



METHOD

An Easily Compatible Eye-tracking System for Freely-moving Small Animals

Kang Huang^{1,2} · Qin Yang¹ · Yaning Han^{1,2} · Yulin Zhang¹ · Zhiyi Wang³ · Liping Wang^{1,2} · Pengfei Wei^{1,2}

Received: 24 May 2021 / Accepted: 3 December 2021 / Published online: 24 March 2022
© Center for Excellence in Brain Science and Intelligence Technology, Chinese Academy of Sciences 2022

Abstract Measuring eye movement is a fundamental approach in cognitive science as it provides a variety of insightful parameters that reflect brain states such as visual attention and emotions. Combining eye-tracking with multimodal neural recordings or manipulation techniques is beneficial for understanding the neural substrates of cognitive function. Many commercially-available and custom-built systems have been widely applied to awake, head-fixed small animals. However, the existing eye-tracking systems used in freely-moving animals are still limited in terms of their compatibility with other devices and of the algorithm used to detect eye movements. Here, we report a novel system that integrates a general-purpose, easily compatible eye-tracking hardware with a robust eye

feature-detection algorithm. With ultra-light hardware and a detachable design, the system allows for more implants to be added to the animal's exposed head and has a precise synchronization module to coordinate with other neural implants. Moreover, we systematically compared the performance of existing commonly-used pupil-detection approaches, and demonstrated that the proposed adaptive pupil feature-detection algorithm allows the analysis of more complex and dynamic eye-tracking data in freely-moving animals. Synchronized eye-tracking and electroencephalogram recordings, as well as algorithm validation under five noise conditions, suggested that our system is flexibly adaptable and can be combined with a wide range of neural manipulation and recording technologies.

Kang Huang, Qin Yang and Yaning Han have contributed equally to this work.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s12264-022-00834-9>.

✉ Pengfei Wei
pf.wei@siat.ac.cn

- ¹ Shenzhen Key Lab of Neuropsychiatric Modulation and Collaborative Innovation Center for Brain Science, Guangdong Provincial Key Laboratory of Brain Connectome and Behavior, CAS Center for Excellence in Brain Science and Intelligence Technology, Brain Cognition and Brain Disease Institute, Shenzhen-Hong Kong Institute of Brain Science-Shenzhen Fundamental Research Institutions, Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China
- ² University of Chinese Academy of Sciences, Beijing 100049, China
- ³ Harbin Institute of Technology Shenzhen, Shenzhen 518055, China

Keywords Eye-tracking · Freely-moving · Head-mounted device · Pupil detection · Adaptive Kalman filter

Introduction

In modern neuroscience, monitoring techniques are becoming increasingly crucial for gaining insight into the mechanisms governing the cognitive functions of the brain [1–3]. One of these technologies, eye-tracking, is used to obtain information about the eye in humans and animals [4–11], including fine and rich eye-movement features, such as saccades, nystagmus, and pupil dilation [12, 13]. These features are reliable and provide stable cognitive readouts as they are usually modulated by specific neural circuits [6]. Thus, eye-tracking has largely promoted various areas of research including the neural mechanisms of attention [14], social interaction [15], learning and memory [16, 17], emotions [18, 19], and arousal states [20, 21], as well as having applications in neuropsychiatric

disorders [22–24], childhood development [25], human-computer interactions [26], advertising [27], and driveability assessment [28]. Although commercial eye-tracking software and hardware systems are currently available for data acquisition and subsequent analysis in humans and non-human primates, for small animals, most custom-built systems can only be applied to awake, head-fixed animals [29]. Many behavioral experiments require small animals to freely move in specific arenas (e.g., the Y-maze-based working memory paradigm allows mice to explore different channels; the preference for open/closed arms on the elevated plus maze demonstrates anxiety levels). Besides, the rapid development and the diversity of neuroscience technologies require that researchers flexibly combine various techniques according to different experimental purposes. Therefore, there is still no flexible and easily-compatible system that incorporates and synchronizes other neural techniques for eye-tracking in freely-moving small animals.

Previous studies have used magnetic sensors to track eye movements to investigate the visual system of freely-behaving chickens and the sleep-wake cycle of mice by implanting a magnet beneath the conjunctiva [9, 30, 31]. However, magnetic sensors alone cannot record pupil dynamics. Mounting miniature cameras and optical components on an animal's head is the optimal strategy for freely-moving eye-tracking. To study the visual attention of great-tailed grackles in flight, Yorzinski [32] mounted a pair of cameras to record eye blinks. But the major limitation is that a device weighing 50 g cannot be used in mice. Wallace *et al.* [33] developed a head-mounted eye-tracking system, but it can only be used in rats due to limitations in size and weight. Meyer *et al.* [34] and Michael *et al.* [35] proposed a multi-functional system that combines behavioral monitoring, electrophysiological recording, and eye-tracking in freely-moving mice. Although the combination of these highly customized modules can play a powerful role in specific experiments, these modules that add extra weight to the system may not be needed in most other eye-tracking experiments. Neuroscience experiments commonly use electrodes, optical fibers, and microlenses implanted in the brain to apply electrophysiology, optogenetics, Ca^{2+} imaging, and miniature two-photon microscope manipulations or measurements to match brain activity in relation to external behavior. These implanted devices vary greatly in shape and size, and even for specific modalities, such as electrodes, there are differences among configurations. Modifying these systems to adapt to different implanted devices is challenging because it requires recalibration or rearrangement of the light source and other components on the device. Thus, there is a strong demand to develop a

general-purpose eye-tracking device that can easily integrate other neural implants.

To overcome these limitations, we developed an easily compatible eye-tracking system for freely-moving small animals to coordinate various neurological devices. First, our head-mounted eye-tracking unit (ETU) integrates ultra-micro components, except for the necessary image sensors and optical components; the peripheral circuits are moved to the remote end. We used a software approach to replace a portion of the hardware, which significantly reduced the device's weight and allowed animals to carry other devices. Second, the ETU's mounting structure fully exposed the top of the skull, facilitating other neural implants with accurate event synchronization. Finally, to accurately track the pupil dynamics, we compared the currently commonly-used pupil extraction methods and adopted complementary advantages. By designing an adaptive Kalman filter (AKF), we integrated the traditional image segmentation algorithm with a method based on machine learning to overcome the noise. We designed an AKF fusion algorithm based on DeepLabCut (DLC) [36] to overcome the noise caused by the large curved eyeballs and the close head-mounted light source, as well as a large eye movement range and increased blinking. These features allow the researcher to combine and synchronize neural recordings and manipulations to comprehensively interpret cognition.

Materials and Methods

Design and Fabrication of the Eye-Tracking Device

We designed a compatible device for recording eye movements in freely-moving small animals (Fig. 1A), which included three main parts (Fig. 1A, B): the ETU, ETU connector, and peripheral acquisition circuit.

ETU

The ETU has two design options (Fig. 1C). The first scheme (Sch1) is designed to be lightweight. The image sensor (ultra-micro camera) is aimed directly at the animal's eye, suitable for experiments with additional devices implanted on the head. Sch1 requires the camera to be aimed at the center of the pupil to obtain the most accurate measurement; however, it may obscure the central area of the animal's visual field. Thus, the second scheme (Sch2) also includes a reflected optical path. A near-infrared (NIR) total reflector reflects the image of the eye to the camera and transmits visible light, allowing the animal to observe the environment. Sch2 does not occlude the visual field, thus allowing the animal to undergo visual cognition experiments.

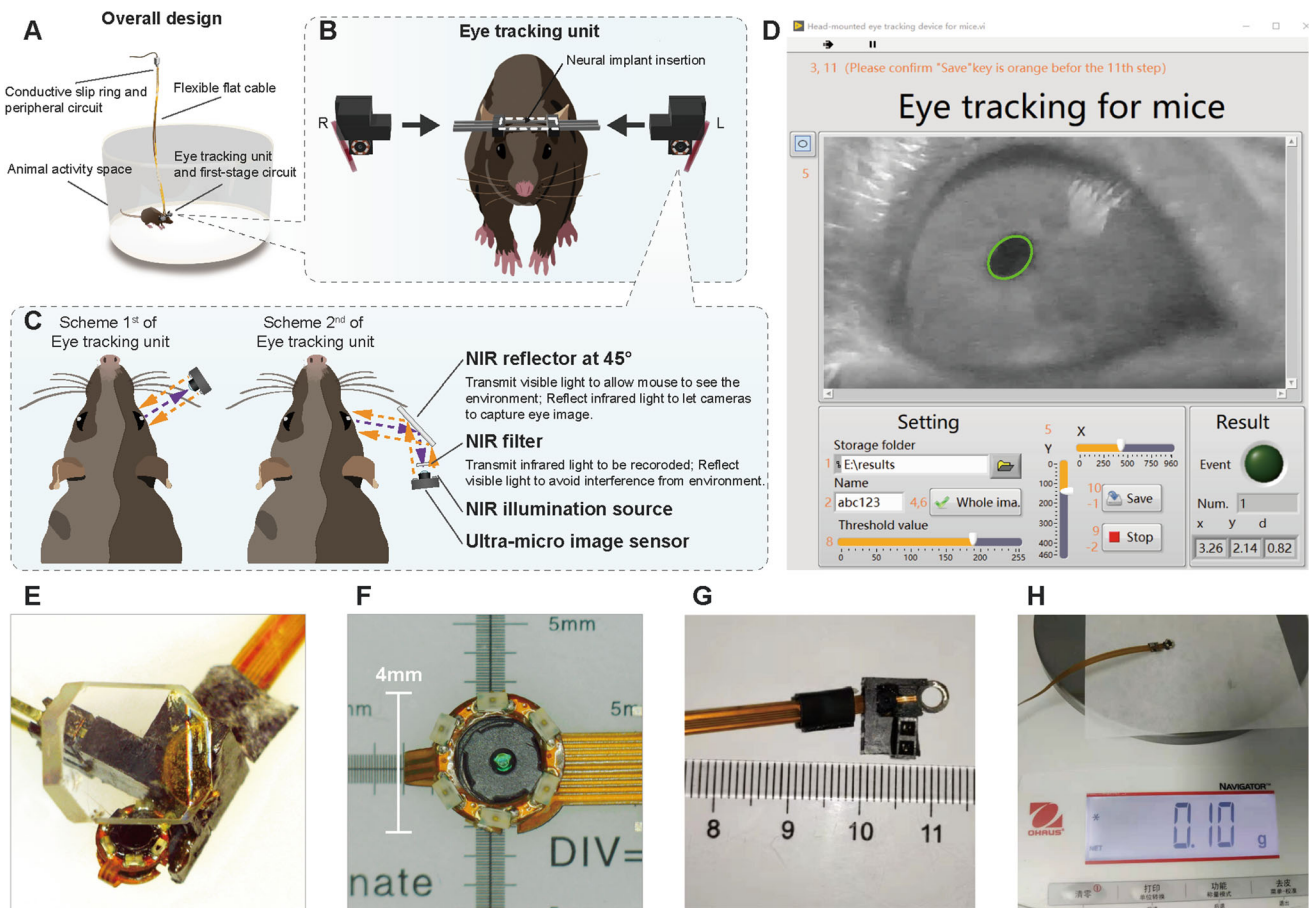


Fig. 1 The miniature eye-tracking system for freely-moving small animals. **A** An overview of the ETU fixed on a freely-moving mouse inside a circular open-field. The ETU is mounted on the animal's head. **B** The detachable design for mounting the ETU. The ETU connector comprising a male pin header and implanted slot is surgically mounted on the head in advance, and the left and right ETUs can be mounted to it detachably. The implant slot allows the head to be exposed to allow insertion of other neural recording or manipulation devices. **C** Two optional ETU schemes. **D** Screenshot of

