Article

A vision chip with complementary pathways for open-world sensing

🔴 Check for updates

Zheyu Yang[1,2,5], Taoyi Wang[1,5], Yihan Lin[1,5], Yuguo Chen[1], Hui Zeng[1], Jing Pei[1], Jiazheng Wang[1], Xue Liu[1], Yichun Zhou[2], Jianqiang Zhang[2], Xin Wang[2], Xinhao Lv[2], Rong Zhao[1,3 ✉] & Luping Shi[1,3,4 ✉]

Image sensors face substantial challenges when dealing with dynamic, diverse and unpredictable scenes in open-world applications. However, the development of image sensors towards high speed, high resolution, large dynamic range and high precision is limited by power and bandwidth. Here we present a complementary sensing paradigm inspired by the human visual system that involves parsing visual information into primitive-based representations and assembling these primitives to form two complementary vision pathways: a cognition-oriented pathway for accurate cognition and an action-oriented pathway for rapid response. To realize this paradigm, a vision chip called Tianmouc is developed, incorporating a hybrid pixel array and a parallel-and-heterogeneous readout architecture. Leveraging the characteristics of the complementary vision pathway, Tianmouc achieves high-speed sensing of up to 10,000 fps, a dynamic range of 130 dB and an advanced figure of merit in terms of spatial resolution, speed and dynamic range. Furthermore, it adaptively reduces bandwidth by 90%. We demonstrate the integration of a Tianmouc chip into an autonomous driving system, showcasing its abilities to enable accurate, fast and robust perception, even in challenging corner cases on open roads. The primitive-based complementary sensing paradigm helps in overcoming fundamental limitations in developing vision systems for diverse open-world applications.

Image sensors[1] play a crucial part in a broad range of applications[2–5]. However, with the increasing prevalence of open-world applications such as autonomous machines[6], robotics[7] and artificial intelligence[8], current image sensors face substantial challenges. Despite the development of algorithms tailored for open-world applications[9], image sensors have difficulty in handling dynamic, diverse and unpredictable corner cases[10] beyond their sensing range, causing algorithm failures[11]. Some of these challenges are aliasing and quantization error, data redundancy and limited dynamic range at the sensing level[11], as well as semantic misalignment, latency problem[12] and domain shift[13] at the perception level[14] (Fig. 1a). To effectively address corner cases, image sensors must possess exceptional performance in spatial resolution, speed, precision and dynamic range simultaneously. Yet, achieving this goal is impeded by the power and bandwidth walls[15]. Traditional sensors exhibit escalated power and bandwidth requirements with higher spatial resolution, speed and precision, leading to constrained capture abilities and excessive data burdens. Moreover, given the rarity of corner cases[16], these sensors are prone to generating redundant data, consequently wasting bandwidth and power resources.

In contrast to the existing image sensors, the human visual system (HVS) stands out for its versatility, adaptability and robustness in open-world environments. The HVS interprets visual stimuli into multiple visual primitives, such as colour, orientation and motion,

and allocates them to the ventral and dorsal pathways in a complementary manner[17]. The cooperation between the two pathways efficiently provides a unified representation of visual scenes[18] (Methods). Various endeavours have been made to replicate specific features of the HVS, including silicon retinas[19,20], neuromorphic vision sensors[21–28], pulse frequency modulation[29–32], time-to-first-spike[33–35] and near-sensor-computing chips[36–39]. However, challenges persist in achieving image sensors with high spatial resolution, high speed, high precision, and large dynamic range in the constraints of limited power and bandwidth.

Here we report a complementary sensing paradigm inspired by the multi-level characteristics of the HVS, along with a vision chip designed based on this paradigm, named Tianmouc. Our paradigm consists of a primitive-based representation and two complementary visual pathways (CVPs) that enable the parsing of a visual scene into primitives and their subsequent assembly into corresponding pathways. As shown in Fig. 1b, these primitives encompass colour, precision, sensitivity, spatial resolution, speed, absolute intensity, spatial difference (SD) and temporal difference (TD), serving as the foundational elements for a comprehensive representation of the scene (Supplementary Note 2). The CVP consists of two distinct pathways: the cognition-oriented pathway (COP) and the action-oriented pathway (AOP), analogous to the ventral and dorsal pathways of the HVS, respectively (Extended
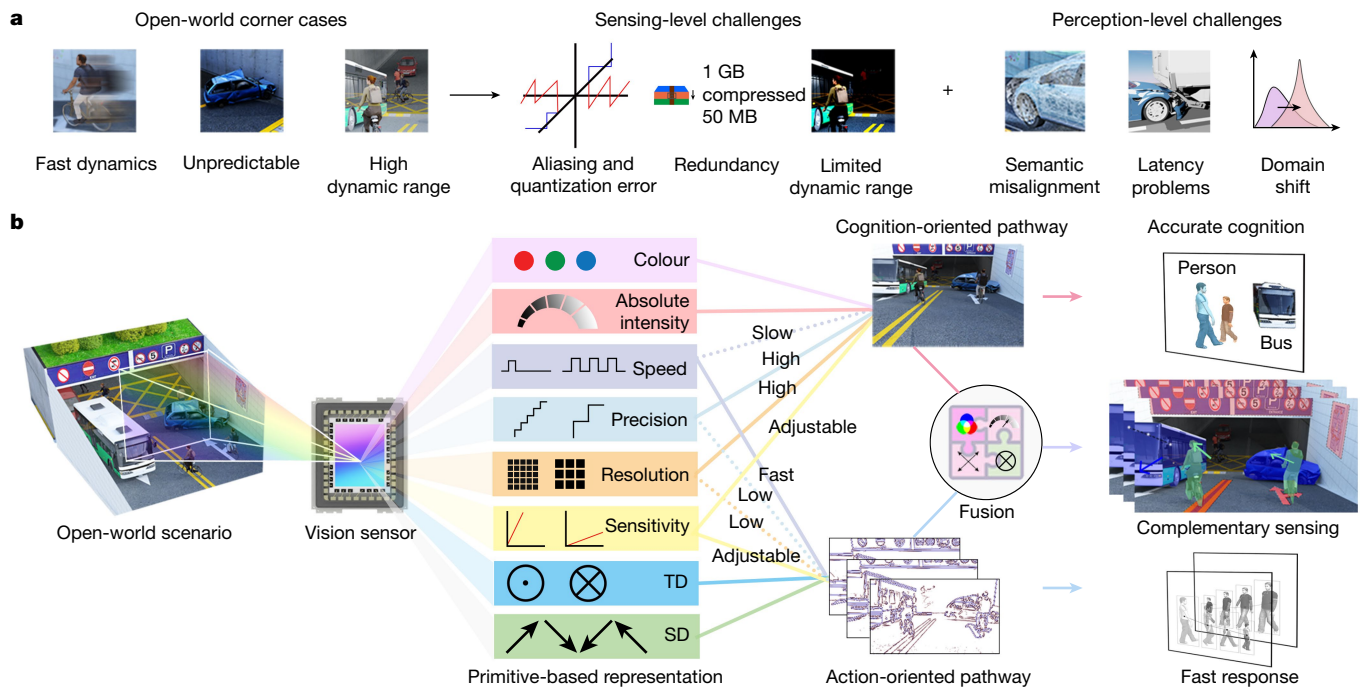
Fig. 1 | The challenges of open-world visual sensing and the solution with the complementary vision paradigm. a, Corner cases of the open world, such as dynamic, diverse and unpredictable situations, pose various challenges across both sensing and perception levels. **b**, The complementary visual paradigm requires the parsing of visual inputs into primitive-based representations, the assembly of these primitives into multiple pathways and their subsequent reorganization to accommodate different scenarios in the open world. Our vision sensor integrates these primitives into the COP and AOP, similar to the ventral and dorsal pathways, respectively, in the human visual system. TD, temporal difference with polarities ($\otimes$ for negative, $\odot$ for positive); SD, spatial difference with directions ($\nearrow$, $\searrow$, $\swarrow$, $\nwarrow$).

Data Table 1). The COP uses primitives of colour, intensity, high spatial resolution and high precision to achieve accurate cognition, minimizing spatial aliasing and quantization errors. By contrast, the AOP uses primitives of SD, TD and high speed to achieve fast responses with robustness and high sparsity, addressing data redundancy and latency problems. The two pathways complement each other in constructing representations of normal and corner cases, thereby achieving high dynamic range and mitigating semantic misalignment and domain shift problems.

Design of the complementary vision chip

Implementing the complementary sensing paradigm in a physical sensing system has several challenges that must be addressed. It is essential to design the pixel array such that it facilitates the simultaneous dissection of optical-to-electrical information conversion for corresponding primitives on the same focal plane. Furthermore, the readout architectures of the two pathways must incorporate heterogeneous building blocks that can encode electrical information with different data distributions and formats.

As shown in Fig. 2a, the Tianmouc chip, fabricated using a 90-nm CMOS back-side illuminated technology, consists of two centrepieces: a hybrid pixel array for converting optical information into electrical signals and a parallel-and-heterogeneous readout architecture for constructing two CVPs. Inspired by photoreceptor cells, the hybrid pixel array comprises cone-inspired and rod-inspired pixels with varying characteristics such as colours, response modes, resolutions and sensitivities. These pixels can parse visual information into specific colours (red, green and blue) and a white spectrum, serving as colour-opponent primitives. They can also be adjusted for four different sensitivities using high and low charge-to-electrical conversion gain, thereby achieving a high dynamic range by using the low noise of the high-gain mode and the high saturation capacity of the low-gain

mode. The cone-inspired pixels are designed with a fine-grained 4-μm pitch for absolute intensity sensing, whereas the rod-inspired pixels feature two larger receptive fields, 8 μm and 16 μm, for sensing TD and SD, respectively. A spatiotemporal consecutive pixel architecture is used to facilitate TD and SD computation through the use of high-density in-pixel memory. Specifically, the rod-inspired pixels buffer historical voltage signals in a ping-pong behaviour to enable continuous computation of TD in the AOP readout. The same memory in rod-inspired pixels across a block can be reorganized to compute the SD, as shown by the operational phase in Fig. 2b. The full hybrid pixel array comprises 320 × 320 cone-inspired pixels and 160 × 160 rod-inspired pixels. Further details about these two types of pixel are provided in the Methods and Extended Data Fig. 2a,b.

