

YORU: social behavior detection based on user-defined animal appearance using deep learning

1 Authors/affiliations

2 Hayato M Yamanouchi^{1,7,8,*}, Ryosuke F Takeuchi^{2,7}, Naoya Chiba³, Koichi Hashimoto⁴,
3 Takashi Shimizu¹, Ryoya Tanaka^{1,5,8,*} & Azusa Kamikouchi^{1,5,6,8,*}

4 ¹Graduate School of Science, Nagoya University, Nagoya, Aichi, 464-8602, Japan

5 ²Graduate School of Pharmaceutical Sciences, Nagoya University, Nagoya, Aichi, 464-8602,
6 Japan

7 ³Cybermedia Education Division, D3 Center, Osaka University, Toyonaka, Osaka, 560-0043,
8 Japan

9 ⁴Graduate School of Information Sciences, Tohoku University, Sendai, Miyagi, 980-8579,
10 Japan

11 ⁵Institute for Advanced Research, Nagoya University, Nagoya, Aichi, 464-8601, Japan

12 ⁶Institute of Transformative Bio-Molecules (WPI-ITbM), Nagoya University, Nagoya, Aichi,
13 464-8602, Japan

14

15 Author list footnotes

16 ⁷These authors contributed equally and share first authorship

17 ⁸These authors contributed equally and share corresponding authorship

18 *Corresponding authors

19

20 Contact info

21 Hayato M YAMANOUCHI: haya.m.yamano.neuro@gmail.com

22 Ryoya TANAKA: tanaka.ryoya.z3@f.mail.nagoya-u.ac.jp

23 Azusa KAMIKOUCHI: kamikouchi@bio.nagoya-u.ac.jp

24

25 **Abstract**

26 The creation of tools using deep learning methodologies for animal behavior analysis has
27 revolutionized neuroethology. They allow researchers to analyze animal behaviors and reveal
28 causal relationships between specific neural circuits and behaviors. However, the application
29 of such annotation/manipulation systems to social behaviors, in which multiple individuals
30 interact dynamically, remains challenging. Here, we applied an object detection algorithm to
31 classify animal social behaviors. Our system, packaged as "YORU" (Your Optimal
32 Recognition Utility), classifies animal behaviors, including social behaviors, based on the
33 shape of the animal as a "behavior object". It successfully classified several types of social
34 behaviors ranging from vertebrates to insects. We also integrated a closed-loop control
35 system for operating optogenetic devices into the YORU package. YORU enables real-time
36 delivery of photostimulation feedback to specific individuals during specific behaviors, even
37 when multiple individuals are close together. We hope that the YORU system will accelerate
38 the understanding of the neural basis of social behaviors.

39

40 **Introduction**

41 Social behaviors such as courtship, aggression, and group formation are important for
42 improving survival rate and reproductive efficiency in a wide range of animals, including
43 invertebrates and vertebrates^{1–3}. To accelerate our understanding of the neural basis of these
44 behaviors, it is necessary to capture information on the location and type of social interactions
45 that occur between individuals⁴. Recent advances in machine learning have led to the
46 establishment of various tools for animal behavior detection, represented by markerless body
47 part tracking^{5–7}. These emerging tools enable real-time behavior analysis, allowing
48 researchers to manipulate neural activity precisely when the animal exhibits the behavior of
49 interest^{8–10}. Such a closed-loop approach promises to clarify the causal relationship between
50 the neural activity and behaviors, and indeed has yielded significant success in uncovering
51 neural basis for single individual behavior^{7,11,12}. However, detection of the social interaction

52 between multiple individuals with low latency remains challenging^{4,6,7}. This difficulty arises
53 from the complexity of defining social behavior based on the coordinates of each individual's
54 body parts, preventing the effective use of a closed-loop approach^{6,7,13}. Therefore, novel
55 methods are required to enable the detection of social interaction and to perform real-time
56 behavior analysis for neural intervention.

57 Social interactions are often accompanied by individuals striking distinctive postures
58 to interact with others, such as wing extension by courting male fruit flies and mounting on
59 females by male mice attempting copulation^{14–16}. In this study, we propose the detection of
60 social behaviors of multiple individuals based on their appearances by defining each as a
61 “behavior object”. To this end, we focused on an object detection algorithm “YOLOv5”, a
62 high-speed object detection algorithm based on a convolutional neural network (CNN)^{17,18}.
63 The algorithm achieves fast object recognition by framing detection as a single regression
64 problem, using a unified architecture that processes the entire image in one forward pass^{17,18}.
65 The YOLO-based detection is robust to variations in object orientation, size and background
66 noise¹⁹. YOLOv5 is therefore a promising algorithm to detect various animals' social
67 behaviors without relying on the body-part coordinates, making it well-suited for adapting
68 behavior analysis and a closed-loop system to investigate social interactions.