online recording software with a user-friendly interface. The main panel shows a mouse eye image captured in real-time, and the display quality (exposure and ROI) can be adjusted by the user within an appropriate range. **E** Photograph of a single ETU: ultra-micro image sensor (camera), NIR illumination source, NIR filter, quartz NIR reflector, ETU base, and the first-stage acquisition circuit. **F–H** Photographs showing the ultra-micro image sensor (diameter: 4 mm), and the width (8 mm) and weight (0.10 g) of a single ETU (**C**).

The cameras in both schemes are based on the ultra-micro CMOS image sensor (resolution: 1280×720 ; OVM9724, OmniVision, China) equipped with a macro lens (Fig. 1E). Six NIR light-emitting diodes (LEDs, OmniVision, GTG1005IRC-940) integrated around the camera provide an infrared source, which is used to illuminate the animal's eye directly or through the NIR total reflector. To avoid interference from the environment, an NIR band-pass filter (940 nm, K9, Wuhan Huaxing Photoelectric Technology Co., Ltd), which transmits infrared light and filters visible light, is mounted in front of the lens. The assembled direction of the NIR total reflector in Sch2 is 45° from the central axis of both the eyeball and the camera. For high imaging quality, quartz is used as the material of the NIR total reflector in Sch2 (Fig. 1C, E), which accounts for the main weight. These

components are assembled on the base of a 2-pin female header made of insulating material (polyvinyl chloride).

ETU Connector

Regarding the mounting of the ETU to the animal's head, we applied a detachable design (Fig. 1B), including a male pin header (2.54-mm pitch) with both ends of the ETU connector horizontally fixed on the head. A square hole was made in the middle of the header as a slot for implanting optical fibers, electrode wires, or other neuronal recording/manipulation devices. The ETU is mounted only when it is needed to avoid hardware damage caused by the animals' chewing or fighting. This detachable design allows the experimenter to quickly plug or unplug the ETU and reuse it in multiple sessions to improve its

utilization. In addition, the ETU can be used in different small-animal species by only adapting the ETU connector without modifying the device itself. The position of the ETU connector may vary depending on the experimenter's surgical requirements and the brain region to be targeted. To include the eye in the field of view of the camera, only two pins of the 3-pin connector are connected, thus facilitating the adjustment of the ETU's position.

Acquisition Circuit

The acquisition circuit involves a two-stage design. The optical image of the eye is collected by the camera of the ETU and encoded as a digital signal *via* the first-stage circuit (frame rate: 30 fps) to improve the anti-noise performance during transmission. The encoded signal is transmitted through a flexible flat cable (50 cm) to the second-stage peripheral circuit and converted to a USB signal. A conductive slip ring (Jinpat Electronics Co., Ltd; LPMS-04A) between the peripheral circuit and the USB cable prevents knotting of the cable during the free movements of the animal.

This two-stage design offers two advantages. First, the additional load on the animal only includes the ETU and the first-stage circuit, which account for 1/7th of the entire acquisition circuit, leading to a 6/7th reduction in load on the animal. The reduced weight and volume are beneficial for the implantation of electrodes, optical fibers, micro-lenses, or other implants. Second, during image acquisition, the circuit produces heat. Thus, moving the main heating components to the peripheral circuit instead of placing them close to the body effectively protects the animal from thermal injury.

Real-Time Online Eye-capture and Pupil Detection

The real-time eye image-acquisition software was implemented based on LabVIEW (National Instruments, Texas, USA). The software obtains the camera data from the USB and displays the real-time image on the main panel. Before the recording starts, the experimenter can adjust the image quality (exposure, gain, and region of interest [ROI]) through the software to adapt to different illumination conditions. After the experiment is started, the researcher can monitor the eye reactions on the software and store the videos on the hard drive (raw image data is zoomed and cropped to 640×360 pixels in the eye region to save storage and computation consumption) with a specified file name. In addition to saving the video, the software can also detect pupils from real-time eye images. The pupil detection algorithm applies an ROI-based threshold segmentation method (ROI-Seg); thus, the experimenter needs to adjust the threshold and ROI manually before the

experiment starts. The detected pupil region is then subjected to ellipse fitting, and the major axis of the ellipse is set as the pupil diameter (Fig. 1D).

Pupil Detection with Deep Learning-based Methods

The eight deep learning-based methods involved belong to two categories, one aims at semantic image segmentation, which includes Fully Convolutional Network (FCN), Unet, ResNet18, ResNet50, InceptionResNetV2, MobileNetV2, and Xception, and the other (DLC) is based on landmark estimation. Both types require generating training sets to train models and estimate pupil size.

Training Set Generation

Each video recorded by the eye-tracking system was down-sampled to 150 frames with k-means clustering [36], thus a total of 600 frames to be labeled were extracted from the four videos. Two types of training sets were generated. For the first type (seven models for semantic segmentation), we used the MatLab (MathWorks, MA, USA) Image Labeler tool to manually label pixels of the pupil region, and the frames that were difficult to recognize were discarded. We then defined the category for each pixel value as “background” or “pupil” to construct semantic segmentation training sets. For the second type, we created a DLC (version: 2.1.8.2) project and configured nine pupil feature points to be detected: eight pupil landmarks including eight different angles and one pupil center point. To train the DLC model, we used DLC graphical user interfaces (GUIs) to manually label 600 selected eye images.

Models Training

Seven models for semantic segmentation were trained in MatLab. Specifically, five networks (ResNet18, ResNet50, InceptionResNetV2, MobileNetV2, and Xception) were created based on deeplabv3plusLayers [37]. The other two networks were based on fcnLayers [38] and unetLayers [39]. The DLC model was trained in the Python3 (www.python.org) environment. These training steps were executed on an NVIDIA GeForce RTX 3090 graphic processing unit. More training parameters are available in the Supplementary Materials.

Pupil Size Estimation

Each frame of the recorded eye videos was input to the pre-trained models for semantic image segmentation, the categorical labels of the image were the output. Then, the converted binary image was used to extract the largest object as pupil region and followed with morphological

operators. Finally, by measuring the properties of the pupil region, the major axis was set as the pupil diameter. For the DLC pre-trained model, the pupil feature points were detected from the frames of the recorded videos. Then, we filtered out landmarks with a likelihood <0.8 and calculated the distances between the retained landmarks and the pupil center as the pupil radii. Finally, the pupil radius was estimated as the average value of the radii. A micrometer was used to calibrate and calculate the conversion coefficient to convert the pupil data obtained in pixels to the actual pupil size in millimeters.

AKF Design

To measure the pupil size as accurately and robustly as possible in the eye-tracking experiment and to fully combine the complementary information of the two pupil measurement approaches (ROI-Seg and DLC), we designed an AKF fusion algorithm (Fig. 5C) to integrate the two types of pupil diameter data obtained separately (Fig. S1). The algorithm was implemented in two main stages: the pre-fusion and the fusion stage.

Pre-fusion Stage

At the pre-fusion stage, the pupil data obtained by the two independent approaches of ROI-Seg and DLC were processed using the same method. The purpose of this stage was to estimate the noise of the two independent pupil measurements, that will be used to calculate the Kalman fusion gain for the next fusion stage.

First, the conventional Kalman filter (KF) estimates the state and yields smooth data after eliminating extreme values. It includes two steps: prediction and update. The following equations are used for the prediction step:

$$r = \frac{\sum(z - \bar{z})^2}{n - 1} \quad (z \geq 0) \tag{1}$$

$$q = \frac{\sqrt{\frac{\sum(z - \bar{z})^2}{n - 1}}}{n} \tag{2}$$

$$\hat{x}_k = a\hat{x}_{k-1} \tag{3}$$

$$p_k = ap_{k-1}a + q \tag{4}$$

where r is the observation noise; q is the process noise; $z \in R^{n \times 1}$ is the observation (two separate pupil measurements); \hat{x}_k is the posteriori state estimation at current time k ; a is the state transition constant from time $k-1$ to k , which can be set to 1 (assuming that the current state is consistent with the previous state); and p_k is the posteriori estimated error of the state estimation and is computed recursively. To initialize the KF, \hat{x}_0 is set to z_0 , and p_0 is set to 1.

The following equations are used for the update step:

$$g_k = \frac{p_k}{p_k + r} \tag{5}$$

$$\hat{x}_k \leftarrow \hat{x}_k + g_k(z_k - \hat{x}_k) \tag{6}$$

$$p_k \leftarrow (1 - g_k)p_k \tag{7}$$

where g_k is the Kalman gain at current time k , as a tradeoff to balance the current observation with the prediction of the previous state.

Second, the covariance of the observation noise largely determines the performance of the KF fusion algorithm by affecting the fusion gain. Therefore, the observation noise is re-calculated dynamically by combining the estimated states with the observation values. By adaptively adjusting the fusion gain, the weight is inclined to a more reliable measurement source. Here, the equation for noise estimation is as follows:

$$e_k = \frac{n|\hat{x}_k - z_k|}{z_k + \varepsilon} \tag{8}$$

where ε is an infinitesimal value such that $z_k + \varepsilon > 0$.