The electrical signals travelling along the two pathways exhibit distinct characteristics, including differences in data distributions and sparsity. These disparities require the use of different methods for encoding signals into digital data with appropriate speed and precision. To address this challenge, the chip adopts parallel-and-heterogeneous readout architectures. For the COP, accurate conversion of absolute intensity signals to dense matrixes is essential. This is achieved through a single-slope analog-to-digital architecture. By contrast, the AOP requires rapid encoding of signals with spatiotemporal differences characterized by symmetrical Laplacian-like distribution[40] and high sparsity. Therefore, a specialized readout architecture is used (Fig. 2c), in which a programmable threshold filter is used to minimize redundancy and noise in the computed TD and SD signals while retaining key information. Subsequently, these signals are quantized using a fast polarity-adaptive digital-to-analog converter with configurable precision. Moreover, a data packetizer is used to achieve lossless compression of the sparse variable-precision TD and SD signals in a unified protocol (Extended Data Fig. 2d). This approach offers adaptive abilities to reduce bandwidth and further enhance the operation speed of the AOP. More details on the readout architectures can be found in the
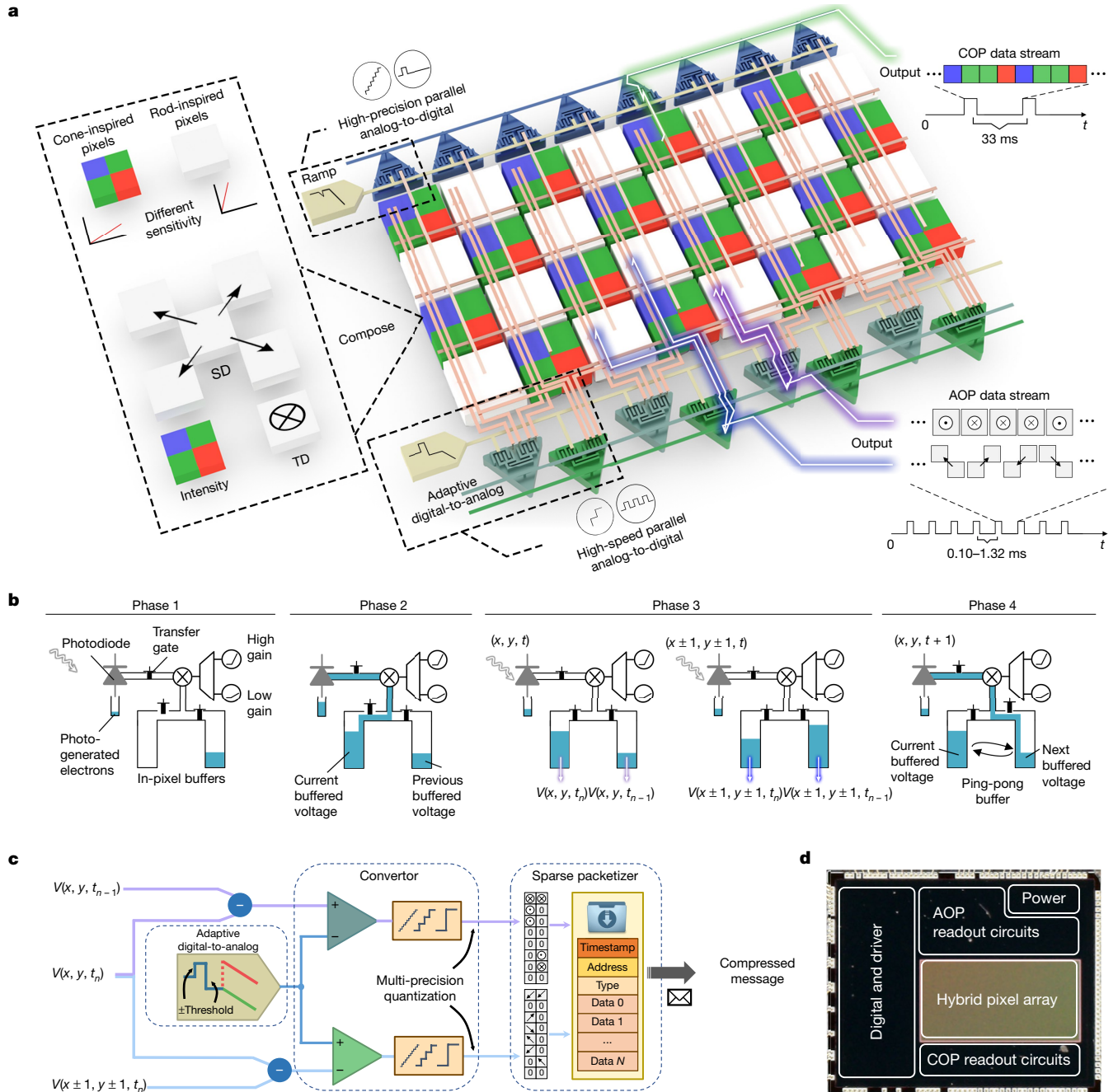
Fig. 2 | The architecture of the Tianmouc chip. a, Schematic of the chip architecture, including the hybrid pixel array and its interaction with multiple pathways. Various primitives such as colour, intensity, TD, SD, sensitivity, spatial resolution, speed and precision are implemented at different levels within the chip. **b**, In phases 1 and 2, photogenerated electrons are integrated into photodiodes and then converted to voltages with high- or low-gain amplification. In phase 3, the voltages buffered in the pixel are read out for spatiotemporal difference computation, whereas the electrons are integrated for the next sampling time. In phase 4, the in-pixel buffer behaves in a ping-pong manner. **c**, Electrical schematics of the AOP readout circuits. The TD and SD are first calculated by a subtractor circuit, filtered by a programmable threshold and then processed by a multi-precision convertor. The sparse TD and SD values are packetized at the final stage to achieve maximum bandwidth reduction. **d**, A microscopic image of the Tianmouc chip.

Methods and Extended Data Fig. 2c,d. An optical micrograph depicting the overall layout of the Tianmouc chip is presented in Fig. 2d.

Characterization of Tianmouc

The performance metrics of the Tianmouc chip, including quantum efficiency, dynamic range, response speed, power and bandwidth, have been evaluated comprehensively. The chip demonstrates high quantum efficiency in both COP and AOP, achieving a maximum of 72% of AOP and 69% of COP at 530 nm (Fig. 3a). It accomplishes a high dynamic range by leveraging the dynamic ranges of different gain modes in the complementary COP and AOP. As plotted in Fig. 3b, an overall dynamic range of 130 dB is achieved by detecting the lowest power density of $2.71 \times 10^{-3} \, \mu W \, cm^{-2}$ and the highest power density of $8.04 \times 10^{3} \, \mu W \, cm^{-2}$, in accordance with a well-established standard (Methods and Extended Data Fig. 4).

The complementary pathways of the Tianmouc chip enable high spatial resolution and precision (Fig. 3c) and high robustness in
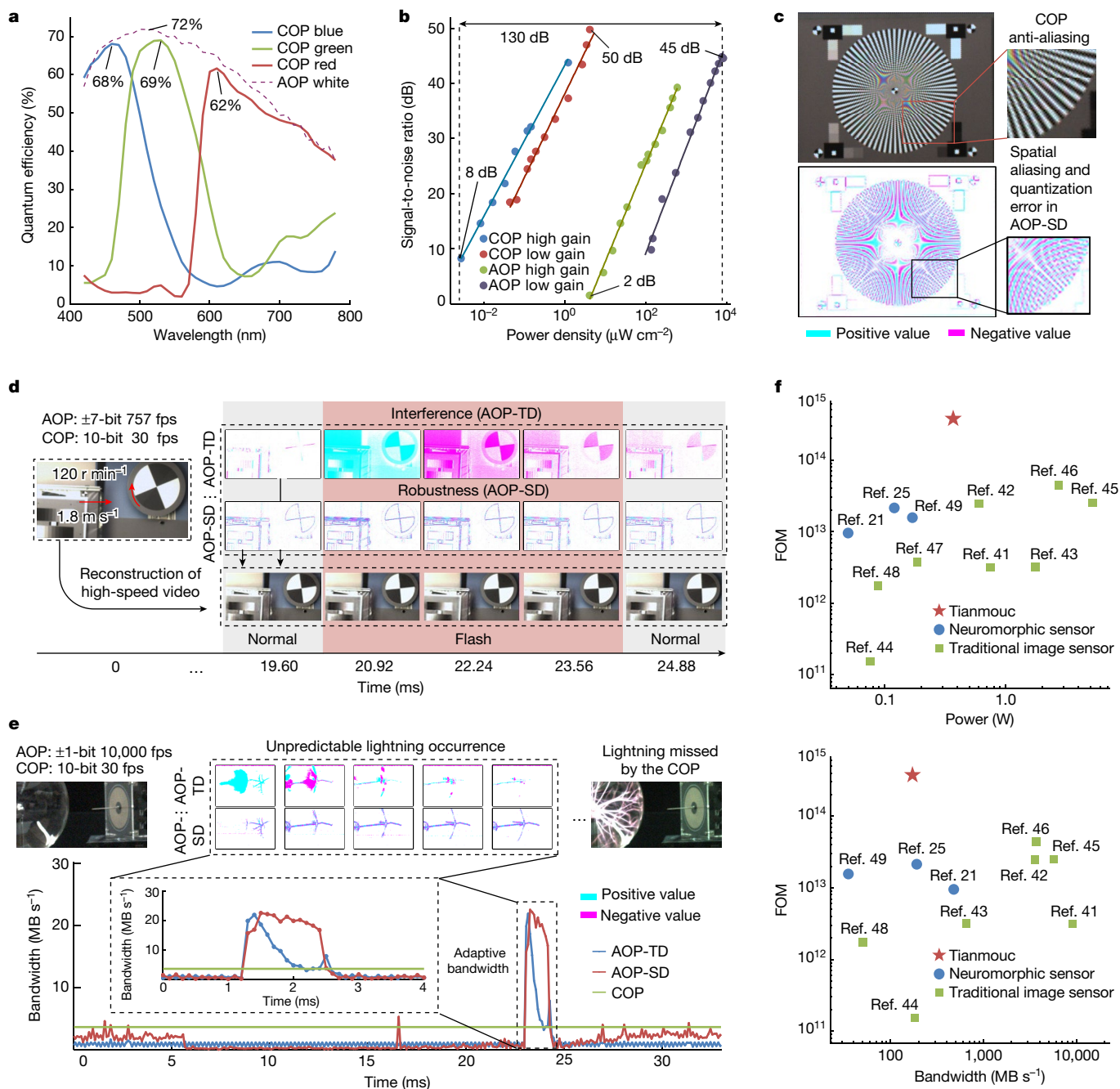
Fig. 3 | Summary of chip evaluation. a, Measured quantum efficiency curves show that both pathways exhibit a broad spectrum of response while maintaining high spectral sensitivities. **b**, Measured signal-to-noise ratios against optical power density show the high dynamic range ability achieved by combining high gain and low gain from the AOP and COP. **c**, Tianmouc eliminates spatial aliasing of the AOP through high spatial resolution and precision of the COP. **d**, Robust reconstruction of high-speed video for scenes with high-speed movements under light flash interference. **e**, Leveraging the high speed of the AOP, Tianmouc can promptly respond to unpredictable and rapidly occurring lightning. The bottom portion shows that the bandwidth consumption of the AOP remains lower than that of the COP, except during lightning events. **f**, A comparison between the Tianmouc chip and traditional image sensors[41–48], as well as neuromorphic vision sensors[21,25,49], on FOM with respect to power (top) and bandwidth consumption (bottom).

unpredictable environments (Fig. 3d). To eliminate spatial aliasing and quantization errors caused by AOP, the Tianmouc chip complementally uses spatial resolution and precision. Although the standard Siemens star chart captured by the AOP-SD in Fig. 3c may appear distorted because of its low resolution, the COP accurately records it. As shown in Fig. 3d, in a scene with horizontally fast-moving and rotating objects, and changing illumination conditions, a sudden flash of light disturbs the AOP-TD, but the AOP-SD remains unaffected. By using the COP image with the AOP-TD and AOP-SD, a frame-by-frame

reconstruction of high-speed video (Methods) enables the recovery of high-speed motion.