69 Here, we established a behavior detection system called “YORU” (Your Optimal
70 Recognition Utility), based on the YOLOv5 algorithm¹⁸. First, to verify the proof of concept,
71 we tested whether YORU could be used to detect social behaviors of various animals ranging
72 from vertebrates to insects. As an application of our behavior analysis, we compared the
73 detection readouts with neural activity imaging in mice to interpret large-scale brain activity.
74 Secondly, we evaluated the inference speed of the detection and feedback latency in YORU's
75 real-time analysis. Finally, as a practical example of the YORU's closed-loop system, we
76 experimented with optogenetic neural manipulation focused on courtship behavior in
77 *Drosophila*. In the presence of multiple flies, we successfully manipulated the neural activity
78 of individuals in an individual selective manner by optogenetic stimulation of the fly
79 exhibiting the behavior of interest.

80

81 **Results**

82 **YORU is a framework for animal behavior recognition**

83 Here we introduce “YORU” (Your Optimal Recognition Utility), an animal behavior
84 detection system with a graphical user interface (GUI). In the YORU system, animal
85 behaviors, either performed by single animals or multiple animals interacting with each other,
86 are classified as a “behavior object” based on their shapes with YOLOv5¹⁸ (Fig. 1a). YORU
87 is an open-source Python software composed of four packages: “Training”, “Evaluation”,
88 “Video Analysis” and “Real-time Process” (Fig. 1b, Figs. 1a-e). The YORU system is
89 designed to work both offline (video analysis) and online (real-time analysis) to detect animal
90 behavior. During video analysis, YORU allows us to quantify animal behavior according to
91 user-defined shapes of social behavior. For real-time analysis, YORU analyzes animal
92 behavior in real-time, which can be used to output trigger signals accordingly to control
93 external devices, such as LEDs for optogenetic control as set by the user.

94 We set three constraints to the workflow of YORU: ease of use for experimenters, low
95 system latency, and high customizability. To make YORU user-friendly, we designed the
96 system to allow easy quantification of animal behaviors without requiring programming
97 knowledge. To achieve low system latency, special efforts were focused on behavior
98 detection and feedback when operating as a closed-loop system. Here, immediate behavior
99 detection is achieved by YOLOv5-based object detection algorithm, which processes
100 generating region proposals and classifying subjects simultaneously, resulting in faster
101 detection^{17,18}. Immediate feedback is achieved by the parallel processing design of YORU
102 real-time process²⁰: image acquisition, object recognition, and hardware (Arduino, DAQs,
103 etc.) manipulation are not processed serially but simultaneously. To achieve high
104 customizability, i.e., adapting YORU system to various experimental systems easily, YORU
105 implements a trigger output for hardware manipulation. The experimenter can customize

106 YORU's parameters with no or minimal user programming. These features enable
107 experimenters to easily use high-performance closed-loop neural feedback.

108 **Detection of social behaviors using object detection algorithm**

109 In previous studies, YOLO-based object detection successfully classified some social
110 behaviors of *Drosophila*, such as offensive behaviors and copulation^{21,22}. In this study, we
111 extended this idea and analyzed a variety of animal behaviors with YORU. The following
112 social behaviors were tested to validate the performance: (1) Fruit flies; wing extension
113 behavior of a male toward a female during courtship¹⁴ (Fig. 2a), (2) Ants; mouth-to-mouth
114 food transfer behavior among workers (trophallaxis)^{23,24} (Fig. 2b), and (3) Zebrafish;
115 orientation behavior toward another individual behind a partition^{25,26} (Fig. 2c). We extracted
116 2000 images from multiple videos and manually labeled their behavior objects with the
117 following definitions.

118 (1) Fruit flies: “wing_extension”; a fly extending one of its wings (Fig. 2a). When the flies
119 were not labeled as “wing_extension”, they were labeled as “fly”.

120 (2) Ants: “trophallaxis”; heads of two ants engaging in a food exchange, with their mouth
121 parts in contact with each other. “no”; the situation of no food exchange although the heads
122 of the two ants were close together (Fig. 2b). No label was given when neither of these two
123 behavior types were detected.

124 (3) Zebrafish: “orientation”; two zebrafish exhibiting orientation behavior as defined in a
125 previous study^{25,26}, “no_orientation”; a zebrafish exhibiting no orientation behavior (Fig. 2c).

