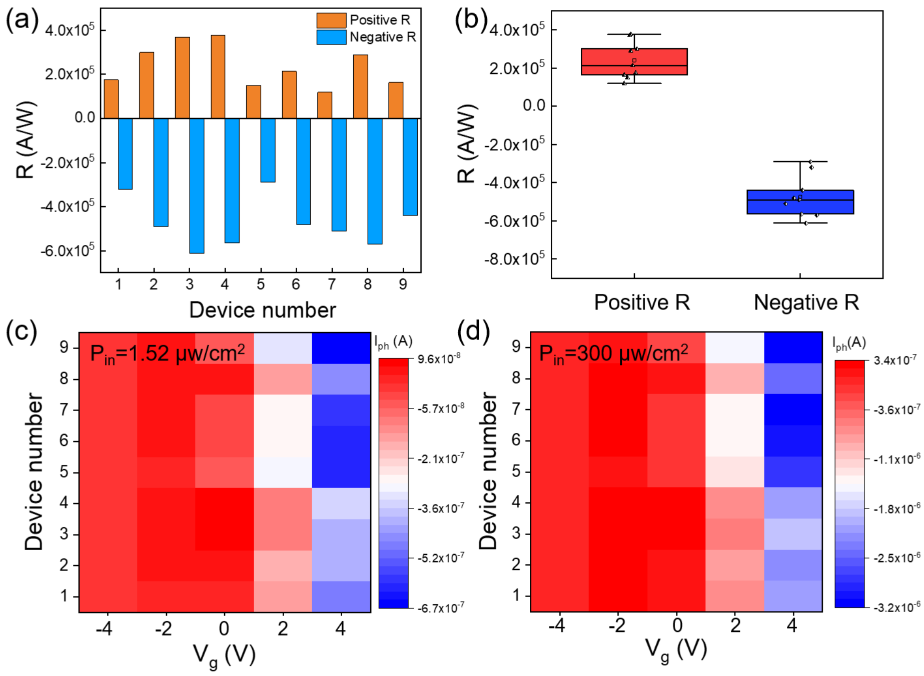


Figure R26. (a) Maximum positive and negative photoresponsivity of the 3×3 $MAPbI_3/Bi_2O_2Se$ phototransistor array. (b) Statistical distribution of maximum positive and negative photoresponsivity across the nine devices. (c) Photocurrent distribution across the nine devices as a function of gate voltage at the light intensity of 1.52 μW/cm². (d) Photocurrent distribution across the nine devices as a function of gate voltage at the light intensity of 300 μW/cm².

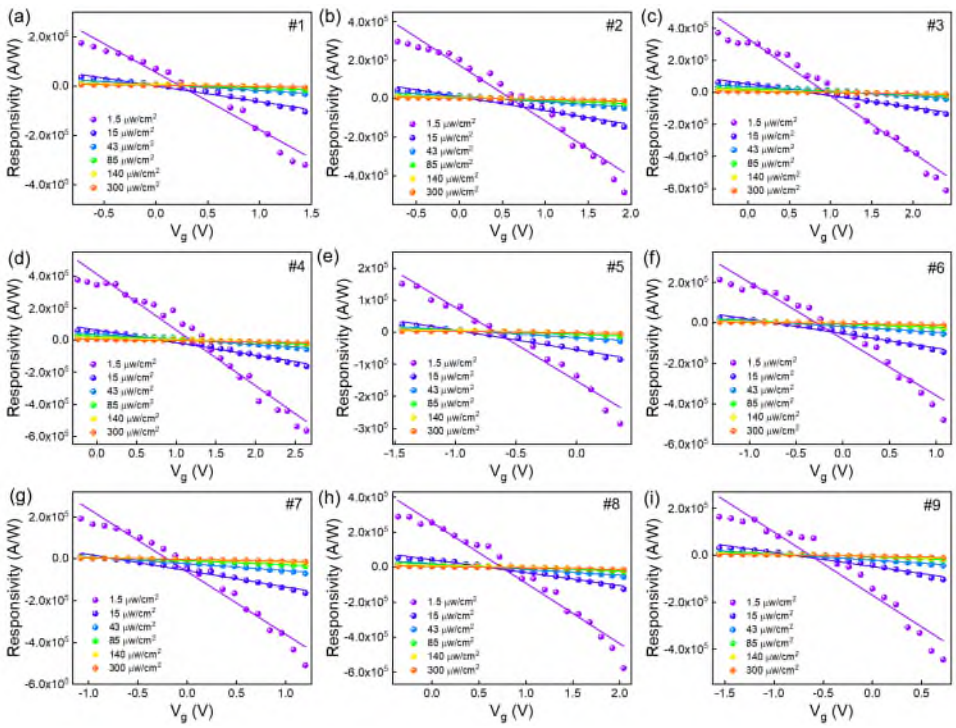
Changes Made:

In the main text (Paragraph 2, Page 8), we have included the following context: "Moreover, we comprehensively analyzed both the photoresponsivity and photocurrent of all nine devices within the 3×3 $MAPbI_3/Bi_2O_2Se$ heterotransistor array under 640 nm illumination (**Supplementary Fig. S9-S10**), revealing that the maximum positive photoresponsivity was consistent in the range of (2.4±0.9) $\times10^5$ A/W and the maximum negative one maintained in the range of (-4.8±1.1) $\times10^5$ A/W (**Supplementary Fig. S11a-b**). In addition, all the nine devices within the heterotransistor array maintained uniform photocurrents at varied gate voltages under both low and high light intensities (**Supplementary Fig. S11c-d**)".

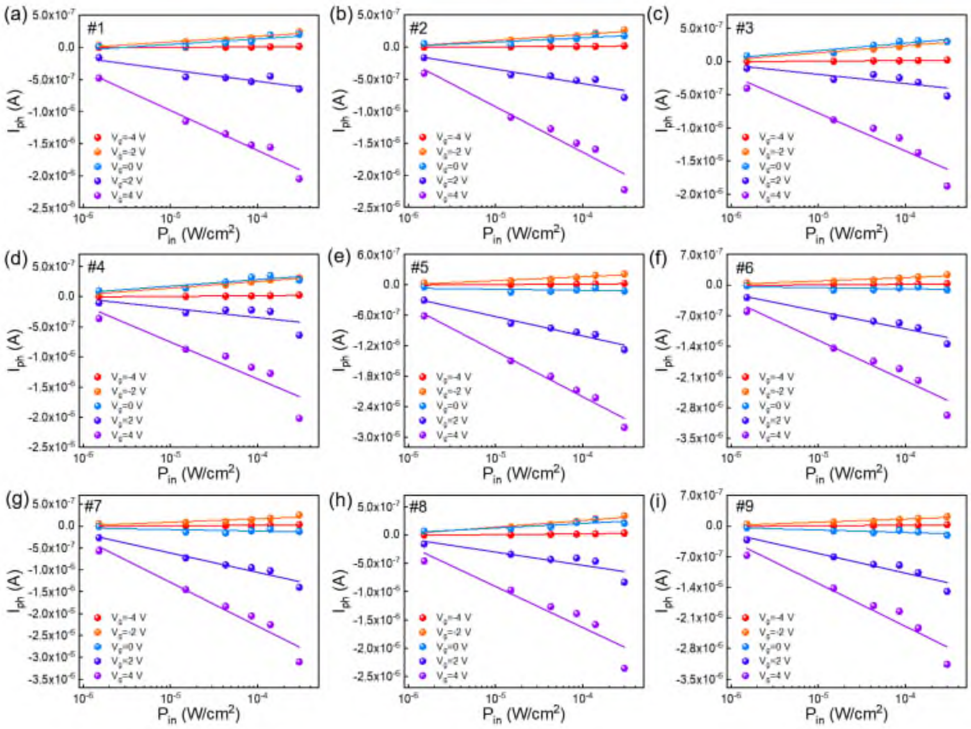
To describe how noise, hysteresis, and device variations are taken into account in the simulation, we have updated the following two notes into the section of "Implementation of

Pattern Classification” in the Methods as follows (highlighted in red): “By integrating impacts of noise and hysteresis into our simulation, and to elucidate the effects of different responsiveness on recognition accuracy, a Gaussian noise with the variance of 0.003 was deployed on the readout photocurrent” and “It is worth noting that, due to the device-to-device variations in the array, which are significant non-ideal factors in neuromorphic computing, we edited the photoresponsivity-gate voltage (R-V_g) data measured from the nine devices into a lookup table. This enables us to accurately simulate the conductance modulation process of these devices”.