Using the AOP, Tianmouc demonstrates rapid response with reconfigurable speeds ranging from 757 frames per second (fps) to 10,000 fps and precisions varying from ±7 bits to ±1 bit. This complements the comparatively slower speed of the COP, which maintains a sustained response at 30 fps and a 10-bit resolution. The high-speed ability of Tianmouc is assessed by a transient lightning test. As shown in Fig. 3e, Tianmouc can operate at 10,000 fps with ±1 bit at a threshold level of
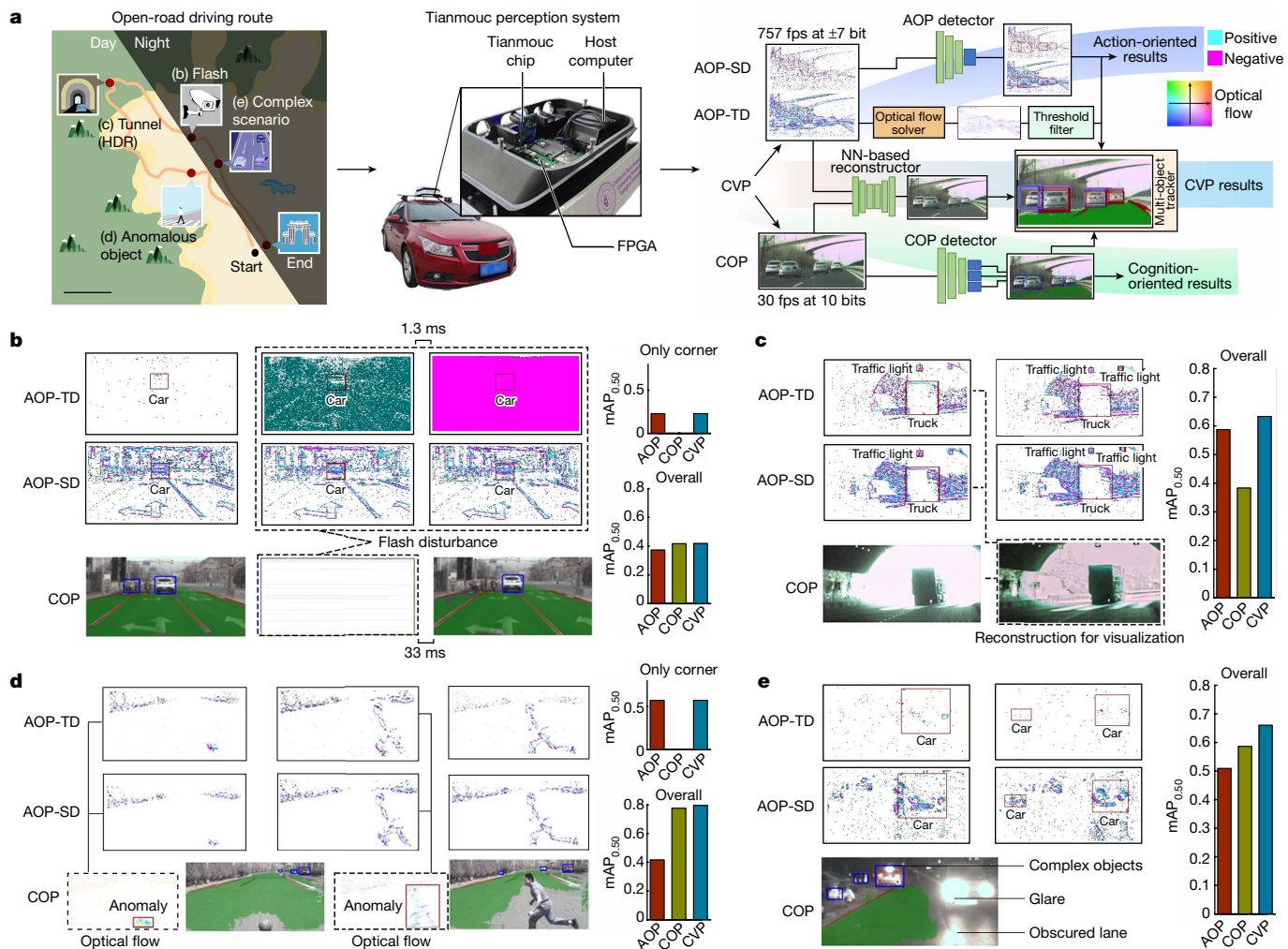
Fig. 4 | Open-world perception experiments. a, A long-distance drive is performed to test the performance of the integrated perception system with a Tianmouc, in which four different types of corner case are encountered. An asynchronous multi-tasking algorithm is designed based on the CVP. The results from the COP (blue detection boxes, green and red masks for segmentation) and those from the AOP (dark red) can be seamlessly aligned and synchronized to provide comprehensive outcomes. The mean average precision ($mAP_{0.50}$) is calculated for the corner cases (Supplementary Note 8). A higher $mAP_{0.50}$ value indicates better algorithm performance. In instances in which corner cases interfere with only a few frames, performance metrics are also calculated specifically for the influenced frames (only corner), apart from the overall performance. **b**, The AOP-SD demonstrates rapid response and strong anti-interference characteristics, effectively mitigating flash disturbances. **c**, Tianmouc supports tasks in high-speed, high-dynamic-range scenes, facilitated by the efficient extension of the dynamic range provided by the CVP. **d**, The AOP supports a high-speed optical flow estimation, effectively coping with anomalous objects. **e**, Tianmouc demonstrates proficiency in concurrently handling multiple corner cases through the cooperation of the CVP. Scale bar, 5 km (**a**).

50 mV to capture fast lightning bolts. It is worth noting that, owing to high sparsity, the AOP consumes only about 50 megabytes per second (MB s^{-1}) of peak bandwidth during transient phenomena, representing a 90% reduction compared with traditional cameras with equivalent spatiotemporal resolution and precision ($640 \times 320 \times 10,000 \times 2$). More demonstrations of high-speed responses and temporal anti-aliasing can be found in Extended Data Fig. 5.

We use a comprehensive figure of merit (FOM), similar to that proposed in ref. 41, to evaluate the overall performance of the Tianmouc chip. This FOM incorporates key performance indicators for open-world sensing, integrating the maximum sampling rate (R_{max}) and dynamic range into a unified metric ($R_{max} \times$ dynamic range). In Fig. 3f, the FOM is plotted against power and bandwidth for various sensors, respectively. The power consumption of Tianmouc varies based on the operation mode (Extended Data Fig. 5b) and averages at 368 mW in a typical mode (±7 bits, 1,515 fps without threshold). As shown in Fig. 3f, Tianmouc achieves an advanced FOM, surpassing the existing neuromorphic sensors and traditional image sensors, while still retaining low power and low bandwidth consumption. Detailed calculations and comparisons can be found in the Methods.

Performance in the open world

The complementary sensing paradigm provides a large space of design possibilities and serves as an exceptional data source for perception algorithms. To assess these abilities in open-world scenarios, we develop an automotive driving perception system (Fig. 4a) integrated with a Tianmouc chip. The assessment is conducted on open roads, involving encounters with various corner cases, such as flash disturbances, high dynamic range scenes, domain shift problems (anomalous objects) and complex scenes featuring multiple corner cases. To make use of the advantages of Tianmouc architecture, we design a multi-pathway algorithm specifically tailored to harness the complementary features of the AOP and COP. At the sensing level, the completeness of primitives

Article

enables the reconstruction of the original scene and adaptation to extreme illumination. Meanwhile, at the perception level, the AOP provides the immediate perception of variations, textures and motion, whereas the COP offers fine semantic details. By synchronizing these outcomes, we achieve a comprehensive understanding of the scene.

The first scenario shown in Fig. 4b evaluates the sensing abilities involving a sudden light flash that causes rapid changes in illumination, potentially affecting sensor robustness. Tianmouc exhibits remarkable resilience to such light flashes while maintaining high perception performance in normal situations. For real-time high dynamic range perception (Fig. 4c), the complementary sensitivity of the two pathways enables the Tianmouc to sense high-brightness contrast without sacrificing speed. At the perception level, the anomaly detection ability is complemented by a high-speed optical flow filter on the AOP, in which the collaboration between AOP-TD and AOP-SD enables precise calculation of motion direction and speed for identifying anomalies (Fig. 4d). Figure 4e shows a complex scene with dim natural illumination, a chaotic traffic environment and sudden disturbances from artificial light, demanding diverse sensing abilities in terms of sampling speed, resolution and dynamic range. The algorithms on CVP provide complementary and diverse results, offering ample room for further decisions in these scenarios. According to the $mAP_{0.50}$ (mean average precision; Supplementary Note 8) bars, the CVP yields superior overall detection performance compared with using only a single pathway across all the cases in Fig. 4. Notably, it achieves this while consuming less than 80 MB s^{-1} bandwidth and an average power consumption of 328 mW. The experimental results demonstrate that Tianmouc can efficiently adapt to extreme light environments and provide domain-invariant multi-level perception abilities. Further details of the experiment setup and algorithms are provided in the Methods and Extended Data Figs. 6–8, whereas the performance evaluation of the algorithm is discussed in Supplementary Notes 7 and 8.

Discussion

Tianmouc excels in capturing intricate details of cognition while rapidly responding to unpredictable emergencies and motion simultaneously. It offers high speed, high dynamic range and high precision, while simultaneously maintaining adaptive low bandwidth. Unlike existing sensing paradigms, our approach overcomes the inefficiencies caused by homogenous representations and accommodates various corner cases in the open world. Compared with contemporary neuromorphic vision sensors, Tianmouc exhibits superior precision and comprehensive information while maintaining a fast and robust response in extreme environments. Importantly, its high scalability allows for advanced spatial resolution through advanced manufacturing, facilitating resolution-sensitive applications with low power and bandwidth requirements. The primitives can also potentially be designed with on-chip reconfiguration and pathway allocation flexibility, enabling active adaptation to different task requirements. The vision sensor with primitive-based complementary pathways provides a unique data source and sensing platform, opening a new avenue for developing advanced computer vision theories, algorithms and systems for open-world applications.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-024-07358-4.

1. Fossum, E. R. CMOS image sensors: Electronic camera-on-a-chip. *IEEE Trans. Electron Devices* **44**, 1689–1698 (1997).
2. Gove, R. J. in *High Performance Silicon Imaging* 2nd edn (ed. Durini, D.) 185–240 (Elsevier, 2019).
3. Yun, S. H. & Kwok, S. J. Light in diagnosis, therapy and surgery. *Nat. Biomed. Eng.* **1**, 0008 (2017).
4. Liu, Z., Ukida, H., Ramuhalli, P. & Niel, K (eds). *Integrated Imaging and Vision Techniques for Industrial Inspection* (Springer, 2015).
5. Nakamura, J. *Image Sensors and Signal Processing for Digital Still Cameras* (CRC Press, 2017).
6. Bogdoll, D., Nitsche, M. & Zöllner, M. Anomaly detection in autonomous driving: a survey. In *Proc. IEEE/CVF International Conference on Computer Vision and Pattern Recognition* 4488–4499 (CVF, 2022).
7. Hanheide, M. et al. Robot task planning and explanation in open and uncertain worlds. *Artif. Intell.* **247**, 119–150 (2017).
8. Sarker, I. H. Machine learning: algorithms, real-world applications and research directions. *SN Comp. Sci.* **2**, 160 (2021).
9. Joseph, K., Khan, S., Khan, F. S. & Balasubramanian, V. N. Towards open world object detection. In *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition* 5830–5840 (CVF, 2021).
10. Breitenstein, J., Termöhlen, J.-A., Lipinski, D. & Fingscheidt, T. Breitenstein, J., Termöhlen, J.-A., Lipinski, D. & Fingscheidt, T. Systematization of corner cases for visual perception in automated driving. In *2020 IEEE Intelligent Vehicles Symposium (IV)* 1257–1264 (IEEE, 2020).
11. Yan, C., Xu, W. & Liu, J. Can you trust autonomous vehicles: contactless attacks against sensors of self-driving vehicle. In *Proc. Def Con* 24, 109 (ACM, 2016).
12. Li, M., Wang, Y.-X. & Ramanan, D. Towards streaming perception. In *European Conf. Computer Vision* 473–488 (Springer, 2020).
13. Quiñonero-Candela, J., Sugiyama, M., Schwaighofer, A. & Lawrence, N. D. *Dataset Shift in Machine Learning* (Mit Press, 2008).
14. Khatab, E., Onsy, A., Varley, M. & Abouelfarag, A. Vulnerable objects detection for autonomous driving: a review. *Integration* **78**, 36–48 (2021).
15. Shu, X. & Wu, X. Real-time high-fidelity compression for extremely high frame rate video cameras. *IEEE Trans. Comput. Imaging* **4**, 172–180 (2017).
16. Feng, S. et al. Dense reinforcement learning for safety validation of autonomous vehicles. *Nature* **615**, 620–627 (2023).
17. Goodale, M. A. & Milner, A. D. Separate visual pathways for perception and action. *Trends Neurosci.* **15**, 20–25 (1992).
18. Nassi, J. J. & Callaway, E. M. Parallel processing strategies of the primate visual system. *Nat. Rev. Neurosci.* **10**, 360–372 (2009).
19. Mahowald, M. & Mahowald, M. in *An Analog VLSI System for Stereoscopic Vision* (ed. Mahowald, M.) 4–65 (Kluwer, 1994).
20. Zaghloul, K. A. & Boahen, K. Optic nerve signals in a neuromorphic chip I: Outer and inner retina models. *IEEE Trans. Biomed. Eng.* **51**, 657–666 (2004).
21. Son, B. et al. 4.1 A 640×480 dynamic vision sensor with a 9 μm pixel and 300 Meps address-event representation. In *2017 IEEE International Solid-State Circuits Conference (ISSCC)* 66–67 (IEEE, 2017).
22. Kubendran, R., Paul, A. & Cauwenberghs, G. A 256×256 6.3 pJ/pixel-event query-driven dynamic vision sensor with energy-conserving row-parallel event scanning. In *2021 IEEE Custom Integrated Circuits Conference (CICC)* 1–2 (IEEE, 2021).
23. Posch, C., Matolin, D. & Wohlgenannt, R. A QVGA 143 dB dynamic range frame-free PWM image sensor with lossless pixel-level video compression and time-domain CDS. *IEEE J. Solid-State Circuits* **46**, 259–275 (2010).
24. Leñero-Bardallo, J. A., Serrano-Gotarredona, T. & Linares-Barranco, B. A 3.6 μs latency asynchronous frame-free event-driven dynamic-vision-sensor. *IEEE J Solid-State Circuits* **46**, 1443–1455 (2011).
25. Prophesee. IMX636ES (HD) https://www.prophesee.ai/event-camera-evk4/ (2021).
26. Brandli, C., Berner, R., Yang, M., Liu, S.-C. & Delbruck, T. A 240×180 130db 3μs latency global shutter spatiotemporal vision sensor. *IEEE J. Solid-State Circuits* **49**, 2333–2341 (2014).
27. Guo, M. et al. A 3-wafer-stacked hybrid 15MPixel CIS + 1 MPixel EVS with 4.6GEvent/s readout, in-pixel TDC and on-chip ISP and ESP function. In *2023 IEEE International Solid-State Circuits Conference (ISSCC)* 90–92 (IEEE, 2023).
28. Kodama, K. et al. 1.22 μm 35.6Mpixel RGB hybrid event-based vision sensor with 4.88 μm-pitch event pixels and up to 10 K event frame rate by adaptive control on event sparsity. In *2023 IEEE International Solid-State Circuits Conference (ISSCC)* 92–94 (IEEE, 2023).
29. Frohmader, K. P. A novel MOS compatible light intensity-to-frequency converter suited for monolithic integration. *IEEE J. Solid-State Circuits* **17**, 588–591 (1982).
30. Huang, T. et al. 1000× faster camera and machine vision with ordinary devices. *Engineering* **25**, 110–119 (2023).
31. Wang, X., Wong, W. & Hornsey, R. A high dynamic range CMOS image sensor with inpixel light-to-frequency conversion. *IEEE Trans. Electron Devices* **53**, 2988–2992 (2006).
32. Ng, D. C. et al. Pulse frequency modulation based CMOS image sensor for subretinal stimulation. *IEEE Trans. Circuits Syst. II Express Briefs* **53**, 487–491 (2006).
33. Culurciello, E., Etienne-Cummings, R. & Boahen, K. A. A biomorphic digital image sensor. *IEEE J. Solid-State Circuits* **38**, 281–294 (2003).
34. Shoushun, C. & Bermak, A. Arbitrated time-to-first spike CMOS image sensor with on-chip histogram equalization. *IEEE Trans. Very Large Scale Integr. VLSI Syst.* **15**, 346–357 (2007).
35. Guo, X., Qi, X. & Harris, J. G. A time-to-first-spike CMOS image sensor. *IEEE Sens. J.* **7**, 1165–1175 (2007).
36. Shi, C. et al. A 1000 fps vision chip based on a dynamically reconfigurable hybrid architecture comprising a PE array processor and self-organizing map neural network. *IEEE J. Solid-State Circuits* **49**, 2067–2082 (2014).
37. Hsu, T.-H. et al. A 0.8 V intelligent vision sensor with tiny convolutional neural network and programmable weights using mixed-mode processing-in-sensor technique for image classification. In *2022 IEEE International Solid-State Circuits Conference (ISSCC)* 1–3 (IEEE, 2022).
38. Lefebvre, M., Moreau, L., Dekimpe, R. & Bol, D. 7.7 A 0.2-to-3.6TOPS/W programmable convolutional imager SoC with in-sensor current-domain ternary-weighted MAC operations for feature extraction and region-of-interest detection. In *2021 IEEE International Solid-State Circuits Conference (ISSCC)* 118–120 (IEEE, 2021).