To make the observation noise of the two measurements comparable, we normalized them as follows:

$$e_{norm} = 20 \log_{10} \left(\frac{e}{n} \right) \tag{9}$$

Fusion Stage

The fusion stage also includes two steps: update and prediction. The equations for the prediction step are the following:

$$R_k = \begin{bmatrix} e_{ROI-Seg_k} & 0 \\ 0 & e_{dlc_k} \end{bmatrix} \tag{10}$$

$$Q = \frac{1}{n(n-1)} \sqrt{\sum(z_{ROI-Seg} - \bar{z}_{ROI-Seg})^2 + \sum(z_{dlc} - \bar{z}_{dlc})^2} \tag{11}$$

$$\hat{X}_k = A\hat{X}_{k-1} \tag{12}$$

$$P_k = AP_{k-1}A^T + Q \tag{13}$$

where $z_{ROI-Seg}$ and z_{dlc} are the two separate pupil measurements obtained by the ROI-seg and DLC methods, and $e_{ROI-Seg_k}$ and e_{dlc_k} are their observation noise each estimated at the pre-fusion stage; R_k is the covariance matrix between each pair of measurements at current time k ; Q is the covariance of the process noise; \hat{X}_k is the posteriori state estimation at current time k ; A is the state transition constant from time $k-1$ to k , which can be also set to 1; and P_k is the posteriori covariance matrix of the estimated states at time k . To initialize, \hat{X}_0 is set to $z_{ROI-Seg_0}$, and P_0 is set to 1.

The equations for the update step are as follows:

$$C = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad (14)$$

$$Z_k = \begin{bmatrix} z_{ROI-Seg} \\ z_{dlc} \end{bmatrix} \quad (15)$$

$$G_k = \frac{P_k C^T}{C P_k C^T + R_k} \quad (16)$$

$$\hat{X}_k \leftarrow \hat{X}_k + G_k (Z_k - C \hat{X}_k) \quad (17)$$

$$P_k \leftarrow (1 - G_k C) P_k \quad (18)$$

where G_k is the Kalman fusion gain at current time k , which is an adaptive weight to tune the contribution of the measurement source to the final fused output X_{fuse} .

Performance Evaluation of Different Pupil Detection Methods

We used seven recorded videos and evaluated the accuracy according to the workflow shown in Fig. S2. We first screened for pupil data with high noise (good; noise < -30 dB) and low noise (poor; noise > -5 dB) by thresholding the estimated noise. According to the data quality of the ROI-Seg and DLC, the ground truth (GT) was generated with the following five conditions: (1) random selection; (2) both ROI-Seg and DLC are good; (3) only ROI-Seg is good; (4) only DLC is good; (5) both ROI-Seg and DLC are poor. A maximum of 150 GTs were generated under each condition. When labeling the GTs for a specific condition, the frames to be labeled were randomly selected from the videos that met the condition. For each of the seven recorded eye-tracking videos, the number of frames to be labeled was proportional to the recording duration. When the total number of eligible frames in one of the conditions was insufficient, the number of GTs was <150. After selecting the frames to be labeled under all conditions, we used a custom-written MatLab script to manually draw an ellipse along the pupil boundary. Frames that were difficult to recognize were discarded. Finally, the major axis of the ellipse in each frame was considered to be the pupil diameter. To measure the accuracy, the GTs' pupil diameter was subtracted from each pupil diameter obtained by different approaches, and we used the mean absolute error as the metric for accuracy evaluation.

We used a 5000-frame pupil video to estimate the pupil detection speed of the ten methods. According to the properties and implementations of these methods, the evaluation was performed in three environments: (1) ROI-Seg and AdaThresh were executed in a MatLab CPU environment; (2) seven semantic segmentation methods were executed in a MatLab GPU environment; (3) DLC was executed in a Python GPU environment.

Experiments

Animals

C57BL/6J mice (8–10 weeks old) were obtained from Hunan SJA Laboratory Animal Co., Ltd (Changsha, China). The mice were maintained in standard cages (22–25°C) with a 12-h light-dark cycle and were given food and water *ad libitum*. All experimental procedures were approved by the Animal Care and Use Committee at the Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences (Shenzhen, China).

ETU Mounting and Implantation of Electroencephalogram (EEG) Electrodes During Implantation Surgery

Mice were anesthetized with sodium pentobarbital (80 mg/kg) and placed on a stereotaxic apparatus (RWD, China); anesthesia was maintained with 1% isoflurane. After disinfection, ophthalmic ointment was applied to the eyes, the skin was incised to expose the skull, and the overlying connective tissue was removed. Then the ETU connector was attached directly to the skull with dental cement.

For EEG/electromyography (EMG) recordings, two stainless-steel screw electrodes were inserted into the skull, 1.5 mm from the midline and 1.5 mm anterior to the bregma, while two other electrodes were inserted 3 mm from the midline and 3.5 mm posterior to the bregma. Two EMG electrodes were inserted into the neck musculature. Insulated leads from the EEG and EMG electrodes were soldered to a pin header, which was secured to the skull using dental cement. After implanting the electrodes, the ETU connector was cemented onto the skull using dental acrylic. Eye-tracking experiments were conducted at least 1 week after the ETU connector was mounted to ensure adequate postoperative recovery. During experiments, the ETU was plugged into the header of the ETU connector.

Behavioral Tests

We conducted three behavioral tests. First, we tested the eye movements and obtained EEG synchronization recording ($n = 1$), which lasted for >1 h. In the second test, we collected eye movement data to evaluate the performance of the AKF algorithm. Both of these experiments were conducted in an acrylic 40 cm × 40 cm × 30 cm square open-field arena with a shelter nest in a corner outside and a door at the bottom of one side. The day before the experiment, the ETU was mounted on the head, and the mouse was habituated to the arena for 10 min. On the day

of the experiment, the mouse was habituated to the arena for another 5 min.

The third test was designed to test the effect of the ETU on mouse movement. The test was performed in two sessions: a mouse without ETU *versus* with ETU. In each session, the mouse was permitted to move freely in a square open-field for ~ 10 min. We used Bonsai [40] (bonsai-rx.org) to track the mouse; customized MatLab scripts were used to visualize and analyze behavioral data.

EEG/EMG Recording

EEG/EMG data were collected as previously described [41]. Briefly, a pre-amplifier unit (Part #8406-SL, Pinnacle Technology, Inc., Lawrence, KS) that was rigidly attached to the head-mount, provided the first-stage amplification ($100\times$) and initial high-pass filtering (first order: 0.5 Hz for EEG and 10 Hz for EMG). These signals were then routed to an 8401 conditioning/acquisition system (Pinnacle Technology, Inc.) *via* a tether and low-torque commutator (Part #8408, Pinnacle Technology, Inc.). The 8401 amplifier/conditioning unit provided an additional $50\times$ signal amplification, additional high-pass filtering, and 8th-order elliptic low-pass filtering (50 Hz for EEG; 200 Hz for EMG). The signals were then sampled at 400 Hz, digitized using a 14-bit A/D converter, and routed *via* USB to be processed by a PC-based acquisition and analysis software package.

Pupil-EEG Correlation Analysis

We first applied band-pass filtering to the raw EEG signal in different frequency bands (delta, 0.5–4 Hz; theta, 4–8 Hz; alpha, 8–12 Hz; beta, 12–30 Hz; gamma, 30–80 Hz). Thereafter, the power was calculated as the squared value of each filtered signal. To analyze the correlation between pupil dilation and EEG power, the synchronized two types of data were binned with a 5-s window. Finally, Pearson's correlation analysis was applied to five pairs of data.

Event Synchronization Module and Precision Validation

We designed a flexible event synchronization module to synchronize eye-tracking data with other neural recordings or external events. This module can be triggered by analog, digital, and optical signals; and a microcontroller (Arduino UNO R3; Arduino, www.arduino.cc) is used to convert these signals into event commands. The online software then captures the event trigger signals and records them as timestamps. The analog trigger channel receives a 5-V signal generated by a waveform generator or other device

or uses a photoresistor (PT908-7C; Everlight Electronics., Ltd) to convert a visual cue into a voltage signal. The digital trigger channel receives a transistor-transistor logic (TTL) signal sent by an external micro-control unit. In our synchronous EEG and eye-movement data recording, the digital TTL trigger was used. The event synchronization module of the eye-tracking system and the EEG amplifier receive the recording starting trigger simultaneously, and these two systems independently record the timestamp for subsequent alignment of data.

The precision of event synchronization refers to the time delay between the event marker timestamps and the laser-on event captured by the eye-tracking camera. To validate the precision of the event synchronization module, we applied the test strategy shown in Fig. 2D: (1) produce the events by laser; (2) obtain the actual events and event markers through two independent approaches (synchronization module and ETU camera); and (3) compare the time delay between actual laser-on events and event markers. Since capturing the eye image and laser-on event use the same camera, these events are considered synchronized. In addition, compared with analog and digital triggers, the delay of the optical signal trigger is the largest; therefore, we can test the limit delay of our system. We first used the ETU's 30-fps camera to capture the laser brightness. Since the 30-fps camera is relatively slow in laser capture, to assess whether the synchronization module can achieve millisecond accuracy, we tested it with a 240-fps camera. In both tests, laser-on events were repeated for ten trials in a 5-s cycle. A customized MatLab script was used to extract the laser-on event timestamps (threshold set to 1/10 of the maximum brightness, and the laser-on event is considered to start when the brightness exceeds the threshold).

Data Analysis

Pupil data extracted by different approaches were imported into MatLab. The AKF fusion algorithm and subsequent visualizations were implemented using custom-written MatLab functions and scripts.

The data used for the accuracy evaluation of AKF are expressed as the mean \pm SD. Statistics were analyzed using Prism 8.0 (GraphPad Software, Inc.). Before hypothesis testing, the data were first tested for normality using the Shapiro-Wilk normality test and for homoscedasticity using the *F* test. One-way analysis of variance with the Kruskal-Wallis test was applied to determine whether the AKF fusion algorithm significantly improved the pupil measurements.