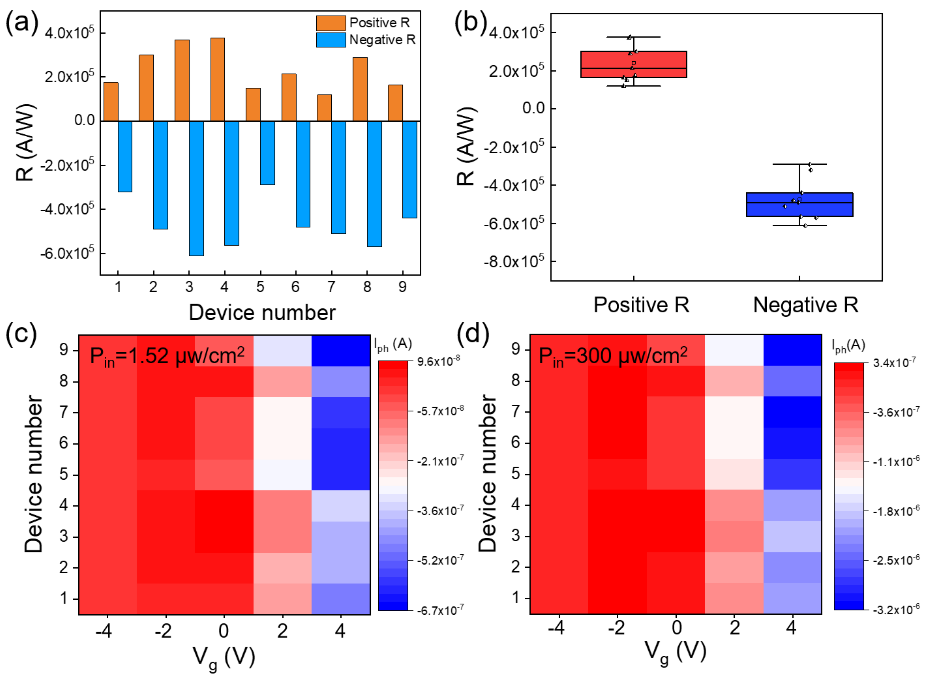
We have included the corresponding **Fig. R24-R26** into the supplementary information as **Supplementary Fig. S9-S11**.



Supplementary Figure S9 | Linear photoresponsivity in response to light intensity across various gate voltages of the 3×3 $MAPbI_3/Bi_2O_2Se$ phototransistor array under the light illumination of 640 nm.



Supplementary Figure S10 | Photocurrent in response to light intensity across various gate voltages of the 3×3 $MAPbI_3/Bi_2O_2Se$ phototransistor array under illumination at 640 nm.



Supplementary Figure S11 | (a) Maximum positive and negative photoresponsivity of the 3×3 $MAPbI_3/Bi_2O_2Se$ phototransistor array. (b) Statistical distribution of maximum positive and negative photoresponsivity across the nine devices. (c) Photocurrent distribution across the nine devices as a function of gate voltage at the light intensity of 1.52 μW/cm². (d) Photocurrent distribution across the nine devices as a function of gate voltage at the light intensity of 300 μW/cm².

===

Comment 3-6: Please explain the algorithm application in more detail for the reader that is not specialized in neuromorphics, perhaps in SI.

Response:

We appreciate the Reviewer's valuable suggestion. To facilitate readers, particularly those not be specialized in neuromorphics to better understand our algorithm, we have enriched our manuscript with more comprehensive explanations of its fundamental components. Our algorithm is delineated into two principal sections: the Implementation of Pattern Classification and the Demonstration of Convolutional Kernel.

(1) Implementation of Pattern Classification

The task of classifying the letters "P", "K", "U" is executed using a Multilayer Perceptron (MLP) model. MLPs are renowned for the efficacy in image classification and recognition. In this task, the architecture encompasses an input layer furnished with 9 neurons $(x_1, x_2, \ldots, x_9)$, mirroring the 9-pixel values of the input letter image, and culminates in an output layer with 3 neurons (y_P, y_K, y_U), each dedicated to one of the target letters. The interconnection between input and output strata is facilitated via a weight matrix W with the output neuron activation a_j determined by $a_j = \sum_{i=1}^{9} W_{ij} \cdot x_i + b_j$, where W_{ij} is the weight from the input neuron i to the output neuron j, and b_j is the bias term for the output neuron j. This configuration enables the precise recognition of letters, as the network discriminates based on the highest activation value among the output neurons.