39. Ishikawa, M., Ogawa, K., Komuro, T. & Ishii, I. A CMOS vision chip with SIMD processing element array for 1 ms image processing. In *1999 IEEE International Solid-State Circuits Conference* 206–207 (IEEE, 1999).

40. Shi, Y.-Q. & Sun, H. *Image and Video Compression for Multimedia Engineering: Fundamentals, Algorithms, and Standards* 3rd edn (CRC Press, 2019).

41. Sakakibara, M. et al. A 6.9-μm pixel-pitch back-illuminated global shutter CMOS image sensor with pixel-parallel 14-bit subthreshold ADC. *IEEE J. Solid-State Circuits* **53**, 3017–3025 (2018).

42. Seo, M.-W. et al. 2.45 e-rms low-random-noise, 598.5 mW low-power, and 1.2 kfps high-speed 2-Mp global shutter CMOS image sensor with pixel-level ADC and memory. *IEEE J. Solid-State Circuits* **57**, 1125–1137 (2022).

43. Bogaerts, J. et al. 6.3 105×65 mm² 391Mpixel CMOS image sensor with >78 dB dynamic range for airborne mapping applications. In *2016 IEEE International Solid-State Circuits Conference (ISSCC)* 114–115 (IEEE, 2016).

44. Park, I., Park, C., Cheon, J. & Chae, Y. 5.4 A 76 mW 500 fps VGA CMOS image sensor with time-stretched single-slope ADCs achieving 1.95e⁻ random noise. In *2019 IEEE International Solid-State Circuits Conference* (ISSCC) 100–102 (IEEE, 2019).

45. Oike, Y. et al. 8.3 M-pixel 480-fps global-shutter CMOS image sensor with gain-adaptive column ADCs and chip-on-chip stacked integration. *IEEE J. Solid-State Circuits* **52**, 985–993 (2017).

46. Okada, C. et al. A 50.1-Mpixel 14-bit 250-frames/s back-illuminated stacked CMOS image sensor with column-parallel *kT*/*C*-canceling S&H and ΔΣADC. *IEEE J. Solid-State Circuits* **56**, 3228–3235 (2021).

47. Solhusvik, J. et al. 1280 × 960 2.8 μm HDR CIS with DCG and split-pixel combined. In *Proc. International Image Sensor Workshop* 254–257 (2019).

48. Murakami, H. et al. A 4.9 Mpixel programmable-resolution multi-purpose CMOS image sensor for computer vision. In *2022 IEEE International Solid-State Circuits Conference (ISSCC)* 104–106 (IEEE, 2022).

49. iniVation. DAVIS 346, https://inivation.com/wp-content/uploads/2019/08/DAVIS346.pdf (iniVation, 2019).

Article

Methods

A brief introduction to HVS

The HVS differs fundamentally from frame-based image sensors, boasting inherent general-purpose functionality, robustness and efficiency in the open world. As shown in Extended Data Fig. 1, the use of primitives and complementary pathways is apparent throughout the HVS, spanning from initial photoreceptor response to high-level visual processing. Notably, the retina exhibits distinct responses to input stimuli: cones detect colour and converge in the fovea with high acuity, whereas rods are colour insensitive and distributed across the retina[50]. Rather than transmitting the entire signal from the retina to the brain, the lateral geniculate nucleus parses the response from the retina to different characteristics and transmits them to the primary visual cortex by the magnocellular and parvocellular (M and P) pathways. The P-pathway maintains a sustained response to colour information, offering a high spatial resolution but a low temporal resolution. Complementarily, the M-pathway, insensitive to colour, are sensitive to transient information, exhibiting a fast temporal response albeit at the expense of spatial resolution. Subsequently, the primary visual cortex (V1 and V2 regions) interprets these inputs into multiple visual primitives, including colour, orientation, direction and depth, which are then re-composed into ventral and dorsal pathways for high-level perception and immediate action, respectively[51]. These primitives and pathways enable the HVS to achieve unified and coherent representations of the open world while significantly reducing data redundancy without ignoring emergencies. Moreover, even in scenarios in which certain primitives or pathways are out of operation in some corner cases, their combination can compensate in a complementary manner. In summary, the HVS serves as a rich source of inspiration when tackling open-world challenges. However, harnessing these advantages still requires the development of mathematical models, theoretical analysis and silicon implementation.

A complementary vision paradigm

Here we present a synthesis of neuroscience findings and the key features of the HVS that are implicated in our complementary sensing paradigm. The HVS comprises two main processing pathways: one dedicated to the vision for perception (ventral pathway) and the other to vision for action (dorsal pathway). Although visual perception and cognition are often associated with semantic attributes that require high precision and high spatial resolution, for vision-guided behaviours, the absolute intensity of colours is deemed less important compared with transient and gradient information with higher temporal resolution. Drawing inspiration from the ventral and dorsal pathways, we introduce the COP and the AOP as shown in Extended Data Table 1. Furthermore, we integrate key features from the photoreceptor level, the lateral geniculate nucleus and the high-level vision into each pathway as primitives. Similar to the HVS, our paradigm offers advantages across multiple levels. At the photoreceptor level, incorporating dual photoreceptors greatly extends sensitivity and dynamic range. At the sensor level, incorporating primitive-based representations enables high dynamic range, high precision, high spatiotemporal resolution and low latency sensing without data redundancy. At the perception level, the complementary sensing paradigm provides the potential for fast responses to emergencies while having high-precision cognition for critical objects. To quantitatively assess representation ability, we introduce the concept of completeness of representation and conduct a comprehensive theoretical analysis of the primitive-based representations (Supplementary Notes 1–3). The results show that the primitive-based representation of Tianmouc maintains completeness compared with a high-speed frame-based representation.

Architecture design

The Tianmouc chip adopts a hybrid pixel array and parallel-and-heterogenous readout architectures to simultaneously implement primitives and complementary pathways on the same focal plane, enabling support for the complementary sensing paradigm. Overall, the COP captures colour intensity accurately through fine-grained cone pixels and high-precision parallel analog-to-voltage conversion, whereas the AOP uses rod pixels, which can be reconfigured for varying high-speed and precision, to facilitate fast, sparse spatiotemporal difference sensing and further compression.

A schematic of the hybrid pixel array illuminated from the back side is shown in Extended Data Fig. 2a. The colour or white filters and microlens array are fabricated on the cone-inspired and rod-inspired pixels, respectively. The three-dimensional diagram of the pixel structure shows the arrangement of the photodiode, high-density storage, transistors and metal wires. An optical micrograph of the hybrid pixel array is provided on the right. Cone-inspired and rod-inspired pixels use the same sensing frontend (Extended Data Fig. 2b) to convert optical information to charges in the photodiode and transmit these charges through the transfer gate controlled in a global shutter. The charges are converted to electrical signals by high or low gain realized by the reset and low-gain reset transistors, along with an additional capacitor. Keeping the reset transistor always on adds the additional capacitor to the signal path to achieve low gain, while keeping the low-gain reset transistor always on achieves the high-gain mode. Different sensing backends (Extended Data Fig. 2b) in cone-inspired and rod-inspired pixels enable the two pathways to read out the voltage in different ways.

Low-noise readout of intensities in cone-inspired pixels is achieved by correlated double sampling (CDS) circuits (Extended Data Fig. 2c). The value processed by the CDS circuit is further compared with a linear ramp generated by a shared digital-to-analog converter (DAC) and converted to digital format by a high-precision counter. As shown in Extended Data Fig. 2c, the spatiotemporal signals stored in rod-inspired pixels are fed simultaneously into the corresponding analog-to-digital converters (ADCs). The TD is calculated based on the conservation of charges. For SD computing, signals in each rod-inspired pixel are first processed in the CDS circuit to reduce noise, and the results of the two rows are further subtracted to generate SD values. For high-speed conversion of the processed Laplacian-like distributed spatiotemporal difference, we adopt a unified and polarity-adaptive DAC with a programmable threshold that filters the spatiotemporal difference to preserve the critical information. Using the reconfigurable slope generated by the DAC, the digital value of TD and SD is quantized to different precisions from ±7 bits to ±1 bit with various speeds from 757 fps to 10,000 fps (see detailed specification in Extended Data Table 2 and Supplementary Note 4).

The data representation of the COP is a dense matrix, whereas the AOP data are sparse with variable length. The different data representation poses challenges for efficient data communication and storage. As shown in Extended Data Fig. 2d, a packetizer using the unified address difference representation to encode the sparse spatiotemporal difference by assembling the timestamp, the address of pixels that generate non-zero values, and the corresponding TD and SD to compact and unified packets with compatibility of different types (TD and SD) and various precisions.