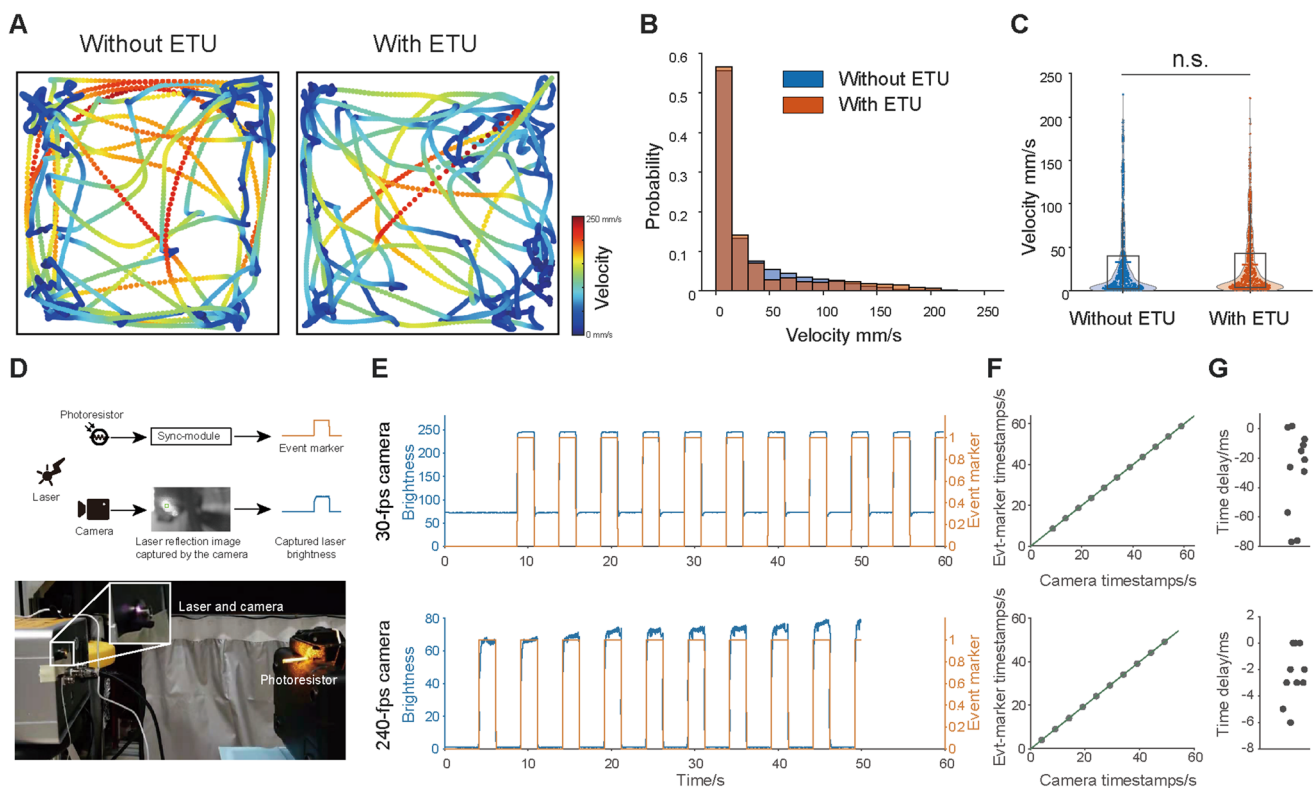


Fig. 2 Effects on animal movement and validation of the precision of event synchronization. **A** Mouse movement trajectories and velocities of the two test sessions (without ETU *versus* with ETU). In both tests, the mice moved freely in a square open-field for ~10 min. **B** Histogram showing consistent velocity distributions between the two tests. **C** Velocity distribution in the two test sessions. The dot on each violin plot represents the average velocity in a 1-s bin. Bin numbers: Without ETU, $n = 614$; With ETU, $n = 590$; $P = 0.0544$, Mann-Whitney test. **D** Diagram (upper) and photograph (lower) of the event synchronization temporal accuracy test. An additional

240-fps camera is used to test whether the event synchronization module can achieve millisecond accuracy. **E–G** Comparisons of laser-on events with their event markers (upper, test by 30-fps camera; lower, test by 240-fps camera). **E** Laser-on events repeated for 10 trials in a 5-s cycle. Blue traces, laser brightness value captured by the camera in the ROI; orange traces, the states of event markers (0: baseline, 1: laser on). **F** Comparison of the camera-captured laser-on events with event marker timestamps. The points around the green diagonal indicate a low time delay. **G** Distribution of time delay.

Results

A Compatible Eye Tracking Device Combined with Additional Neural Implant for Use in Freely-Moving Animals

To combine eye tracking with various neural implant devices for recording and manipulating unconstrained small animals, we designed and implemented a novel eye-tracking device that is compatible with such neural devices. Because the maximum load on a mouse's head is ~3 g, combining eye-movement recording with other devices in free-moving mice requires reducing the weight of the ETU as much as possible. We measured the physical parameters of the ETU with its connector (Fig. 1F–H). The total width and total weight a single Sch1 ETU were 8 mm and 0.10 g, respectively, which applied to most adult mice. We applied a detachable design for ETU mounting, allowing the ETU to be mounted immediately before an

experiment. During postoperative recovery or in the home-cage, the mouse only needed to have the ETU connector mounted, which significantly increased the utility of the ETU. Although the ultra-micro image sensor we used is smaller than conventionally-used sensors, it has better imaging quality and provides a resolution of 1280×720 pixels.

We also developed a user-friendly GUI software to facilitate setting up, monitoring, and recording the eye-tracking data (Fig. 1D). The software allowed real-time pupil detection, and the detected pupil features, along with the synchronization event information of the neural recording or manipulation, could be saved.

Performance Validation and Synchronized EEG–Eye Movement Recording

Two essential factors determined the compatibility and coordination of the eye-tracking system with other neural

implants. First, the animal's behavior should not be significantly affected after the ETU is mounted. Second, the system accurately records the timestamps of external events. Therefore, we first assessed the effect of the ETU on mouse behavior. We visualized the velocity and trajectory heatmaps of two behavioral sessions (without ETU *versus* with ETU, Fig. 2A) and compared their velocity distribution (Fig. 2B, 2C). We found that the histograms of the velocity distribution of the two sessions closely overlapped. We also found no significant differences between the binned velocities (1-s time window) of the two tests (without ETU: 33.1 ± 39.0 mm/s; with ETU: 30.1 ± 47.7 mm/s, $P = 0.0544$, mean \pm SD).

We then tested the temporal precision of the event synchronization module (Fig. 2D). Besides the 30-fps camera (ultra-micro camera of the ETU), an additional 240-fps camera was used to test the limit delay of our system. We compared the laser brightness with the event marker. With the laser on, the brightness of the ROI of the laser in the captured video increased immediately, and the event marker was recorded instantly (video 2). The recorded traces of the laser brightness and event marker status were also well aligned in both the 30-fps and 240-fps tests (Fig. 2E). We extracted the event timestamps from the cameras that captured laser brightness (Fig. 2F) and calculated the time delays between the camera timestamps and captures (Fig. 2G). The delays of the 30-fps and 240-fps were 29.3 ± 8.5 ms and 2.4 ± 0.6 ms, respectively. Since there was a delay of $\sim 1/30$ s for the 30-fps camera, we demonstrated that our system can achieve a precision of ~ 2.4 ms through the 240-fps test. While most experiments accept a 1/33-s delay, the ETU offers a camera with a higher frame rate for experiments requiring greater time accuracy.

Finally, to test whether our device was robustly compatible with other neural implants in freely-moving mice during eye-movement experiments, we first mounted the ETU connectors on the mice, and the implant slot allowed us to expose the skull. Thereafter, the EEG electrodes were implanted through this slot (Fig. 3A). The same ETU was used repeatedly for recording from all tested mice. The simultaneous recording lasted >1 h, and the mice moved freely in the square open-field behavioral chamber (Fig. 3B), without any attached cable or fiber winding. The collected synchronized data and preliminary analysis (Fig. 3C, D) showed that pupil dilation strongly correlated with the alpha (8–12 Hz) and beta (12–30 Hz) frequency bands of EEG. These results suggest that our system can be used to study the neural activity mechanisms of brain state in freely-moving animals [42].

Comparison of Performance of Different Pupil Detection Methods

To accurately and efficiently extract pupil dynamics from the recorded eye video, we applied ten commonly-used pupil-extraction methods based on image processing and deep learning (Fig. 4A). These methods include two traditional image segmentation methods ROI-Seg and adaptive threshold segmentation (AdaThresh), and seven deep neural network-based semantic segmentation methods: FCN, Unet, ResNet18, ResNet50, InceptionResNetV2, MobileNetV2, Xception, and DLC-based pupil landmark estimation.

The performance comparison included two aspects. First, we evaluated the accuracy of these methods by comparing the extracted pupil diameter with the manually-labeled GT (randomly selected). Among these methods (Fig. 4B and Table 1), DLC had the best performance with an average error of 3.84 ± 5.75 pixels (mean \pm SD). The two lightweight networks Xception (21.36 ± 12.41 pixels) and MobileNetV2 (20.51 ± 15.92 pixels) had the worst performance with the same number of training sets. For the two traditional image-processing methods, although AdaThresh was more accurate than ROI-Seg in some cases, the overall error was larger than ROI-Seg due to mis-detection of the eyelid as the pupil. Then, we compared the detection speed of the different methods (Fig. 4B and Table 2). Since ROI-Seg used real-time online detection, it performed fastest with an average frame rate >1000 fps. AdaThresh used local first-order statistics to determine thresholds [43], and was slower than ROI-Seg (154 fps on average). In the deep learning-based methods, DLC performed best, while the other methods were too slow to achieve real-time processing of 30 fps eye video.

The image-processing method depends on the image features, while deep learning methods rely on human knowledge. Therefore, to fully test these methods, we used ROI-seg and DLC as representatives, and accuracy was further compared under five noise conditions. The results (Fig. 4C and Table 1) showed that the pupil regions estimated by ResNet18, ResNet50, InceptionResNetV2, MobileNetV2, and Xception were relatively larger than GT, while FCN and Unet were relatively smaller. The errors of ROI-Seg, AdaThresh, and DLC were evenly distributed around zero. These results showed that with the same number of training sets, although most deep learning-based methods located the pupil region correctly, it was difficult to accurately segment the pupil boundary. In addition, due to the large eye movements and complex illumination in the freely-moving eye tracking, ROI-Seg had a higher average error. But when the pupil was in the range of the ROI, ROI-Seg was the most accurate method to detect the pupil boundary.