In a typical hardware realization, this neural construct can be emulated using an array of bipolar sensors, epitomizing in-sensor computing. In this work, the photoresponsivity R_{ij} of $MAPbI_3/Bi_2O_2Se$ heterotransistor array are tailored through the gate voltage $V_{g_{ij}}$ to reflect the synaptic weights W_{ij} of a neural network. Upon exposure to light encoding the input data P_{ij}, each device in the array generates a photocurrent I_{ij} proportional to the product of the input signal and its responsivity: $I_{ij} = R_{ij} \cdot P_{ij}$. The total current output $I_{out} = \sum_{i,j}^{i=3,j=3} I_{ij} =$

$\sum_{i,j}^{i=3,j=3} R_{ij} \cdot P_{ij}$, embodies the multiply-accumulate (MAC) operation within the sensor array, harnessing Kirchhoff's law to execute computation.

For the classification task, a specialized dataset comprising 100 instances of Gaussian-noise afflicted “P”, “K”, and “U” letter patterns (i.e., noise standard deviation of 0.2) is segmented into training and testing sets in a 7:3 ratio. The letters are illuminated onto the sensor array with peak incident light intensity set at 0.1 μW/cm^2. Comparative analysis is conducted across three sensor arrays, differentiated by the photoresponsivity of 10^2 A/W, 10^4 A/W, and 10^6 A/W, respectively. The photoresponsivity of each heterotransistor is all initialized randomly at the beginning. The projection of the letter image induces a readout current $I_{out} = \sum_{i,j}^{i=3,j=3} I_{ij} = \sum_{i,j}^{i=3,j=3} R_{ij} \cdot P_{ij}$, which is activated by the softmax function $S(I_{out}) = \frac{e^{\xi \cdot I_{out}}}{\sum_{k=1}^{3} e^{\xi \cdot R_k P_k}}$, where k is the elements within the “P”, “K”, “U” set, and ξ is the scaling factor. A loss function $\psi = -\frac{1}{k}\sum_{k=1}^{3} y_m \log\,[S(I_{out})]$ is defined to quantify the error for training, where y_m is the label of each input letter. Through iterative training, employing the backpropagation algorithm with a learning rate of 0.001, the photoresponsivity R_{ij} of each sensor in the array is fine-tuned by modulating the gate voltage V_g, thereby minimizing the loss function and enhancing classification accuracy.

(2) Recognition of Traffic-Light Color

For the image processing and convolutional neural networks (CNNs), convolutional kernels play a pivotal role in tasks such as image blurring, sharpening, edge detection, and the extraction of salient features (e.g., edges, textures, and patterns). These kernels are instrumental in discerning specific attributes within images, thereby facilitating a nuanced understanding and manipulation of visual data. **Fig. 5a** illustrates the implementation of three quintessential convolutional kernels. The inaugural function, inversing, serves as a stylization technique, transforming the visual aesthetics of an image. Subsequently, the denoising function, characterized by its operator weights arranged in accordance with the 2D Gaussian distribution $G = \frac{1}{2\Pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}$, is colloquially termed Gaussian blur due to its efficacy in blurring and noise reduction. The terminal function, dedicated to edge detection, employs a discrete form of

Laplacian operator $L(x,y)=\frac{\partial^2 I}{\partial x^2}+\frac{\partial^2 I}{\partial y^2}$, where $L(x,y)$ is the operator and I is the pixel intensity of images. The Laplacian operator, characterized by its 2D isotropic measure of the second-order spatial derivative, exhibits high sensitivity to noise, thereby necessitating a preliminary denoising phase to optimize the clarity and precision of edge detection.

We employed a 3×3 sensor array to formulate a Gaussian filter operator, aimed at denoising images marred by noise. For the task of traffic-light detection, where the input images feature three RGB channels, we executed a convolution of each RGB channel individually with the Gaussian kernel. The images in the LISA Traffic Lights dataset possess a resolution of 1280×720 pixels, and to simulate realistic conditions, we introduced Gaussian noises to these images prior to their use in the traffic light recognition task. The Gaussian smoothing convolution operator in the size of 3×3 was designed based on the distribution patterns of a 2D Gaussian function, ensuring that the center pixel exerts the most significant influence on the outcome, with a progressively diminishing contribution from the peripheral pixels. This design facilitates a smoothing effect, with a typical discrete approximation of the weight distribution depicted in **Fig. 5a**, displaying the center: edge: corner weight ratio of 4:2:1 for the 3×3 matrix.