Experimental setup for chip characterization

The system setup for chip characterization is shown in Extended Data Fig. 3. The test board is shown in Extended Data Fig. 3a. As shown in Extended Data Fig. 3b, the digital output of the chip is processed by a commercial FPGA board (AMD-Xilinx, EK-U1-ZCU106-G-J) and transmitted to the host computer (Nvidia, Jetson AGX Orin) through Peripheral Component Interconnect Express (PCIe) protocol for post-processing. To miniaturize the system size, the FPGA board is replaced with a smaller board (Milianke, MLK-H4-KU040) in the autonomous driving perception system.

The characterization of Tianmouc is conducted in two setups. The first is based on the European Machine Vision Association 1288

(EMVA1288) standard[52]. As shown in Extended Data Fig. 4a–c, a uniform light is generated by disk-shaped LEDs and monochromators, and then projected on the focal plane of the sensor placed in the machine (Looglook, ez1288-RD95) for quantum efficiency measurement. The digital output of the COP is processed in the host computer. Because the AOP generates only spatiotemporal differences, a high-speed ADC acquisition card (ART Technology, PCIE8914M) is used to record the analog output of the intensity signals of rod pixels to be compatible with EMVA1288. The intensity images of both pathways are analysed by EMVA1288 standard-based algorithms to generate measurement data.

The signal-to-noise ratio (SNR) curve for dynamic range measurement in Fig. 3b is evaluated using a customized optical setup (Extended Data Fig. 4d,e) because the light source in the standard EMVA1288 machine has a limited dynamic range and cannot be programmed. Collimated light from a laser (Fisba, READYBeam) and a collimator are filtered by a high-frequency filter consisting of an objective and a pinhole, and then expanded by a lens to form a uniform light spot. This spot is then projected onto an optical power meter (Thorlabs, PM100D with S120C for high laser power characterization and PM160 for low laser power measurement) and the sensor chip on the same optical path. In both experiments, AOP data are collected by triggering the laser to generate flickers projected on the chip and recording the TD data with ±7-bit precision.

The key to calculating sensitivity and SNR curve is the calculation of SNR using intensity images. The calculation of SNR is outlined below

$$\mathrm{SNR} = \frac{\mu - \mu_{\mathrm{dark}}}{\sigma_I}, \qquad (1)$$

where μ is the mean value of two consecutive normal images I_{t_n} and $I_{t_{n+1}}$ exposed under a light source, and μ_{dark} is the mean value of two consecutive dark images $I_{t_n}^{\mathrm{dark}}$ and $I_{t_{n+1}}^{\mathrm{dark}}$, σ is the standard deviation of the noise. Considering an image with a resolution of $M \times N$, μ and μ_{dark} are calculated by averaging across all rows m and columns n of the two images

$$\mu = \frac{1}{2MN} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} (I_{t_n}(m,n) + I_{t_{n+1}}(m,n)),$$

$$\mu_{\mathrm{dark}} = \frac{1}{2MN} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} (I_{t_n}^{\mathrm{dark}}(m,n) + I_{t_{n+1}}^{\mathrm{dark}}(m,n)). \qquad (2)$$

The σ is calculated using the mean values as

$$\sigma^2 = \frac{1}{2MN} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} (I_{t_n}(m,n) - I_{t_{n+1}}(m,n))^2. \qquad (3)$$

Chip and system characterization

Extended Data Fig. 5a demonstrates the ability of Tianmouc to effectively capture unpredictable fast-moving ping-pong balls shot by a machine in random directions. Tianmouc accurately captures the static background when no balls are shot. The ping-pong balls are ejected suddenly by the machine and their trajectory can be detected by both AOP-TD and AOP-SD. Moreover, clear textures of other objects are captured by the AOP-SD.

The power consumption of the Tianmouc chip is measured using multimeters (Fluke, 17B+ digital multimeter) for different operating modes, ranging from 328.4 mW to 419.7 mW (Extended Data Fig. 5b). Tianmouc achieves relatively low power consumption compared with traditional high-speed image sensors[41]. A breakdown of power for a typical operating mode is provided in Extended Data Fig. 5b.

The Tianmouc effectively addresses temporal aliasing by using the high speed of AOP. As shown in Extended Data Fig. 5c, the COP records

the rotation of a car wheel with aliasing, which is then accurately captured by AOP because of its ability for fast response. Through a reconstruction algorithm, an anti-aliasing reconstructed video is achieved, recovering the actual rotation at a speed of 757 fps.

As a supplement to Fig. 3e, we present the texture of lightning captured by Tianmouc under different operating modes. As shown in Extended Data Fig. 5d, the transient lightning with detailed texture is recorded by AOP at 1,515 fps, ±7 bits and a threshold of 50 mV. By applying an on-chip threshold filter, the peak bandwidth can be reduced to about 50 MB s⁻¹ at a mode of 10,000 fps with ±1-bit precision (Fig. 3e), and 55 MB s⁻¹ at a mode of 1,515 fps with ±7-bit precision (Extended Data Fig. 5d). This presents an 80–90% reduction compared with traditional cameras with equivalent spatiotemporal resolution and precision $(640 \times 320 \times 1,515 \times 8)$. Further details on the calculation of FOM can be found in Supplementary Note 5.

Scene reconstruction based on the CVP

For data reconstruction, a neural-network-based reconstructor is trained based on a self-supervised model. Two adjacent colour images from the COP and the AOP data stream between these images are used to reconstruct the original colourful scene. This process requires the same sampling rate as the AOP-SD and AOP-TD and the same resolution as the COP. The training process is adapted from ref. 53, incorporating the COP image at t_0 $(I_{t_0}^{\mathrm{C}})$, the accumulation of AOP-TD from t_0 ~ t_f $(\sum_{t=t_0 \sim t_f} I_t^{\mathrm{TD}})$ and the AOP-SD at the initial and final time points $(I_{t_0}^{\mathrm{SD}}, I_{t_f}^{\mathrm{SD}})$ to train a neural network F. The objective of F is to reconstruct the colour image at the target time $t_f (I_{t_f}^{\mathrm{C}})$, that is, $F(I_{t_0}^{\mathrm{C}}, \sum_{t_i=t_0}^{t_f} I_{t_i}^{\mathrm{TD}}, I_{t_0}^{\mathrm{SD}}, I_{t_f}^{\mathrm{SD}}) = \widehat{I_{t_f}^{\mathrm{C}}}$ and the time-reversed process is also considered, as $F(I_{t_f}^{\mathrm{C}}, -\sum_{t_i=t_0}^{t_f} I_{t_i}^{\mathrm{TD}}, I_{t_0}^{\mathrm{SD}}, I_{t_f}^{\mathrm{SD}}) = \widehat{I_{t_0}^{\mathrm{C}}}$.

The reconstruction network comprises three parts: a convolutional block attention module (CBAM)-based[54] tiny-UNet[55], an optical flow estimator and a fusion network, as shown in Extended Data Fig. 6a. The CBAM-based tiny-UNet takes the original COP data and $\sum_{t_i=t_0}^{t_f} I_{t_i}^{\mathrm{TD}}$ as input. The optical flow estimator, based on SpyNet[56], is modified by replacing the colour images with $I_{t_f}^{\mathrm{SD}}$ and $I_{t_0}^{\mathrm{SD}}$ as input. Then, $\sum_{t_i=t_0}^{t_f} I_{t_i}^{\mathrm{TD}}$ is added as a complement to the motion information, and the network structure is modified corresponding to the number of data channels, as shown in Extended Data Fig. 6b. The fusion network is also a UNet. Further details of the network structure can be found in Supplementary Table 1. The entire network is trained in an end-to-end manner, as shown in Extended Data Fig. 6c. By adjusting t_f, the reconstruction result of an arbitrary frame can be obtained, as shown in Extended Data Fig. 6d. Using this approach, at least 757 fps reconstructed colour images with a resolution of 640 × 320 can be achieved. Moreover, the network can be generalized to high dynamic range scenes, as shown in Fig. 4c. Details of the evaluation process of the reconstruction algorithm can be found in Supplementary Note 7.

Perception algorithms used in the experiments

For real-time perceptual tasks, we build a parallel and complementary data process pipeline, as shown in Extended Data Fig. 7. The system supports the streaming process of the Tianmouc chip to support CVP-based perception processes. To accomplish detection and tracking tasks in the open world, two neural network (NN)-based algorithms run parallelly in our system, including a multi-tasks network modified on YOLOPv1 (refs. 57,58), a high-speed detector modified on YOLOv5s originated from YOLOv4 (ref. 59), and a multiple-object tracker (MOT) based on Kalman filter and an optical flow filter. The complete system contains a Tianmouc chip, an FPGA and a host computer (NVIDIA Jetson AGX Orin). The Tianmouc serves as the core of the whole system, which senses visual signals, converts them into two pathways, encodes the signals from the two pathways in compressive digital format and outputs the digital data. The output data are transmitted from the chip to the FPGA by parallel or serial interfaces and then forwarded to the host computer using a PCIe interface without modifications. All perceptual

Article

algorithms are executed on the host computer. The training details and the evaluation method can be found in Supplementary Notes 6 and 7, with detailed evaluation results provided in Supplementary Note 8 using annotated Tianmouc datasets referenced in Fig. 4.

For complementary perception, the results are synchronized and integrated using a MOT based on a Kalman filter. Detection results from multiple pathways are time-stamped and transmitted through different buffers into a tracking thread. The MOT is set to record the historical results for about 150 ms and updates tracking results at the same speed as the detector on the AOP. It also gives a tracking trace by drawing the centre points of tracked targets. Detection results given by different pathways are synchronized in the same trace.

Moreover, an optical flow solver and optical flow filter are introduced to the detection-tracking task. We adopt the Horn–Schunck method[60] for real-time dense optical flow calculation. On getting the dense optical flow (u_{ij}, v_{ij}) for each point (i, j), we calculate the average of all non-zero optical flow values in the FOV to obtain the global optical flow $(u_{mean}, v_{mean}) = \left(\frac{\sum_{i=1}^{W}\sum_{j=1}^{H} u_{ij}}{N}, \frac{\sum_{i=1}^{W}\sum_{j=1}^{H} v_{ij}}{N} \right)$ to approximate the camera motion, where N is the number of pixels with non-zero optical flow vectors, and W and H are the width and height of the AOP frame, respectively. Based on this, targets that introduce inconsistencies with global motion can be filtered out as potential out-of-distribution obstacles through morphological operations. For more details, see Supplementary Note 9.

Bandwidth and data visualization in different open-world scenarios

To demonstrate the efficiency of Tianmouc, we calculate the bandwidth consumption for 10 different scenarios, while maintaining consistent precision and speed settings of the sensor. The results are presented in Extended Data Fig. 8a–d, corresponding to Fig. 4d–g, respectively. In Extended Data Fig. 8e, the performance of the algorithm is showcased while travelling on a tree-shaded road. Extended Data Fig. 8f shows a challenging scenario in which our vehicle passes over a speed bump, including notable camera shaking and large-amplitude vibration, a corner case in our evaluation. Despite these challenges, Tianmouc effectively tracks the target object owing to the high response speed of AOP. In Extended Data Fig. 8g, our vehicle navigates through roads with heavy traffic, resulting in dense AOP-TD, whereas in Extended Data Fig. 8h, the system operates on a highway with almost zero relative speed with other vehicles, leading to very sparse AOP-TD. Extended Data Fig. 8i shows the performance of Tianmouc when entering and leaving a short tunnel with a large variation in light on the same target. Extended Data Fig. 8j simulates an artificial anomaly in which many people are playing basketball at a tunnel exit and the vehicle keeps static. Here, the AOP-TD does not respond effectively, whereas the AOP-SD still provides a clear description of the bright part.

On the right of each case, we count and average the bandwidth consumption of the AOP-TD, AOP-SD, COP and their combination across the entire sample. The actual average bandwidth of Tianmouc ranges from 50 MB s^{-1} to 80 MB s^{-1}, with peak bandwidth generally below 80 MB s^{-1}, significantly less than traditional high-speed high-dynamic-range cameras. The main bandwidth consumption of Tianmouc is caused by the AOP-TD and AOP-SD because of their high sampling speed. However, the data distribution of the AOP-TD and AOP-SD and efficient coding methods reduce the bandwidth requirements of the AOP across all tested environments.