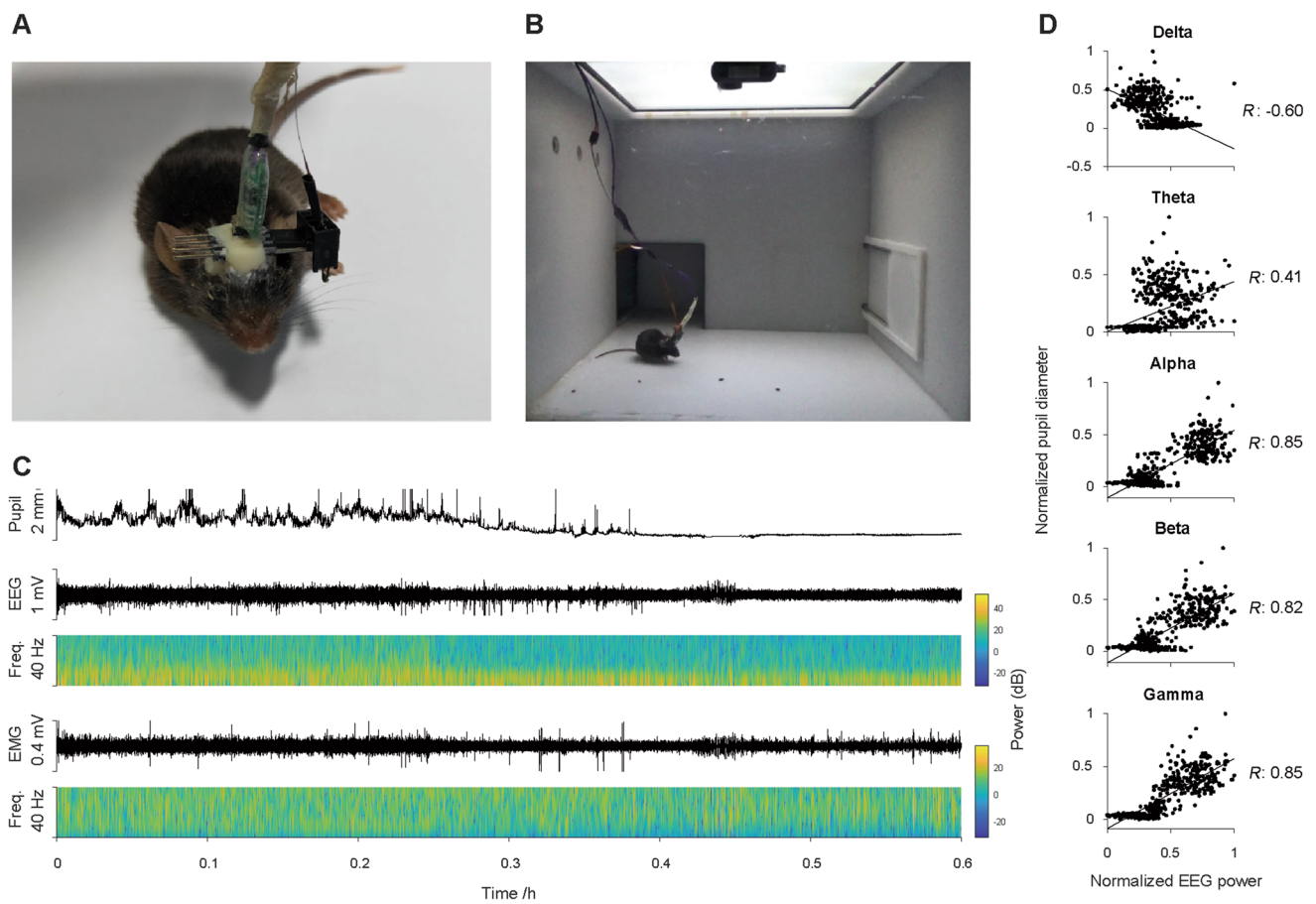


Fig. 3 Simultaneous recording of eye movements and electroencephalogram (EEG) in a freely-moving mouse. **A** ETU and the combined EEG head-stage mounted on the head. **B** Synchronized eye-tracking and EEG recording from a freely-moving mouse in a square behavioral chamber. **C** Example of synchronized data collected over

0.6 h showing the extracted pupil diameter, along with EEG and EMG raw signals and their spectrograms. **D** Representative plots of Pearson's correlation between pupil diameter and EEG power in each frequency band. Dashed line in each plot is the linear fit.

Table 1 Accuracy evaluation for ten different pupil detection methods (Error in pixels, Mean \pm SD at 640 \times 360 pixels resolution).

Conditions						
Methods	Random	Both good	ROI-Seg good & DLC poor	DLC good & ROI-Seg Poor	Both poor	Average
ROI-Seg	10.49 \pm 13.59	4.86 \pm 5.99	12.49 \pm 9.61	42.69 \pm 21.00	43.01 \pm 30.48	22.66 \pm 25.18
AdaThresh	14.06 \pm 21.85	8.54 \pm 15.68	21.38 \pm 20.66	14.18 \pm 21.58	37.06 \pm 31.01	18.13 \pm 24.68
FCN	10.93 \pm 12.87	8.09 \pm 8.36	32.48 \pm 30.21	20.12 \pm 19.74	31.21 \pm 27.67	18.57 \pm 21.95
Unet	18.59 \pm 21.56	13.51 \pm 14.05	32.71 \pm 25.57	32.08 \pm 26.05	32.41 \pm 31.12	24.45 \pm 25.10
ResNet18	14.79 \pm 13.53	13.27 \pm 11.00	15.62 \pm 13.00	17.00 \pm 16.96	24.03 \pm 20.00	16.83 \pm 15.65
ResNet50	13.34 \pm 11.18	11.98 \pm 8.74	12.75 \pm 11.67	11.37 \pm 11.25	25.83 \pm 23.84	15.05 \pm 15.26
InceptionResNetV2	16.81 \pm 17.04	13.81 \pm 10.84	20.10 \pm 15.59	19.09 \pm 17.58	26.19 \pm 21.00	18.78 \pm 17.08
MobileNetV2	20.51 \pm 15.92	17.45 \pm 13.37	20.07 \pm 17.10	17.66 \pm 17.16	24.84 \pm 21.80	19.95 \pm 17.25
Xception	21.36 \pm 12.41	19.95 \pm 12.45	22.99 \pm 16.22	20.50 \pm 15.24	25.38 \pm 20.48	21.81 \pm 15.45
DLC	3.84 \pm 5.75	2.73 \pm 3.34	17.30 \pm 30.75	4.55 \pm 5.90	13.19 \pm 31.88	7.05 \pm 18.70
Fuse	4.10 \pm 4.83	3.93 \pm 5.26	11.50 \pm 9.47	4.34 \pm 4.85	9.65 \pm 14.85	6.03 \pm 8.92

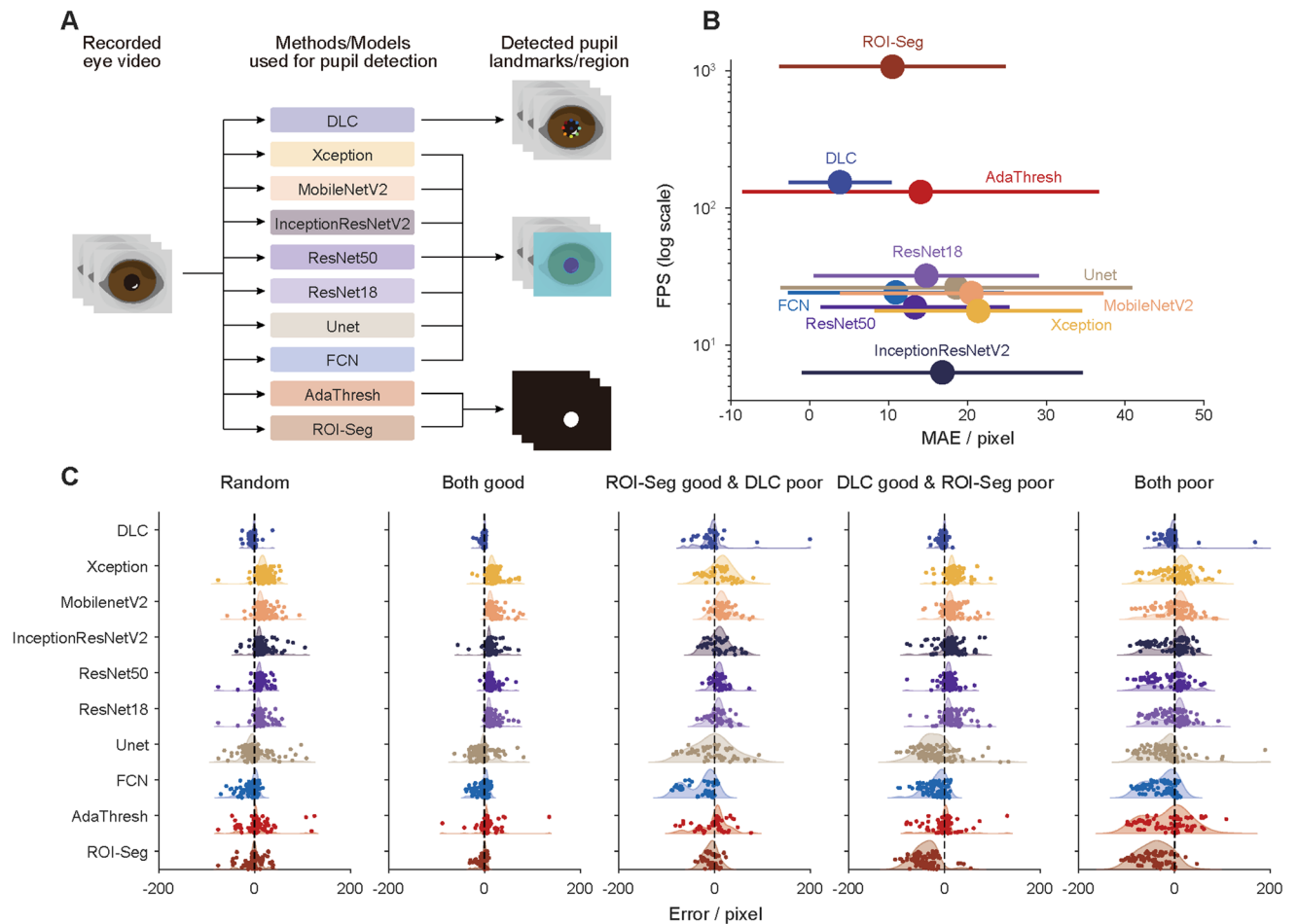


Fig. 4 Performance of different pupil-detection methods. **A** Diagram showing the use of ten different methods to detect the pupil region. The two traditional image segmentation methods output the binarized image of the segmented pupil region; seven deep neural network-based semantic segmentation methods output a categorical label image to indicate the pupil region; the DLC method outputs pupil region landmarks. **B** Pupil detection error and speed in the various models. The speed was estimated by recording the processing time of a 5000-frame pupil video. fps: frames per second. Error bars are presented as the mean \pm standard deviation (SD) of mean absolute error. GTs were randomly selected (117 GT samples).