Given the tri-channel nature of the input images, we deployed three distinct 3×3 operators tailored for the red, green, and blue channels, respectively. For the red channel, a 780 nm wavelength was chosen, adjusting the biases for the central, edge, and corner pixels within the 3×3 sensor array to V_g=-1.3 V, -0.8 V, and -0.7 V, respectively. This resulted in varying photoresponsivity R of 1.0×10^7 A/W, 5.0×10^6 A/W, and 2.5×10^6 A/W, respectively. Similarly, wavelengths of 532 nm and 400 nm were selected for the green and blue channels, with the sensor arrays configured accordingly to serve as convolution kernels in image processing. In subsequent simulations, the input images were dissected into their constituent RGB channels, with pixel values linearly transformed from the 0-255 range to correspond to an incident light intensity in the range of 1.5-150 μW/cm². A convolutional kernel was then applied by sliding it across the image, commencing with the initial 3×3 window, and employing a stride of 1. For each 3×3-pixel sub-window, the simulation computed the photocurrent $I_{ph}=\Sigma RP$, where P is the product of pixel value. The output value of each sliding window was then used as the pixel value for the corresponding position in the denoised output image".

Changes Made:

We have updated this section of "Implementation of Pattern Classification" in the Methods as follows (highlighted in red): "The task of classifying the letters "P", "K", "U" is executed using a Multilayer Perceptron (MLP) model. In this task, the architecture encompasses an input layer furnished with 9 neurons $(x_1, x_2, \dots, x_9)$, mirroring the 9-pixel values of the input letter image, and culminates in an output layer with 3 neurons (y_P, y_K, y_U), each dedicated to one of the target letters. The interconnection between input and output strata is facilitated via a weight matrix W with the output neuron activation a_j determined by $a_j = \sum_{i=1}^{9} W_{ij} \cdot x_i + b_j$, where W_{ij} is the weight from the input neuron i to the output neuron j, and b_j is the bias term for the output neuron j. This configuration enables the precise recognition of letters, as the network discriminates based on the highest activation value among the output neurons.

In a typical hardware realization, this neural construct can be emulated using an array of bipolar sensors, epitomizing in-sensor computing. In this work, the photoresponsivity R_{ij} of $MAPbI_3/Bi_2O_2Se$ heterotransistor array are tailored through the gate voltage $V_{g_{ij}}$ to reflect the synaptic weights W_{ij} of a neural network. Upon exposure to light encoding the input data P_{ij}, each device in the array generates a photocurrent I_{ij} proportional to the product of the input signal and its responsivity: $I_{ij} = R_{ij} \cdot P_{ij}$. The total current output $I_{out} = \sum_{i,j}^{i=3,j=3} I_{ij} = \sum_{i,j}^{i=3,j=3} R_{ij} \cdot P_{ij}$, embodies the multiply-accumulate (MAC) operation within the sensor array, harnessing Kirchhoff's law to execute computation.

For the classification task, a specialized dataset comprising 100 instances of Gaussian-noise afflicted "P", "K", and "U" letter patterns (i.e., noise standard deviation of 0.2) is segmented into training and testing sets in a 7:3 ratio. The letters are illuminated onto the sensor array with peak incident light intensity set at 0.1 μW/cm^2. Comparative analysis is conducted across three sensor arrays, differentiated by the photoresponsivity of 10^2 A/W, 10^4 A/W, and

10^6 A/W, respectively. The photoresponsivity of each heterotransistor is all initialized randomly at the beginning. The projection of the letter image induces a readout current $I_{out} = \sum_{i,j}^{i=3,j=3} I_{ij} = \sum_{i,j}^{i=3,j=3} R_{ij} \cdot P_{ij}$, which is activated by the softmax function $S(I_{out}) = \frac{e^{\xi \cdot I_{out}}}{\sum_{k=1}^{3} e^{\xi \cdot R_k P_k}}$, where k is the elements within the "P", "K", "U" set, and ξ is the scaling factor. By integrating impacts of noise and hysteresis into our simulation, and to elucidate the effects of different responsiveness on recognition accuracy, a Gaussian noise with the variance of 0.003 was deployed on the readout photocurrent. A loss function $\psi = -\frac{1}{k}\sum_{k=1}^{3} y_m \log\left[S(I_{out})\right]$ is defined to quantify the error for training, where y_m is the label of each input letter. Through iterative training, employing the backpropagation algorithm with a learning rate of 0.001, the photoresponsivity R_{ij} of each sensor in the array is fine-tuned by modulating the gate voltage V_g, thereby minimizing the loss function and enhancing classification accuracy. It is worth noting that, due to the device-to-device variations in the array, which are significant non-ideal factors in neuromorphic computing, we edited the photoresponsivity-gate voltage (R-V_g) data measured from the nine devices into a lookup table. This enables us to accurately simulate the conductance modulation process of these devices".