Data availability

The data supporting the findings of this study are available in the main text, Extended Data, Supplementary Information, source data and Zenodo (https://doi.org/10.5281/zenodo.10602822)[61]. Source data are provided with this paper.

Code availability

The algorithms and codes supporting the findings of this study are available at Zenodo (https://doi.org/10.5281/zenodo.10775253)[62].

50. Kandel, E. R., Koester, J. D., Mack, S. H. & Siegelbaum, S. A. *Principles of Neural Science* 4th edn (McGraw-Hill, 2000).
51. Mishkin, M., Ungerleider, L. G. & Macko, K. A. Object vision and spatial vision: two cortical pathways. *Trends Neurosci.* **6**, 414–417 (1983).
52. Jähne, B. EMVA 1288 Standard for machine vision: Objective specification of vital camera data. *Optik Photonik* **5**, 53–54 (2010).
53. Reda, F. A. et al. FILM: Frame Interpolation for Large Motion. In *Proc. IEEE/CVF International Conference on Computer Vision* 250–266 (ACM, 2022).
54. Woo, S., Park, J., Lee, J.-Y. & Kweon, I. S. CBAM: Convolutional Block Attention Module. In *Proc. European Conference on Computer Vision (ECCV)* 3–19 (2018).
55. Ronneberger, O., Fischer, P. & Brox, T. U-Net: convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference Proc.* Part III Vol. 18 (eds Navab, N. et al.) 234–241 (Springer, 2015).
56. Ranjan, A. & Black, M. J. CBAM: Convolutional Block Attention Module. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition* 4161–4170 (CVF, 2018).
57. Redmon, J., Divvala, S., Girshick, R. & Farhadi, A. You Only Look Once: unified, real-time object detection. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition* 779–788 (IEEE, 2016).
58. Wu, D. et al. YOLOP: You Only Look Once for Panoptic Driving Perception. *Mach. Intell. Res.* **19**, 550–562 (2022).
59. Bochkovskiy, A., Wang, C.-Y. & Liao, H.-Y. M. YOLOv4: optimal speed and accuracy of object detection. Preprint at https://arxiv.org/abs/2004.10934 (2020).
60. Horn, B. K. & Schunck, B. G. Determining optical flow. *Artif. Intell.* **17**, 185–203 (1981).
61. Wang, T. Tianmouc dataset. *Zenodo* https://doi.org/10.5281/zenodo.10602822 (2024).
62. Wang, T. Code of "A vision chip with complementary pathways for open-world sensing". *Zenodo* https://doi.org/10.5281/zenodo.10775253 (2024).
63. iniVation. *Understanding the Performance of Neuromorphic Event-based Vision Sensors* White Paper (iniVation, 2020).
64. iniVation. *DAVIS 346 AER* https://inivation.com/wp-content/uploads/2023/07/DAVIS346-AER.pdf (iniVation, 2023).