C Evaluation of different methods under five different noise conditions based on ROI-Seg and DLC performance: (1) randomly-selected (117 samples, corresponding to the Random column in Table 1); (2) samples obtained by both approaches have a low noise level (137 samples); (3) only samples obtained by DLC have high noise (57 samples); (4) only samples obtained by ROI-Seg have high noise (112 samples); (5) samples obtained by both approaches have a high noise level (108 samples). The dots represent the error for each detection under specific methods and conditions. Area charts represent error distributions.

Table 2 Average pupil detection speed of various methods (fps at 640×360 pixels resolution).

Methods	ROI-Seg	AdaThresh	FCN	Unet	ResNet18	ResNet50	InceptionResNetV2	MobileNetV2	Xception	DLC
Speed	1075.68	131.48	24.40	26.52	32.31	18.99	6.35	24.09	17.88	154.79

The AKF-Based Algorithm Improves the Measurement of Pupil Size

The systematic comparison of pupil extraction methods indicated that DLC had the best performance balance in accuracy and speed, while real-time online ROI-Seg extracted the pupil more accurately under stable illumination and weak pupil shaking. Therefore, we proposed an

AKF algorithm (Fig. 5A–C) that efficiently fused multiple pupil data sources obtained from the DLC and ROI-Seg approaches, thus improving the tracking quality of pupil data in freely-moving mice (Fig. S1).

The conventional Kalman fusion algorithm statically estimates the measurement noise from a global perspective [44]. However, the performance of the KF is sensitive to noise estimation [45–47]. If the measurement noise was

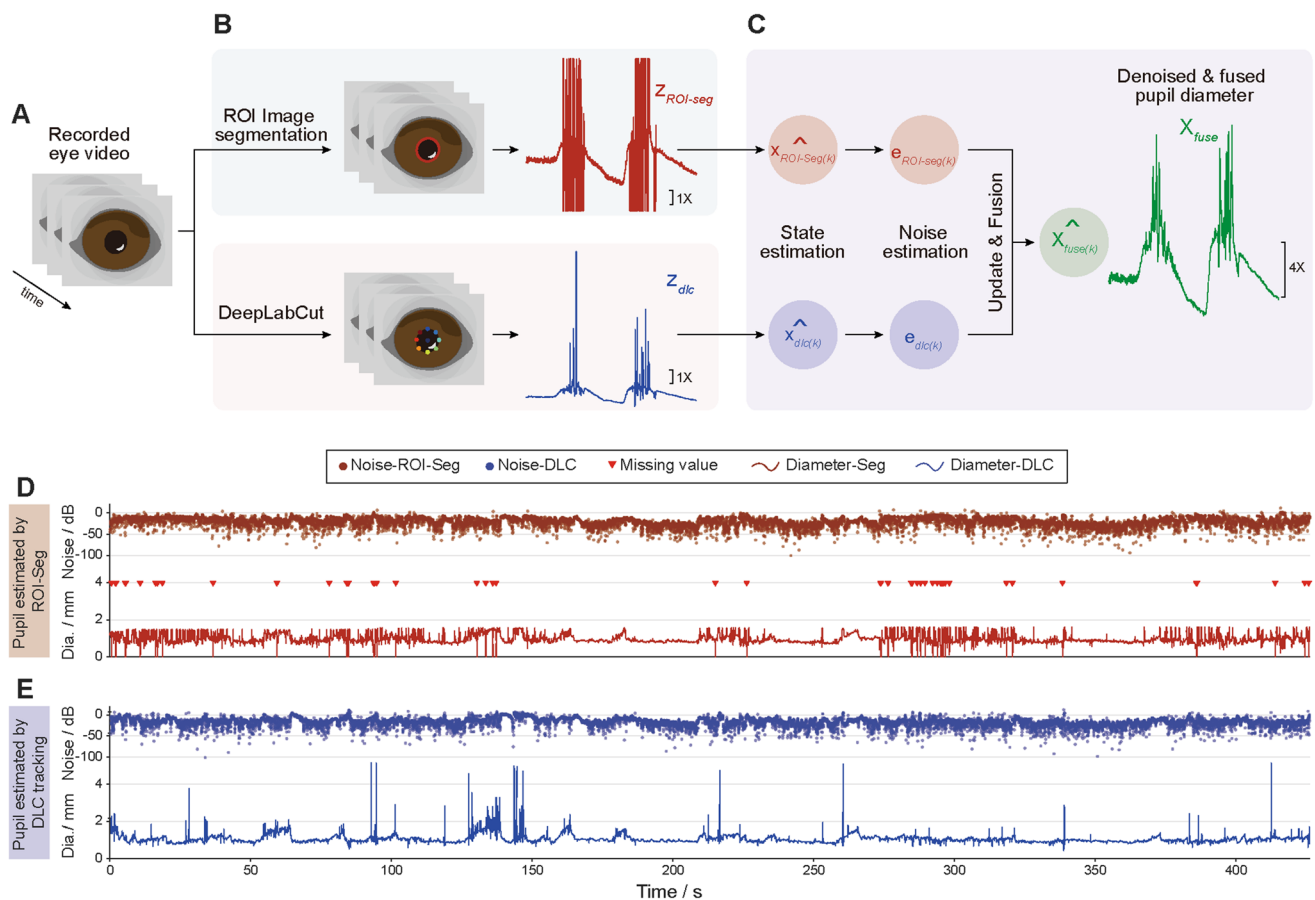


Fig. 5 Diagram of AKF-based measurement algorithm and noise estimation of two pupil extraction approaches. **A** Mouse eye video recorded online by the eye-tracking unit. **B** Extraction of pupil diameter by ROI-Seg and DLC approaches. Upper, ROI-Seg method to detect pupil region, ellipse fitting on the boundary, and estimation of the pupil diameter ($x_{ROI-Seg}$, red curve) as the major axis of the ellipse; lower, pre-trained DLC model to detect the landmark points from different angles and the center. Pupil diameter ($x_{ROI-Seg}$, red curve) is estimated based on the maximum axis of the fitted ellipse or the average of eight radii. **C** Denoising and fusion pupil diameters extracted from two independent processes by AKF. First, using AKF

to estimate current state ($\hat{x}_{ROI-Seg}(k)$ and $\hat{x}_{DLC}(k)$) and current noise ($e_{ROI-Seg}(k)$ and $e_{DLC}(k)$) of each data source according to previous observation and noise; second, calculating the fusion gain adaptively based on the current prediction and observation to obtain the denoised and fused pupil diameter (\hat{x}_{fuse} , green curve). **D**, **E** Raw pupil diameter traces extracted by ROI-Seg (**D**, red) and DLC (**E**, blue). The scattered data on the top of each panel represent the estimated measurement noise at each time point, and larger values indicate higher noise. The downward-pointing triangles mark the values missing from the pupil diameter estimation.

regarded as constant, it was difficult to achieve the optimal performance for dynamic and complex pupil tracking data in freely-moving animals. Therefore, we first applied a novel noise estimation method to calculate the Kalman gain adaptively. Compared with the conventional KF, which considers the measurement noise as constant, our algorithm was based on the difference between KF state estimation and the measurement that dynamically calculated the noise at each sampling point. We used the algorithm to calculate the measurement noise at each data point for a representative single recording (Fig. 5D, E). The measurement noise for DLC was -19.249 ± 34.639 dB and that for ROI-Seg was -25.768 ± 11.024 dB (mean \pm SD), indicating that the pupil measurement quality by DLC was superior to that by ROI-Seg. In addition, the rate of missing

values for the pupil diameter extracted by ROI-Seg was 0.97%; this was usually due to off-tracking caused by blinking or sudden illumination changes. The measurement noise affects the fused data by adjusting the Kalman gain. When there is a large difference between the measurement and estimation for one of the measurement sources, the fusion gain inclines to the source with lower noise. In addition, when a missing value appears, a conventional noise estimation method, such as the standard deviation, assumes that the noise is zero. In such situations, our algorithm estimated the noise as infinite. Therefore, the fused data directly relied on another data source.

We obtained fused pupil diameter data extracted by two individual approaches using the AKF. As shown in Fig. 6A and video 1, the representative data (~ 7 min) were

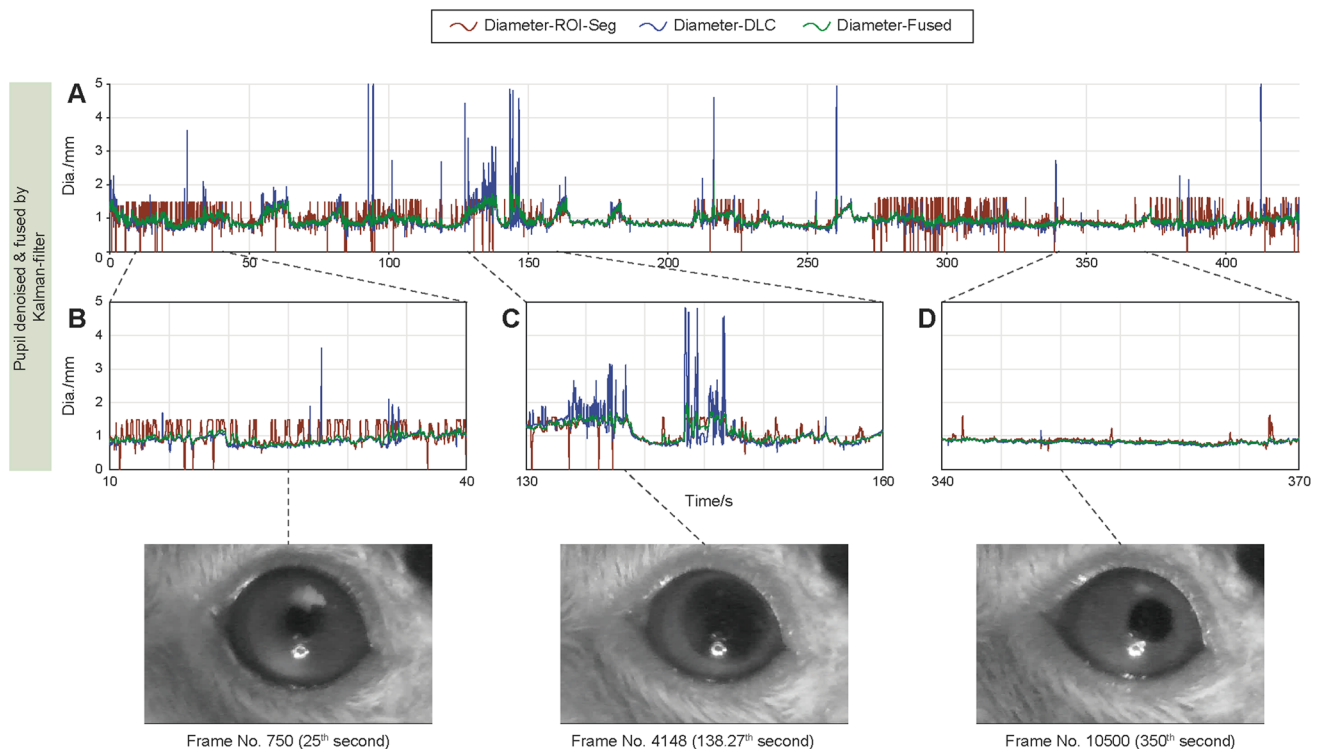


Fig. 6 Representative example of using the AKF algorithm for pupil data denoising. **A** Comparison of the raw pupil diameter shown as red and blue traces, and AKF fused pupil diameter in green. **B–D** Three

selected magnified charts and the corresponding eye images under different noise conditions.