We have also included the correlated context into the Supporting Information as **Supplementary Note 3**: "For the image processing of convolutional neural networks (CNNs), convolutional kernels play a pivotal role in tasks such as image blurring, sharpening, edge detection, and the extraction of salient features (e.g., edges, textures, and patterns). These kernels are instrumental in discerning specific attributes within images, thereby facilitating a nuanced understanding and manipulation of visual data. **Fig. 5a** illustrates the implementation of three quintessential convolutional kernels. The inaugural function, inversing, serves as a stylization technique, transforming the visual aesthetics of an image. Subsequently, the denoising function, characterized by its operator weights arranged in accordance with the 2D Gaussian distribution $G = \frac{1}{2\Pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}$, is colloquially termed Gaussian blur due to its efficacy in blurring and noise reduction. The terminal function, dedicated to edge detection, employs a discrete form of Laplacian operator $L(x,y) = \frac{\partial^2 I}{\partial x^2} + \frac{\partial^2 I}{\partial y^2}$, where $L(x,y)$ is the operator and I is the pixel intensity of images. The Laplacian operator, characterized by its

2D isotropic measure of the second-order spatial derivative, exhibits high sensitivity to noise, thereby necessitating a preliminary denoising phase to optimize the clarity and precision of edge detection.

We employed a 3×3 sensor array to formulate a Gaussian filter operator, aimed at denoising images marred by noise. For the task of traffic-light detection, where the input images feature three RGB channels, we executed a convolution of each RGB channel individually with the Gaussian kernel. The images in the LISA Traffic Lights dataset possess a resolution of 1280×960 pixels, and to simulate realistic conditions, we introduced Gaussian noises to these images prior to their use in the traffic light recognition task. The Gaussian smoothing convolution operator in the size of 3×3 was designed based on the distribution patterns of a 2D Gaussian function, ensuring that the center pixel exerts the most significant influence on the outcome, with a progressively diminishing contribution from the peripheral pixels. This design facilitates a smoothing effect, with a typical discrete approximation of the weight distribution depicted in **Fig. 5a**, displaying the center: edge: corner weight ratio of 4:2:1 for the 3×3 matrix.

Given the tri-channel nature of the input images, we deployed three distinct 3×3 operators tailored for the red, green, and blue channels, respectively. For the red channel, a 780 nm wavelength was chosen, adjusting the biases for the central, edge, and corner pixels within the 3×3 sensor array to V_g=-1.3 V, -0.8 V, and -0.7 V, respectively. This resulted in varying photoresponsivity R of 1.0×10^{7} A/W, 5.0×10^{6} A/W, and 2.5×10^{6} A/W, respectively. Similarly, wavelengths of 532 nm and 400 nm were selected for the green and blue channels, with the sensor arrays configured accordingly to serve as convolution kernels in image processing. In subsequent simulations, the input images were dissected into their constituent RGB channels, with pixel values linearly transformed from the 0-255 range to correspond to an incident light intensity in the range of 1.5-150 μW/cm². A convolutional kernel was then applied by sliding it across the image, commencing with the initial 3×3 window, and employing a stride of 1. For each 3×3-pixel sub-window, the simulation computed the photocurrent $I_{ph}=\Sigma RP$, where P is the product of pixel value. The output value of each sliding window was then used as the pixel value for the corresponding position in the denoised output image".

==

Comment 3-7: What does the multilayer structure represent in Fig. 4b?

Response:

We regret any misunderstanding brought by the initial version of our schematic, and we have undertaken a thorough revision of **Fig. 4b** to enhance its clarity and convey its intended message more effectively.

As shown in **Fig. R27**, the diagram illustrates a neuromorphic computing architecture of in-sensor computing neural network. The uppermost layer presents a 3×3 pixel of the letter "P" embedded within noises, which is subsequently projected onto the 3×3 heterotransistor bipolar sensor array via a laser matrix. The intermediate layer depicts the $MAPbI_3/Bi_2O_2Se$ heterotransistor array, tasked with preprocessing the projected image. The photocurrents generated by this preprocessing are then read and transmitted to a computer for the final recognition task. The bottom layer shows the output from the computer, which successfully identifies the input image as the letter “P”. Through this illustration, we aim to elucidate the capability of our bipolar $MAPbI_3/Bi_2O_2Se$ heterotransistor array to efficiently preprocess and recognize images through an integrated sensing and computing functionalities.