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41586-024-07358-4.
Correspondence and requests for materials should be addressed to Rong Zhao or Luping Shi.
**Peer review information** *Nature* thanks Craig Vineyard and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.
Reprints and permissions information is available at http://www.nature.com/reprints.
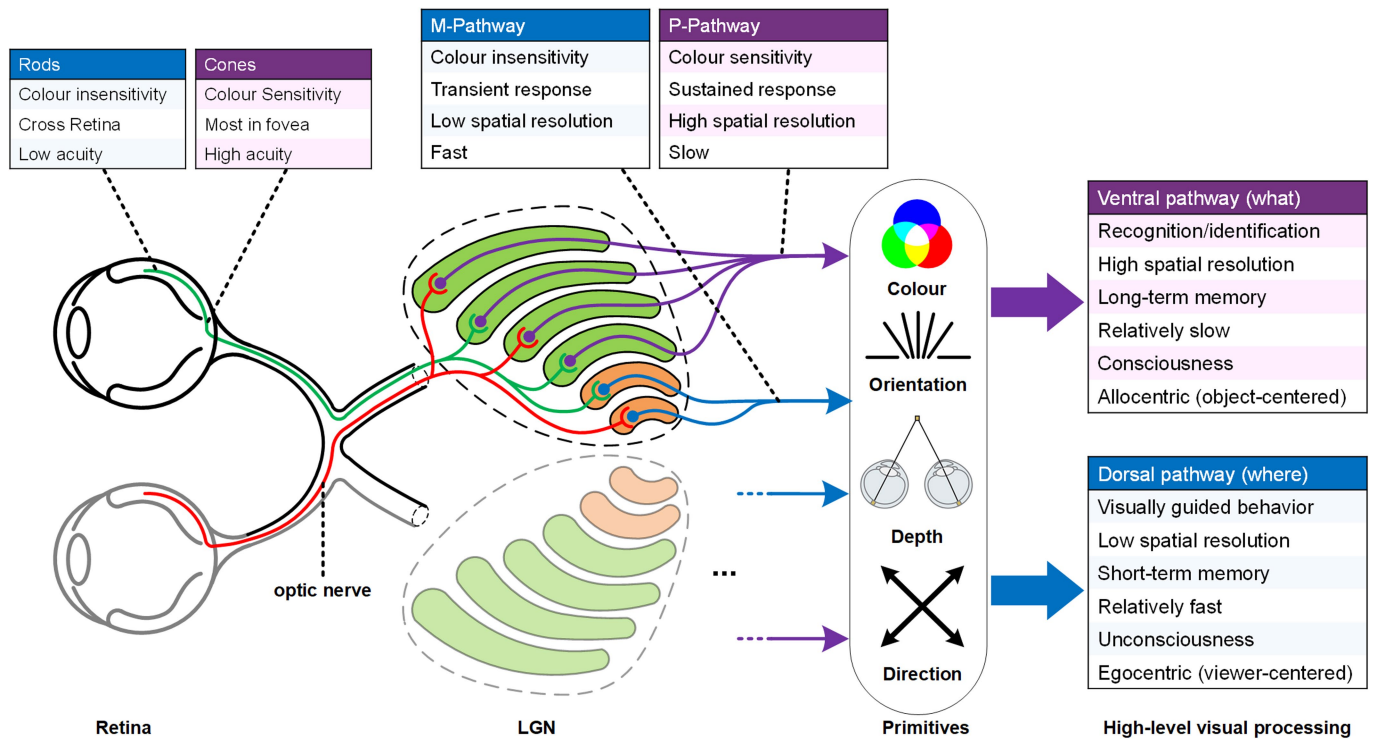
Extended Data Fig. 1 | The complementarity of the Human Vision System (HVS). The retina is composed of rod and cone cells that operate in an oppositional manner to expand the sensitivity range. At the next level, in the lateral geniculate nucleus (LGN), the M-pathway and P-pathway encode information in a complementary manner. The output information from the LGN is then reorganized into a series of primitives, including colour, orientation, depth, and direction at the V1 region. Finally, these primitives are transmitted separately to the ventral and dorsal pathways to facilitate the recognition of objects and visual-guided behavior.
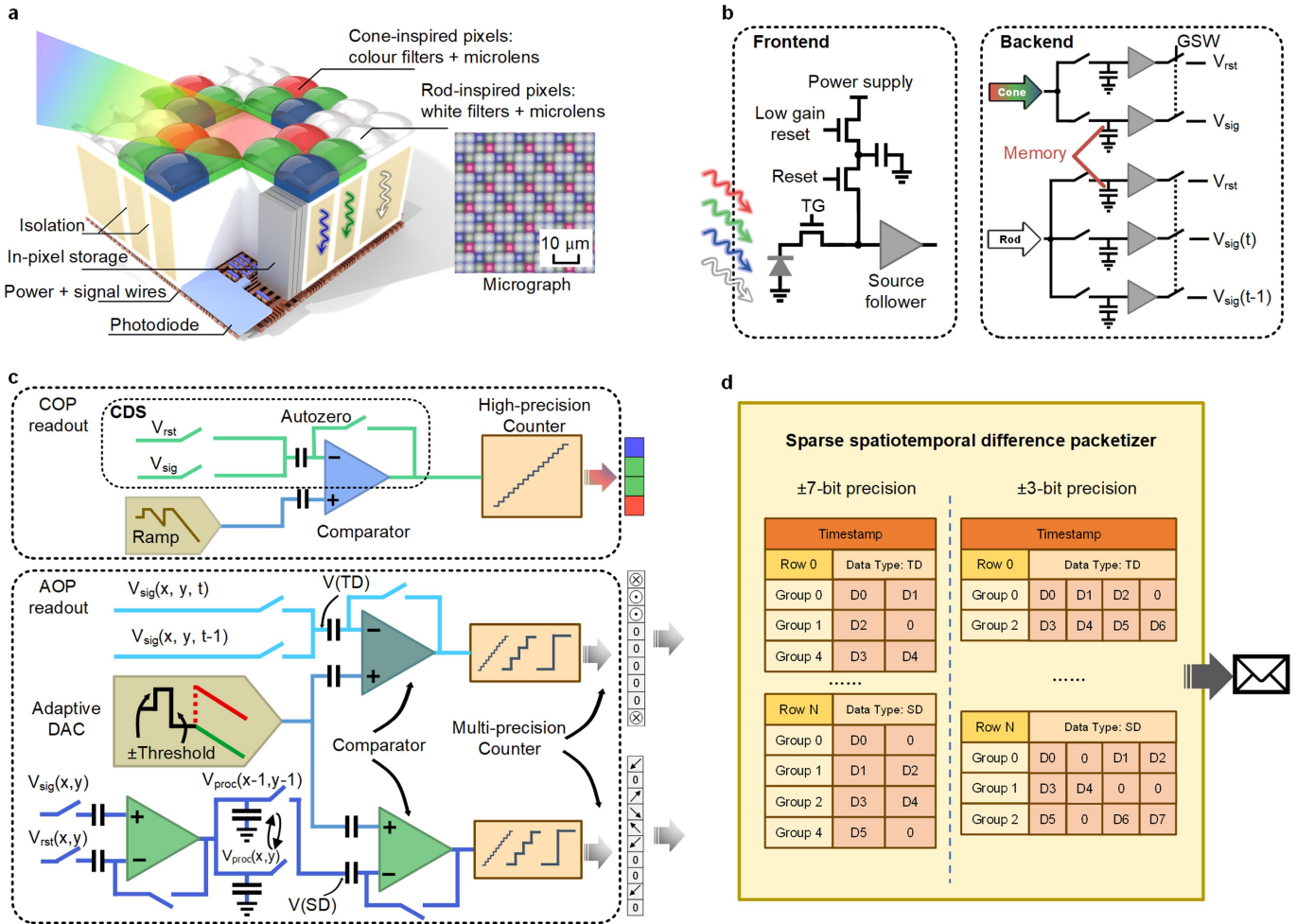
Extended Data Fig. 2 | Tianmouc architecture. a, Schematic of the pixel structure in the back-side illuminated hybrid pixel array. **b**, Schematic of the cone-inspired and rod-inspired pixels. **c**, Schematic of readout circuits of the COP and AOP. **d**, Schematic of compressed packets generation process through the sparse spatiotemporal difference packetizer.
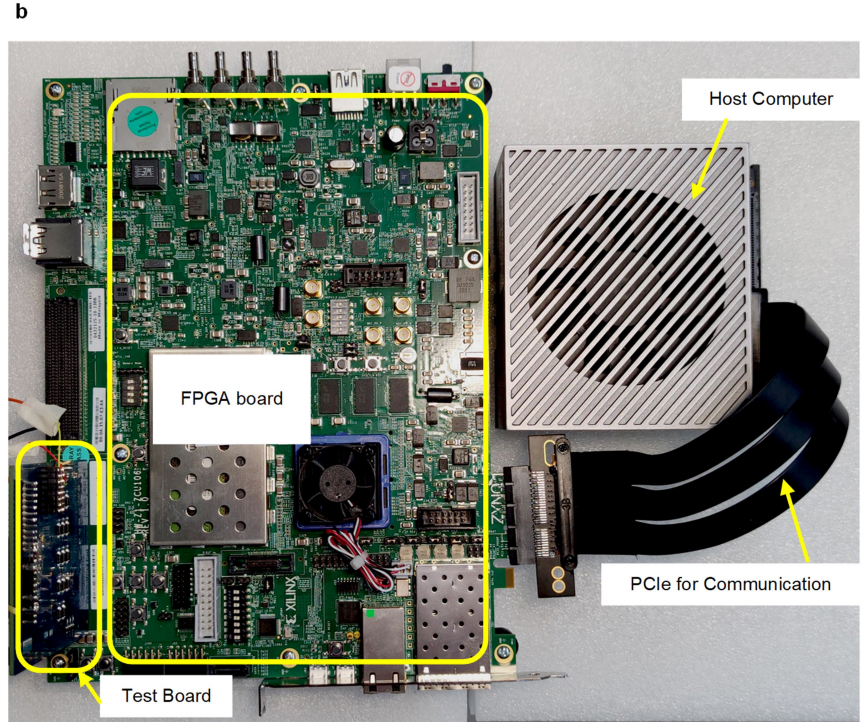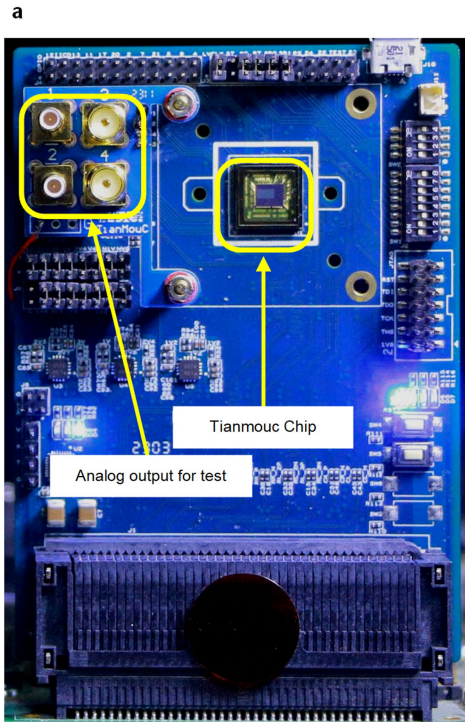
Extended Data Fig. 3 | Tianmouc chip testing systems. a, Testing boards equipped with a Tianmouc chip. **b**, The full system to process the output data of Tianmouc chip. The data is first transmitted to the FPGA board, where it collects raw data before transferring it to the host computer through PCIe. Subsequently, the host takes the charge of data processing for test and other tasks.
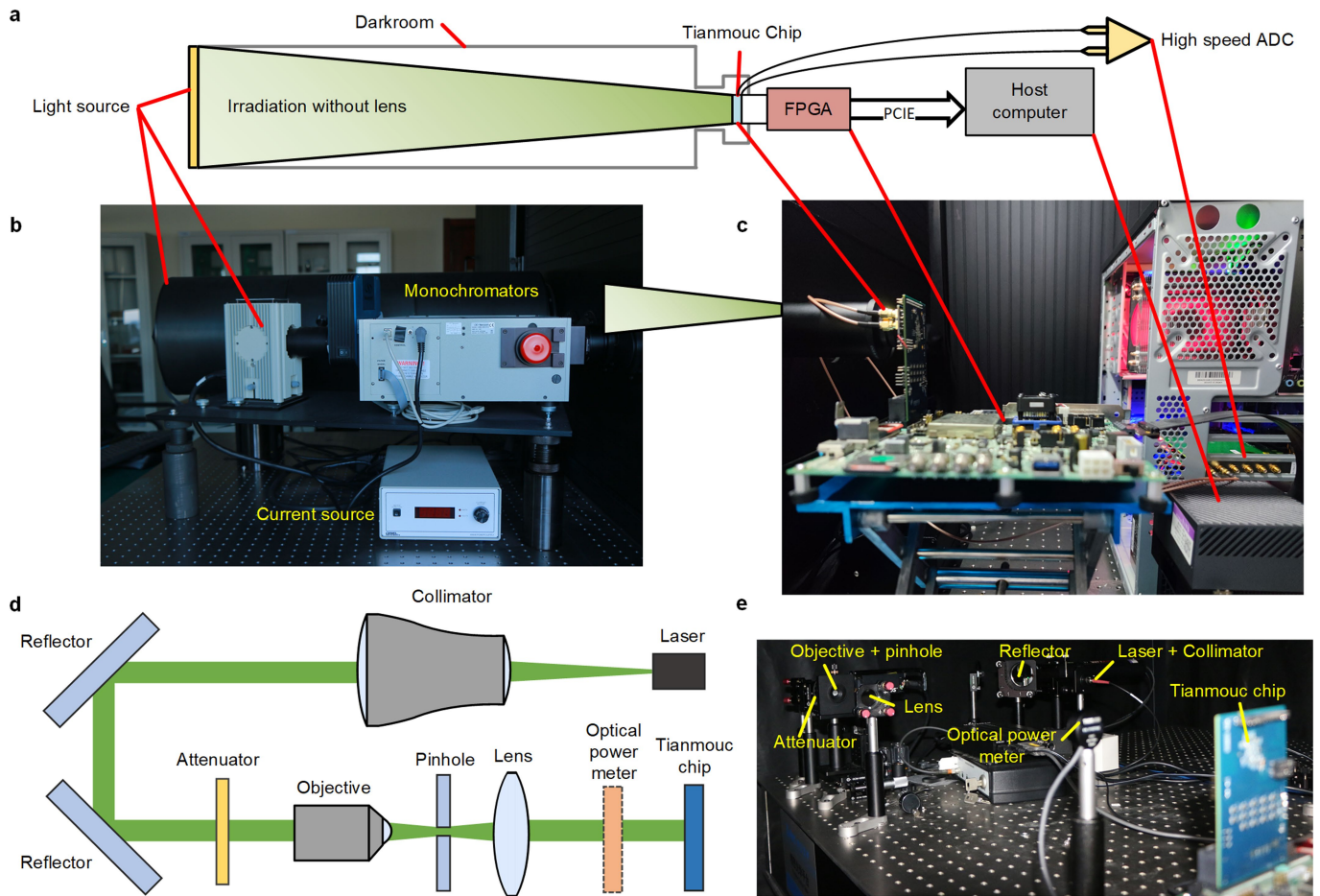
Extended Data Fig. 4 | Experimental setup for chip characterization.
a, Schematic illustration of the experimental set-up for the chip evaluation based on EMVA1288. **b**, A photograph of the optical setup. **c**, Photograph of the chip evaluation system including chip test board, FPGA board, host computer and the high-speed ADC acquisition card. **d**, Schematic illustration of the optical set-up for dynamic range measurement. **e**, A photograph of the optical setup for dynamic range measurement.
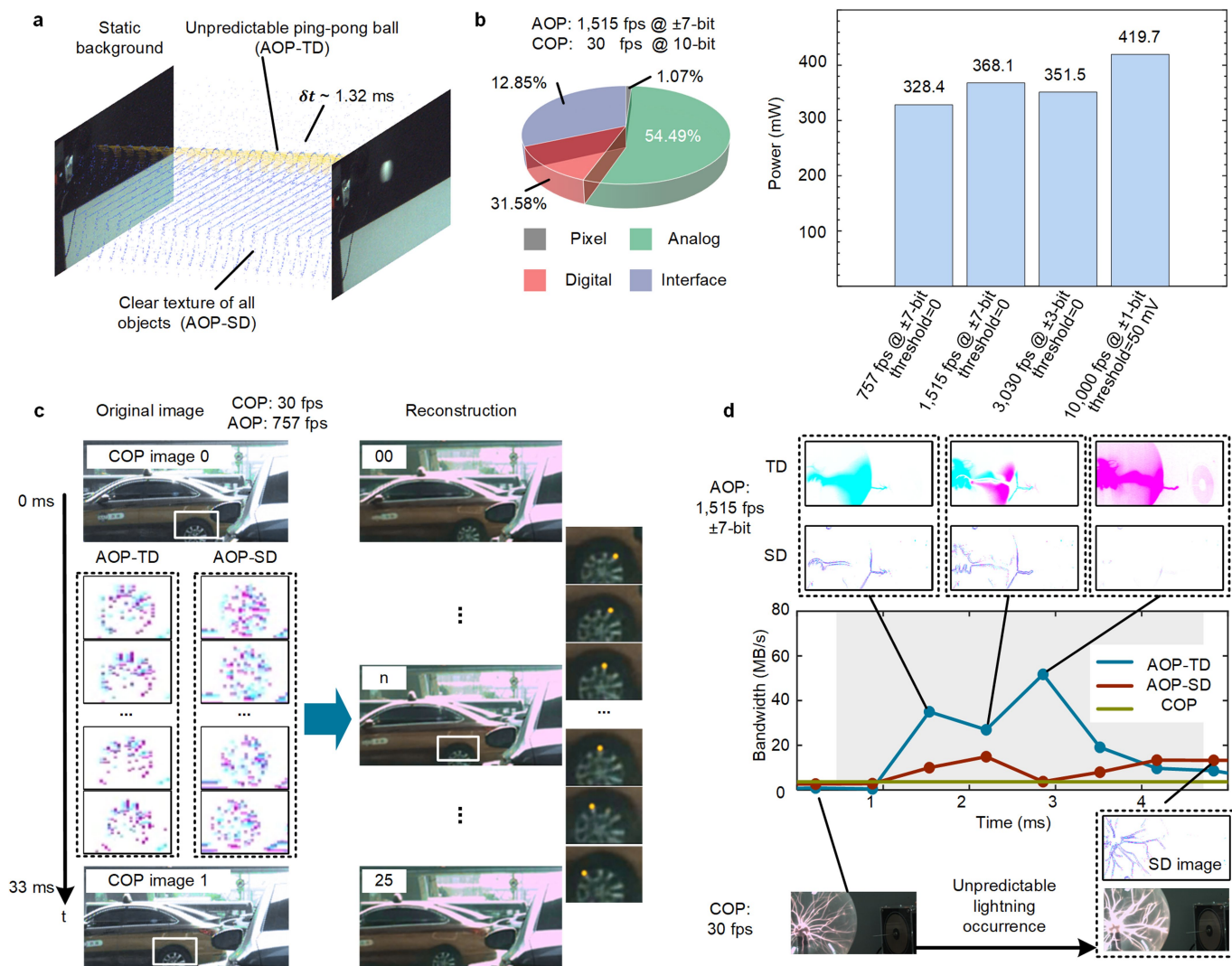
Extended Data Fig. 5 | Chip characterization. a, High-speed recording of an unpredictable and fast-moving ping-pong ball shot by a machine. **b**, Power consumption of Tianmouc. The left half depicts the distribution of different modules including pixel, analog, digital and interface circuits. The right illustrates the total power consumption under different modes. **c**, Anti-aliasing reconstruction of the rotation of a wheel. The alias in the wheel recorded by COP can be eliminated by the high-speed AOP. **d**, the AOP of Tianmouc is able to capture lightning that is missed by COP and record details of textures.
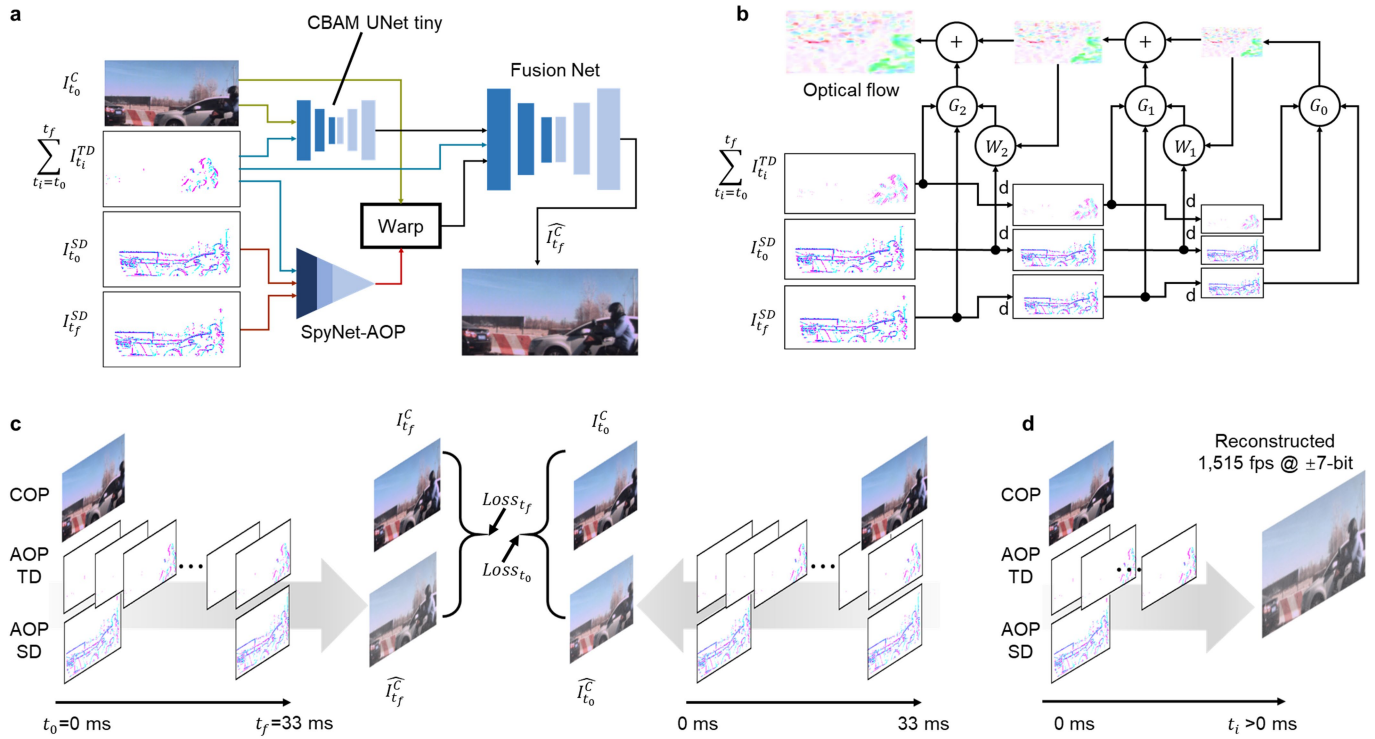
Extended Data Fig. 6 | The reconstruction pipeline. a, The structure of the whole reconstruction network. **b**, The light-weight optical flow estimator modified from SpyNet, using multi-scale residual flow calculation. In this figure, d means down-sampling operation. **c**, A self-supervised training pipeline, where we use the two colour images and the difference data between these two images to provide two training samples. **d**, At the inference stage, we adjust the amount of input data to obtain high-speed colour images at any time point.
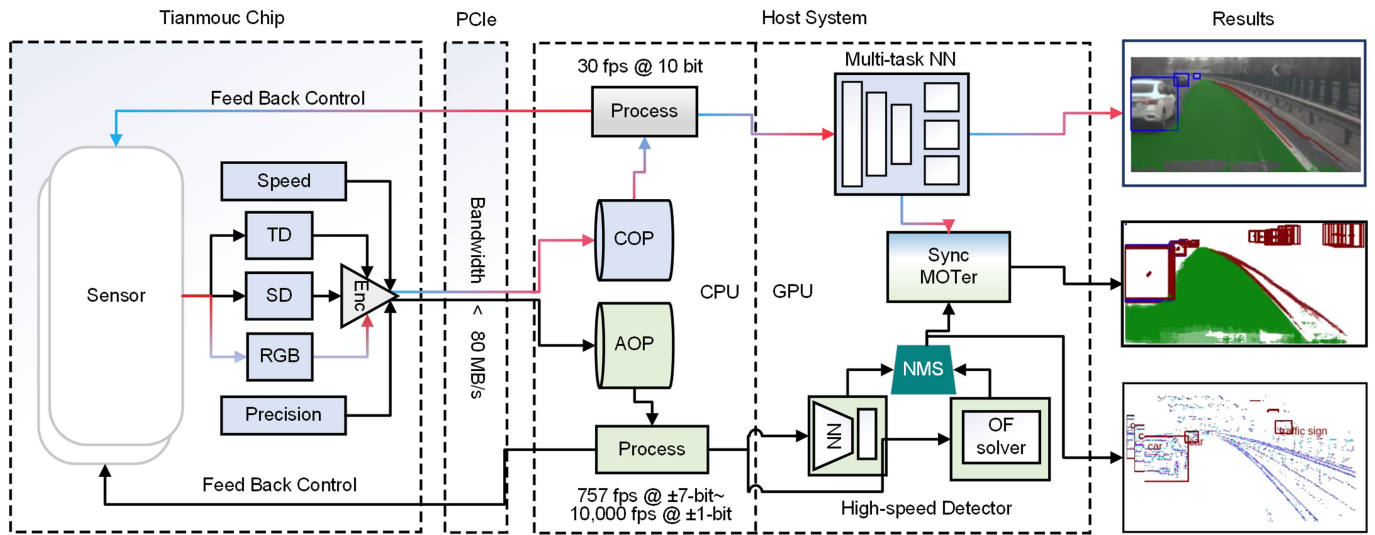
Extended Data Fig. 7 | The streaming perception pipelines for the open-world automotive driving tasks. In Tianmouc, different primitive combinations are encoded to form the AOP and COP. These two pathways maintain separate buffers and support independent feedback control. The processed data of the AOP and the COP are then sent to different NN or an optical flow solver. Subsequently, the inference results are integrated in a multi-object tracker. This approach optimally leverages the CVP at a semantic level, preserving both low-latency response ability and high performance simultaneously.
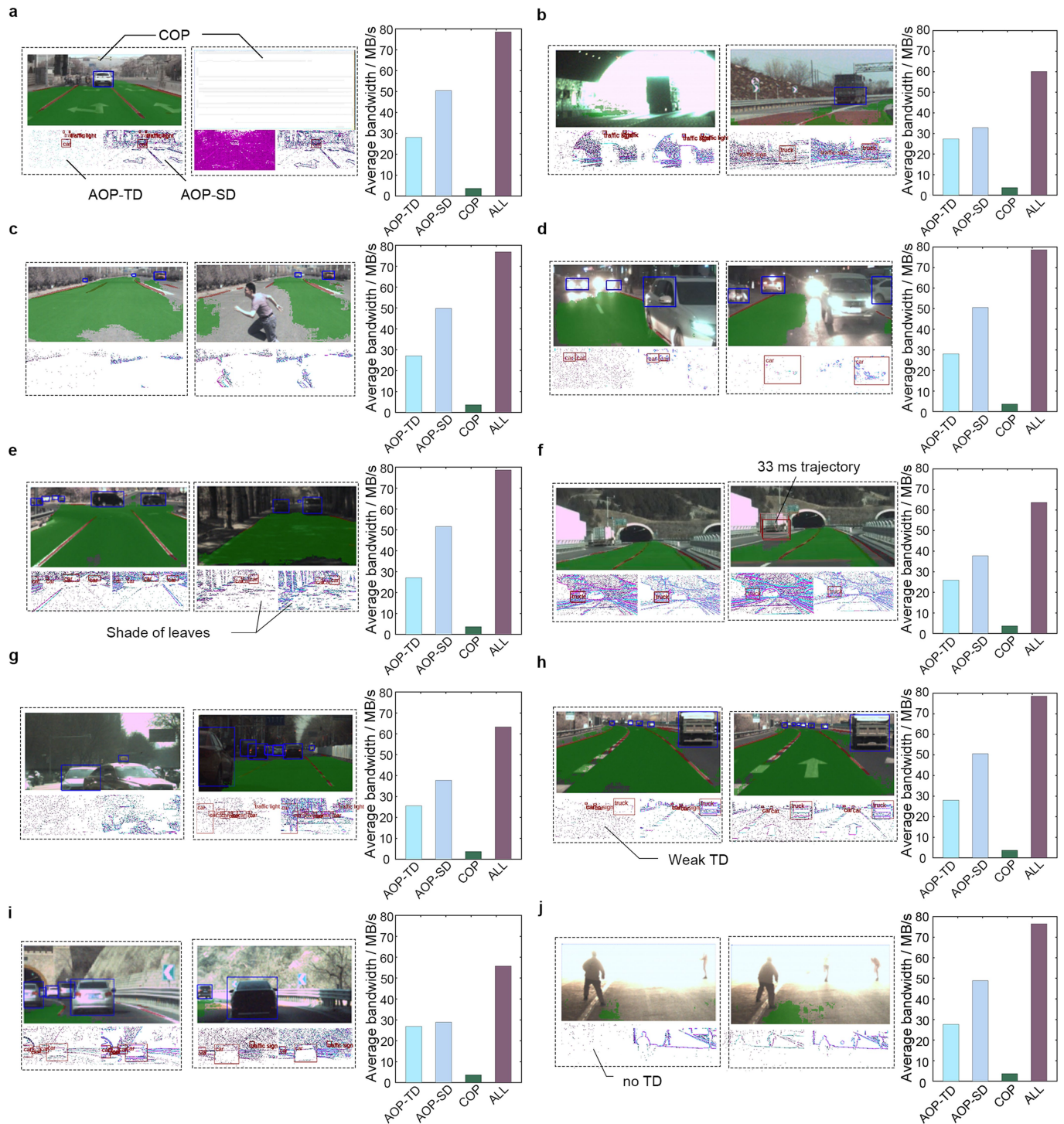
Extended Data Fig. 8 | More cases demonstrate the efficiency of Tianmouc in adapting to the open world. The sparse data in the AOP, coupled with the encoding method, enables Tianmouc to adaptively adjust its transmission bandwidth, typically maintaining it at a bandwidth below 80 MB/s in most scenarios. With the complementary perception paradigm, this bandwidth proves adequate for efficiently addressing diverse corner cases.