Table 3 Hardware comparison.

References	Present design	Wallace <i>et al.</i> , Nature, 2013	Meyer <i>et al.</i> , Neuron, 2018
Species	Mice	Rats	Mice
Resolution	1280×720	752×480	640×480
Frame rates	30 Hz	40–50 Hz	90 Hz
Dimensions (single side)	12×12 mm	50×40 mm	22×15 mm
Weight	0.8 g	3.7 g	2.6 g (only camera system)
Applications	Eye tracking, EEG recording	Eye tracking	Eye tracking, head movement, electrophysiology
Highlights	Lightest hardware, detachable design, robust pupil detection	First eye tracking device on freely-moving rodents	Multimodal recording

smoother after fusion and showed a stronger anti-noise capacity. We manually examined the details of the fused data with their corresponding pupil image: (1) when the data extracted by ROI-Seg were noisy, the fused data had a higher level of trust in DLC (Fig. 6B); (2) when the data extracted by DLC were noisy, the fused data were more dependent on ROI-Seg (Fig. 6C); and (3) when both data had a low noise level, the two data sources were averaged using a compromise strategy (Fig. 6D). We further evaluated the accuracy of AKF under five noise conditions. The evaluation results show that the average error of AKF (Fused: 6.03 ± 8.92) was lower than DLC (7.05 ± 18.70)

and ROI-Seg (22.66 ± 25.18). These results show that our algorithm can effectively improve the measurement of pupil diameter.

Discussion and Conclusion

We developed a compatible eye-tracking system to monitor the eye movements of freely-moving animals, contributing to both the hardware and software development of this novel system. We reduced the size and weight of the hardware, while improving the imaging quality. We

integrated an ultra-micro camera, Lan ED infrared illumination source, an acquisition circuit, and a signal transmission circuit on a thin flexible printed circuit. The lightweight hardware and detachable design allowed our device to be flexibly combined with other neural implants. Our hardware focused on a lightweight design (Table 3). When only one eye was recorded, the heaviest design, Sch2, was 1.01 g, which is lighter than a previously-developed eye-tracking device for freely-moving animals [34]. The detachable design between the ETU and the ETU connector enabled the flexible combination of various neural implant devices for stable and long-term studies.

Accurately extracting pupil dynamics to match neural activity is essential for many neuroscience studies. Conventional methods based on image processing, such as adaptive thresholding [43] or edge detection (Else) [48] followed by ellipse fitting [49], circular binary feature (CBF) detection [50], exclusive curve selector (ExCuSe) [51], and the Haar-like feature-based methods, which can achieve 1000 fps pupil detection [52], have been widely used. Recent studies have addressed these challenges by training deep neural networks with thousands of manually labeled data, such as PupilNet [53], Unet-based pupil segmentation [54], DeepVOG [55], and Tiny convolution neural network (TinyCNN) [56]. In this study, we applied DLC to detect the pupils and systematically compared its performance with previous methods and different datasets. The results showed that DLC had significant advantages in accuracy, speed, and the number of training sets (Supplementary Materials and Fig. S3 and Table S1). We demonstrated that although traditional ROI-Seg performs badly when the pupil shakes or under sudden illumination changes, it can more easily obtain accurate pupil features without introducing human bias in a stable condition [57]. Therefore, we proposed an efficient data fusion algorithm to complement the advantages of these data sources. This algorithm optimized pupil feature detection by robust adaptability under various conditions. In both the test and practical application, there was no need to specify the parameters for each dataset, and the algorithm was able to dynamically estimate the optimal pupil features.

Accurate and robust eye-tracking and eye-movement feature detection in freely-moving animals is beneficial in many aspects. Our system is a general-purpose tool that can be applied across species and in human studies. The miniature camera sensor is superior to the bulky and heavy devices commonly designed. The two-stage acquisition circuit and detachable design allow users to mount the device on monkeys and humans with minimal modifications to our hardware. Our pupil detection algorithm can also be used with other commercial custom-built pupilometers. For mice, we will make the pre-trained model and open-source code publicly available. Taken together,

these advantages will facilitate the integration of our system with those of two-photon imaging in freely-moving animals [58–60], electrophysiology [61], optogenetics [62, 63], virtual reality [64], and machine learning-based behavioral analysis [65], enabling it to become an important readout toolbox for neuroscience research.

Limitations of the Study

The current version of our device is limited in terms of the frame rate, and we are considering upgrading the acquisition circuit for other applications with higher temporal resolution. Nevertheless, most recent studies on pupil dilation in small animals focus on changes in the time domain. Even for spectral analysis, related reports have shown that pupil oscillations in mice and humans are usually around 1 Hz [66], and pupil dilation power is <3 Hz in most cases [67, 68]. Thus, our current version can meet the specifics of most applications. Besides, currently, DLC pupil estimation is still offline. The next step will consider integrating the DLC model and AKF into online recording software.

Resource Availability

The source code for the data analysis of this paper can be found at <https://github.com/huangkang314/PupilLab>. The pre-trained DLC model along with the labeled datasets (mice and PupilNet datasets), and pixel labeled data used in the comparison of different methods are available in the Zenodo repository <https://doi.org/10.5281/zenodo.5602387>. Any other relevant data are available upon reasonable request.

Acknowledgements We thank Dr. Feng Wang for her comments on our manuscript and providing the animals. This work was supported in part by the National Key R&D Program of China (2021ZD0203902 and 2018YFA0701403), the Key Area R&D Program of Guangdong Province (2018B030338001 and 2018B030331001), the National Natural Science Foundation of China (31500861, 31630031, 91732304, and 31930047), the Chang Jiang Scholars Program and the Ten Thousand Talent Program, the International Big Science Program Cultivating Project of the Chinese Academy of Science (CAS) (172644KYS820170004), the Strategic Priority Research Program of the CAS (XDB32030100), the Youth Innovation Promotion Association of the CAS (2017413), Shenzhen Government Basic Research Grants (JCYJ20170411140807570, JCYJ20170413164535041), the Science, Technology and Innovation Commission of Shenzhen Municipality (JCYJ20160429185235132), a Helmholtz-CAS joint research grant (GJHZ1508), the Guangdong Provincial Key Laboratory of Brain Connectome and Behavior (2017B030301017), the Guangdong Special Support Program, the Key Laboratory of the CAS (2019DP173024), the Shenzhen Key Science and Technology Infrastructure Planning Project (ZDKJ20190204002).

Conflict of Interest The authors declare no competing interests.