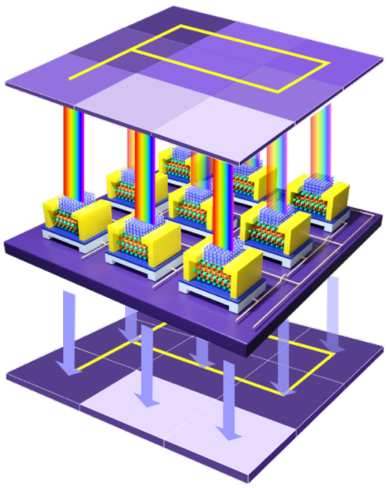


Figure R27. Illustration of our reconfigurable heterotransistor array for in-sensor neural network computing.

Changes Made:

We have revised **Fig. 4b** by redrawing the illustration that accurately displays the multilayer network architecture.

REVIEWERS' COMMENTS

Reviewer #1 (Remarks to the Author):

The manuscript by He et al. described an ultrasensitive and bipolar phototransistor based on $MAPbI_3/Bi_2O_2Se$ van der Waals heterojunction, which enables the implementation of neuromorphic vision sensing and classification under dim light. The revised manuscript demonstrates substantial improvements with lots of supplementary data/results. With a few minor revisions, the manuscript could be recommended for publication in Nature Communications.

1. Light-intensity dependence of specific detectivity (D*) in Fig. 3c should be updated to ensure the consistency with the claim of "photoresponsivity of 6×10^7 A/W, accompanied by the specific detectivity of 5.2×10^11 Jones" in abstract and main text.

Reviewer #2 (Remarks to the Author):

The authors have addressed my comments. I can recommend the publication of this manuscript.

Reviewer #3 (Remarks to the Author):

The authors have implemented significant revisions that have greatly enhanced the quality and clarity of the manuscript. They have thoroughly addressed all my previous comments and concerns, demonstrating a thoughtful and rigorous approach to their revisions. The manuscript is now well-organized, coherent, and presents its findings in a clear and compelling manner. Therefore, I strongly recommend that the manuscript be accepted for publication.

Responses to Review Comments

We thank all the Reviewers for recommending the publication of our manuscript. And we have addressed the minor revision suggested by the Reviewer 1.

Comments from Reviewer #1

The manuscript by He et al. described an ultrasensitive and bipolar phototransistor based on $MAPbI_3/Bi_2O_2Se$ van der Waals heterojunction, which enables the implementation of neuromorphic vision sensing and classification under dim light. The revised manuscript demonstrates substantial improvements with lots of supplementary data/results. With a few minor revisions, the manuscript could be recommended for publication in Nature Communications.

1. Light-intensity dependence of specific detectivity (D*) in Fig. 3c should be updated to ensure the consistency with the claim of "photoresponsivity of 6×10^7 A/W, accompanied by the specific detectivity of 5.2×10^{11} Jones" in abstract and main text.

Response:

We are sorry for the mistake. We have updated the light-intensity dependence of specific detectivity (D^*) in **Fig. 3c** by the experimentally-measured values to ensure consistency with the values of photoresponsivity (6×10^7 A/W) and specific detectivity (5.2×10^{11} Jones) as described in the abstract and main text. Thank you again for your careful checking on our manuscript.