Extended Data Table 1 | The primitive-based representation and complementary sensing paradigm in Tianmouc

Primitives	Cognition-oriented pathway (COP)	Action-oriented pathway (AOP)
Response mode	Sustained response to intensity $g(x,y,t)$	Transient response to difference $\{\frac{\partial g}{\partial x}, \frac{\partial g}{\partial y}, \frac{\partial g}{\partial t}\}$
Wavelength	Colour $\{R, G, B\}$	Monochromatic
Precision	High	Low
Temporal resolution	Low	High
Spatial resolution	High	Low
Sensitivity	Low	High
Function	Object cognition	Vision guide behaviour

Different primitives are combined into specific data modalities, which are sampled and transmitted by the pathways.

Extended Data Table 2 | Comparison of Tianmouc with existing vision sensors

Platform	Tianmouc				DAVIS 346[49]		ISSCC'23[27]		ISSCC'23[28]		IMX636 ES[25]	JSSC'18[41]
Functionality	Intensity	TD and SD			Intensity	TD	Intensity	TD	Intensity	TD	TD	Intensity
Technology	90 nm				180 nm		40 nm + 65 nm + 40nm		90 nm + 22 nm		90 nm + 40 nm	90nm + 65nm
Pixel size (µm)	4	8			18.5		2.2	8.8	1.22	4.48	4.86	6.9
Resolution	320×320	TD: 160×160 SD: 159×160+159×159			346×260		4,096×3,680	1,032×928	7,680×4,336	1,920×1,084	1,280×720	1,632×896
Frame rate or latency	30 fps	757-1,515 fps	1,515-3,030 fps	10k fps	40 fps	20 µs*	18 fps	100 µs @1k Lux 200 µs @100 Lux	59 fps	10k fps	1,000 µs @5 Lux 220 µs @1k Lux	660 fps
Precision (bit)	10	±7	±3	±1	8	±1	N.A.	±1	N.A.	±1	±1	14
Maximum dynamic range (dB)	130				56.7	120	N.A.	N.A.	67.8	N.A.	86 @ 5 lux	70.2
Bandwidth (Gbps)	0.03	<1.4	<1.4	<2.3	0.028	>0.264†	N.A.	N.A.	N.A.	N.A.	<1.6	76
Power (mW)	328.4 ~ 419.7				140	30	N.A.	64	525		84	746
FOM (10^{12})	375.4				15.6		N.A.		N.A.		21.3	3.1

Notes:

* The minimum latency of DAVIS 346 is measured at its best conditions for a few pixels. The total latency of the sensor is collectively influenced by multiple factors, such as pixel bias configuration, illumination, and the event readout scheme[63].

† The bandwidth of DAVIS 346 is calculated by multiplying the maximum event rate and the bus width of each event including row and column address[64].