References

1. Vázquez-Guardado A, Yang YY, Bandodkar AJ, Rogers JA. Recent advances in neurotechnologies with broad potential for neuroscience research. *Nat Neurosci* 2020, 23: 1522–1536.
2. Parker PRL, Brown MA, Smear MC, Niell CM. Movement-related signals in sensory areas: Roles in natural behavior. *Trends Neurosci* 2020, 43: 581–595.
3. Byrom B, McCarthy M, Schueler P, Muehlhausen W. Brain monitoring devices in neuroscience clinical research: The potential of remote monitoring using sensors, wearables, and mobile devices. *Clin Pharmacol Ther* 2018, 104: 59–71.
4. Mathôt S. Pupillometry: psychology, physiology, and function. *J Cogn* 2018, 1: 16.
5. Lim JZ, Mountstephens J, Teo J. Emotion recognition using eye-tracking: Taxonomy, review and current challenges. *Sensors (Basel)* 2020, 20: E2384.
6. Joshi S, Gold JI. Pupil size as a window on neural substrates of cognition. *Trends Cogn Sci* 2020, 24: 466–480.
7. Kret ME, Sjak-Shie EE. Preprocessing pupil size data: Guidelines and code. *Behav Res Methods* 2019, 51: 1336–1342.
8. Dennis EJ, El Hady A, Michael A, Clemens A, Tervo DRG, Voigts J. Systems neuroscience of natural behaviors in rodents. *J Neurosci* 2021, 41: 911–919.
9. Payne HL, Raymond JL. Magnetic eye tracking in mice. *Elife* 2017, 6: e29222.
10. Tresanchez M, Pallejà T, Palacín J. Optical mouse sensor for eye blink detection and pupil tracking: Application in a low-cost eye-controlled pointing device. *J Sensors* 2019, 2019: 3931713.
11. Fuhr W, Tonsen M, Bulling A, Kasneci E. Pupil detection for head-mounted eye tracking in the wild: An evaluation of the state of the art. *Mach Vis Appl* 2016, 27: 1275–1288.
12. van der Wel P, van Steenbergen H. Pupil dilation as an index of effort in cognitive control tasks: A review. *Psychon Bull Rev* 2018, 25: 2005–2015.
13. Maier SU, Grueschow M. Pupil dilation predicts individual self-regulation success across domains. *Sci Rep* 2021, 11: 1–18.
14. Hu YZ, Jiang HH, Liu CR, Wang JH, Yu CY, Carlson S, *et al.* What interests them in the pictures? —Differences in eyetracking between rhesus monkeys and humans. *Neurosci Bull* 2013, 29: 553–564.
15. Cheng YH, Liu WJ, Yuan XY, Jiang Y. The eyes have it: Perception of social interaction unfolds through pupil dilation. *Neurosci Bull* 2021, 37: 1595–1598.
16. Rusch T, Korn CW, Gläscher J. A two-way street between attention and learning. *Neuron* 2017, 93: 256–258.
17. Clewett D, Gasser C, Davachi L. Pupil-linked arousal signals track the temporal organization of events in memory. *Nat Commun* 2020, 11: 4007.
18. Itti L. New eye-tracking techniques may revolutionize mental health screening. *Neuron* 2015, 88: 442–444.
19. Olpinska-Lischka M, Kujawa K, Wirth JA, Antosiak-Cyrak KZ, Maciaszek J. The influence of 24-hr sleep deprivation on psychomotor vigilance in young women and men. *Nat Sci Sleep* 2020, 12: 125–134.
20. Stütt I, Zhou ZC, Radtke-Schuller S, Fröhlich F. Arousal dependent modulation of thalamo-cortical functional interaction. *Nat Commun* 2018, 9: 2455.
21. Milton R, Shahidi N, Dragoi V. Dynamic states of population activity in prefrontal cortical networks of freely-moving macaque. *Nat Commun* 1948, 2020: 11.
22. Lawson RP, Mathys C, Rees G. Adults with autism overestimate the volatility of the sensory environment. *Nat Neurosci* 2017, 20: 1293–1299.
23. Constantino JN, Kennon-McGill S, Weichselbaum C, Marrus N, Haider A, Glowinski AL, *et al.* Infant viewing of social scenes is under genetic control and is atypical in autism. *Nature* 2017, 547: 340–344.
24. Lotankar S, Prabhavalkar KS, Bhatt LK. Biomarkers for Parkinson's disease: Recent advancement. *Neurosci Bull* 2017, 33: 585–597.
25. Katus L, Hayes NJ, Mason L, Blasi A, McCann S, Darboe MK, *et al.* Implementing neuroimaging and eye tracking methods to assess neurocognitive development of young infants in low- and middle-income countries. *Gates Open Res* 2019, 3: 1113.
26. Lio G, Fadda R, Doneddu G, Duhamel JR, Sirigu A. Digit-tracking as a new tactile interface for visual perception analysis. *Nat Commun* 2019, 10: 5392.
27. Simmonds L, Bellman S, Kennedy R, Nenycz-Thiel M, Bogomolova S. Moderating effects of prior brand usage on visual attention to video advertising and recall: An eye-tracking investigation. *J Bus Res* 2020, 111: 241–248.
28. Guo J, Kurup U, Shah M. Is it safe to drive? An overview of factors, metrics, and datasets for driveability assessment in autonomous driving. *IEEE Transactions on Intelligent Transportation Systems* 2020, 21: 3135–3151.
29. Privitera M, Ferrari KD, von Ziegler LM, Sturman O, Duss SN, Floriou-Servou A, *et al.* A complete pupillometry toolbox for real-time monitoring of locus coeruleus activity in rodents. *Nat Protoc* 2020, 15: 2301–2320.
30. Meng QS, Tan XR, Jiang CY, Xiong YY, Yan B, Zhang JY. Tracking eye movements during sleep in mice. *Front Neurosci* 2021, 15: 616760.
31. Schwarz JS, Sridharan D, Knudsen EI. Magnetic tracking of eye position in freely behaving chickens. *Front Syst Neurosci* 2013, 7: 91.
32. Yorzinski JL. A songbird inhibits blinking behaviour in flight. *Biol Lett* 2020, 16: 20200786.
33. Wallace DJ, Greenberg DS, Sawinski J, Rulla S, Notaro G, Kerr JND. Rats maintain an overhead binocular field at the expense of constant fusion. *Nature* 2013, 498: 65–69.
34. Meyer AF, Poort J, O'Keefe J, Sahani M, Linden JF. A head-mounted camera system integrates detailed behavioral monitoring with multichannel electrophysiology in freely moving mice. *Neuron* 2018, 100: 46-60.e7.
35. Michael AM, Abe ET, Niell CM. Dynamics of gaze control during prey capture in freely moving mice. *Elife* 2020, 9: e57458.
36. Mathis A, Mamidanna P, Cury KM, Abe T, Murthy VN, Mathis MW, *et al.* DeepLabCut: markerless pose estimation of user-defined body parts with deep learning. *Nat Neurosci* 2018, 21: 1281–1289.
37. Chen LC, Zhu YK, Papandreou G, Schroff F, Adam H. Encoder-decoder with atrous separable convolution for semantic image segmentation. *Comput Vis – ECCV* 2018, 2018: 833–851.
38. Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* 2015, 2015: 3431–3440.
39. Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. *Med Image Comput Comput Assist Interv – MICCAI* 2015, 2015: 234–241.
40. Lopes G, Bonacchi N, Frazão J, Neto JP, Atallah BV, Soares S, *et al.* Bonsai: an event-based framework for processing and controlling data streams. *Front Neuroinform* 2015, 9: 7.
41. Naylor E, Aillon DV, Gabbert S, Harmon H, Johnson DA, Wilson GS, *et al.* Simultaneous real-time measurement of EEG/EMG and L-glutamate in mice: A biosensor study of neuronal activity during sleep. *J Electroanal Chem (Lausanne)* 2011, 656: 106–113.

42. Yüzgeç Ö, Prsa M, Zimmermann R, Huber D. Pupil size coupling to cortical states protects the stability of deep sleep via parasympathetic modulation. *Curr Biol* 2018, 28: 392–400.e3.
43. Bradley D, Roth G. Adaptive thresholding using the integral image. *J Graph Tools* 2007, 12: 13–21.
44. Gao B, Hu G, Gao S, Zhong Y, Gu C. Multi-sensor optimal data fusion based on the adaptive fading unscented Kalman filter. *Sensors (Basel)* 2018, 18: 488.
45. Akhlaghi S, Zhou N, Huang Z. Adaptive adjustment of noise covariance in Kalman filter for dynamic state estimation. *IEEE Power & Energy Society General Meeting 2017, 2017: 1–5.*
46. Mourikis AI, Roumeliotis SI. A multi-state constraint Kalman filter for vision-aided inertial navigation. *Proceedings 2007 IEEE International Conference on Robotics and Automation, 2007: 3565–3572.* DOI: <https://doi.org/10.1109/robot.2007.364024>.
47. Lai ZL, Lei Y, Zhu SY, Xu YL, Zhang XH, Krishnaswamy S. Moving-window extended Kalman filter for structural damage detection with unknown process and measurement noises. *Measurement* 2016, 88: 428–440.
48. Fuhl W, Santini T, Kübler TC, Kasneci E. ElSe: Ellipse selection for robust pupil detection in real-world environments. *Proceedings of the Ninth Biennial ACM Symposium on Eye Tracking Research & Applications 2016: 123–130.*
49. Li DH, Winfield D, Parkhurst DJ. Starburst: A hybrid algorithm for video-based eye tracking combining feature-based and model-based approaches. *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Workshops 2005: 79.*
50. Fuhl W, Geisler D, Santini T, Appel T, Rosenstiel W, Kasneci E. CBF: Circular binary features for robust and real-time pupil center detection. *Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications 2018, Article 8: 1–6.*
51. Fuhl W, Kübler T, Sippel K, Rosenstiel W, Kasneci E. Excuse: Robust pupil detection in real-world scenarios. *International Conference on Computer Analysis of Images and Patterns 2015: 39–51.*
52. Fuhl W. 1000 pupil segmentations in a second using haar like features and statistical learning. 2021: arXiv: 2102.01921[[ees-IV](https://arxiv.org/abs/2102.01921)]. <https://arxiv.org/abs/2102.01921>
53. Fuhl W, Santini T, Kasneci G, Kasneci E. PupilNet: convolutional neural networks for robust pupil detection. 2016: arXiv: 1601.04902[[cs.CV](https://arxiv.org/abs/1601.04902)]. <https://arxiv.org/abs/1601.04902>
54. Kitazumi K, Nakazawa A. Robust pupil segmentation and center detection from visible light images using convolutional neural network. *2018 IEEE International Conference on Systems, Man, and Cybernetics 2018: 862–868.*
55. Yiu YH, Aboulatta M, Raiser T, Ophey L, Flanagan VL, Zu Eulenburg P, *et al.* DeepVOG: Open-source pupil segmentation and gaze estimation in neuroscience using deep learning. *J Neurosci Methods* 2019, 324: 108307.
56. Fuhl W, Gao H, Kasneci E. Tiny convolution, decision tree, and binary neuronal networks for robust and real time pupil outline estimation. *ACM Symposium on Eye Tracking Research and Applications. 2020, Article 5: 1–5.*
57. Bushnell M, Umino Y, Solessio E. A system to measure the pupil response to steady lights in freely behaving mice. *J Neurosci Methods* 2016, 273: 74–85.
58. Zong WJ, Wu RL, Li ML, Hu YH, Li YJ, Li JH, *et al.* Fast high-resolution miniature two-photon microscopy for brain imaging in freely behaving mice. *Nat Methods* 2017, 14: 713–719.
59. Klioutchnikov A, Wallace DJ, Frosz MH, Zeltner R, Sawinski J, Pawlak V, *et al.* Three-photon head-mounted microscope for imaging deep cortical layers in freely moving rats. *Nat Methods* 2020, 17: 509–513.
60. Griffiths VA, Valera AM, Lau JY, Roš H, Younts TJ, Marin B, *et al.* Real-time 3D movement correction for two-photon imaging in behaving animals. *Nat Methods* 2020, 17: 741–748.
61. Juavinett AL, Bekheet G, Churchland AK. Chronically implanted Neuropixels probes enable high-yield recordings in freely moving mice. *Elife* 2019, 8: e47188.
62. Karl D. Optogenetics. *Nat Methods* 2011, 8: 26–29.
63. Zhang F, Wang LP, Brauner M, Liewald JF, Kay K, Watzke N, *et al.* Multimodal fast optical interrogation of neural circuitry. *Nature* 2007, 446: 633–639.
64. Stowers JR, Hofbauer M, Bastien R, Griessner J, Higgins P, Farooqui S, *et al.* Virtual reality for freely moving animals. *Nat Methods* 2017, 14: 995–1002.
65. Huang K, Han YN, Chen K, Pan HL, Zhao GY, Yi WL, *et al.* A hierarchical 3D-motion learning framework for animal spontaneous behavior mapping. *Nat Commun* 2021, 12: 2784.
66. Lamirel C, Ajasse S, Moulignier A, Salomon L, Deschamps R, Gueguen A, *et al.* A novel method of inducing endogenous pupil oscillations to detect patients with unilateral optic neuritis. *PLoS One* 2018, 13: e0201730.
67. Naber M, Alvarez GA, Nakayama K. Tracking the allocation of attention using human pupillary oscillations. *Front Psychol* 2013, 4: 919.
68. Naber M, Roelofzen C, Fracasso A, Bergsma DP, van Genderen M, Porro GL, *et al.* Gaze-contingent flicker pupil perimetry detects scotomas in patients with cerebral visual impairments or glaucoma. *Front Neurol* 2018, 9: 558.