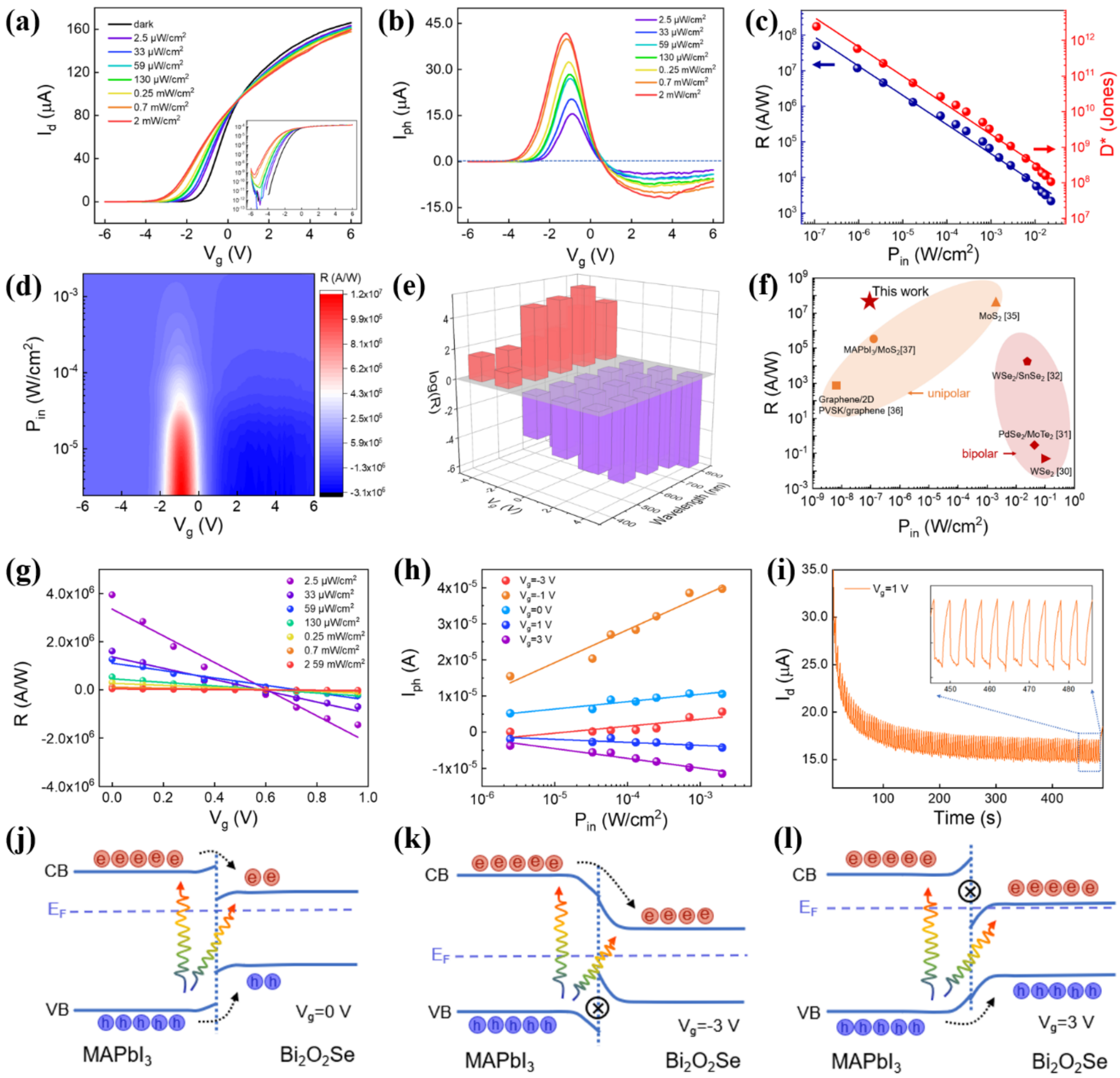


Figure 3. Characterization and mechanism of ultrasensitive reconfigurable $MAPbI_3/Bi_2O_2Se$ heterotransistors. (a) Transfer curves and (b) net photocurrents of $MAPbI_3/Bi_2O_2Se$ heterotransistor under the light illumination of 532 nm with varied light intensities. (c) Light-intensity dependence of the photoresponsivity R and the specific detectivity D^* under the light illumination of 532 nm at the gate voltage of 0 V. (d) Pseudo-color plots of the photoresponsivity under the light illumination of 532 nm with varied light intensities, in which the gate voltage was tuned from -6 V to 6 V. (e) Dependence of bipolar photoresponsivity on the gate voltage in visible illumination range of 400-800 nm. (f) Benchmark of the photoresponsivity in various unipolar and bipolar photosensors. (g) Linear dependence of the bipolar photoresponsivity on the gate voltage under the light illumination of 532 nm. (h) Dependence of the net photocurrent on the light intensity under the light illumination of 532 nm with the gate

voltage varying from -3 V to 3 V. (i) Photocurrent curves measured under the on/off recycles for negative photoresponses. (j)-(l) Illustration of band alignment and interlayer carrier transfer of the $MAPbI_3/Bi_2O_2Se$ heterotransistor under the gate voltage of 0 V, -3 V, and 3 V, respectively.

Comments from Reviewer #2

The authors have addressed my comments. I can recommend the publication of this manuscript.

Response:

Thank you for recommending the publication of our work. No revision is required by the reviewer.

Comments from Reviewer #3

The authors have implemented significant revisions that have greatly enhanced the quality and clarity of the manuscript. They have thoroughly addressed all my previous comments and concerns, demonstrating a thoughtful and rigorous approach to their revisions. The manuscript is now well-organized, coherent, and presents its findings in a clear and compelling manner. Therefore, I strongly recommend that the manuscript be accepted for publication.

Response:

Thank you for your recommendation of publishing our manuscript. No revision is required by the reviewer.