126 We created models and compared their detection accuracies with human annotations using
127 multiple videos that were not used for model creation. The Accuracy score for the model's
128 detection of fruit flies, ants, and zebrafish behaviors were 93.3%, 98.3% and 90.5%,
129 respectively, when compared to human manual annotations (Fig. 2d-f, Supplementary Table
130 1). We also compared zebrafish orientation analyses between the previous method based on
131 body part tracking²⁵ and human annotations, yielding an Accuracy score of 81.2%

132 (Supplementary Table 1, Fig. 2f). YORU can thus potentially detect zebrafish orientation
133 quicker and with higher accuracy than the previous tracking methods. These results suggest
134 that YORU's detection can detect social behaviors with a similar to human annotations.

135 In general, two major factors to consider for the practical application of deep learning-
136 based analyses are the amount of training data (the number of labels) and the selection of
137 based networks ^{5,27}. To find optimal conditions in our dataset, we evaluated the accuracy of
138 models with different numbers of training images (200, 500, 1000, 1500, and 2000) and
139 YOLOv5 networks (YOLOv5n, YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x). Two
140 metrics were used in this evaluation: “Precision”, the ratio of correctly predicted detections
141 to all predicted detections, and “Recall”, the ratio of correctly predicted detections to all
142 ground-truths. All models detected the target social behavior with Precision greater than 85%
143 and Recall greater than 90%, even when only 200 images were used to create models
144 (Supplementary Table 2). The Precision and Recall scores increased with the number of
145 training images, finally exceeding 90% and 95%, respectively, when 1,000 or more images
146 were used for training (Supplementary Table 2). We then evaluated typical object detection
147 model indices, intersection over union (IOU) and average precisions (AP) (Fig. 2g-i, Figs.
148 2a-c) ²⁸. These two indexes can be used to compare the accuracy of the model between
149 conditions; the IOU shows the difference in location information between the ground truth
150 and detected bounding boxes while the AP reflects the accuracy of the behavior classes,
151 which includes Precision and Recall values ²⁸. For each behavior, neither the number of
152 images used for training nor the YOLOv5 models tested affected IOU values (Figs. 2a-c).
153 On the other hand, AP values increased with the number of training images, in which the
154 same number of images gave similar AP values across YOLOv5 based models (Fig. 2g-i).
155 These evaluations suggest the primary factor influencing accuracy is the amount of training
156 data, at least in these cases.

157 Next, we increased the number of individuals to test the performance of YOLOv5
158 models in multi-individual conditions. In the first condition, a group of flies was analyzed to
159 detect three objects: “wing extension”, “copulation”, and other behaviors labeled as “fly”

160 (Fig. 2j, k). In the second condition, a group of ants was used to detect two objects:
161 “trophallaxis” and “no_trophallaxis” (Fig. 2l, m). In the group of flies, Precision and Recall
162 scores were over 95% even when using only 200 images for training; the scores increased
163 with the number of training images (Supplementary Table 2). In the group of ants, Precision
164 and Recall scores exceeded 90% with 500 or more training images (Supplementary Table 2).
165 These findings suggest that YORU’s detection of social behaviors is also applicable to
166 multiple individuals (more than two individuals) with high accuracy (Fig. 2 j-m, Figs. 3a, b,
167 Supplementary Table 2), highlighting its usefulness for analyzing various types of social
168 behaviors. The detection accuracy does not change significantly with different YOLOv5 pre-
169 trained models and instead depends strongly on the number of training images.

170 **The relationship between behavioral readouts and neural activity interpretation**

171 One of the main questions in neurophysiology is the interpretation of which sensations
172 and animal behaviors can explain observed neural activity. One promising methodology to
173 address this question is to combine time series analyses of multiple behavioral types *via* video
174 analysis with neural activity measurements. To test the performance of YORU for such tasks,
175 we recorded dorsal cortex-wide neural activity with wide-field calcium imaging from mice
176 running in a virtual reality (VR) system which provides visual feedback coupled to their
177 locomotion²⁹ (Fig. 3a). A mouse on a treadmill in VR typically exhibits multiple behaviors,
178 such as running, grooming, and eye blinking (Fig. 3b). First, to validate the behavior
179 classification performance of YORU, we estimated the time series of eight behavior classes
180 from video analysis: “Running”, “Stop”, “Whisker-On”, “Whisker-Off”, “Eye-Open”, “Eye-
181 Closed”, “Grooming-On”, and “Grooming-Off”. Precision and Recall scores for the model's
182 detection of these behaviors were 91.8% and 92.7%, respectively, validating this model to
183 detect mouse behaviors as accurately as human manual annotations (Supplementary Table
184 3). For all classes detected by this model, the average IOUs and AP@50 (i.e., the AP value
185 at IOU=50) were above 0.60 and 0.55, respectively (Fig. 4, Supplementary Table 3). Previous
186 studies have reported that rodents actively move their whiskers to seek and identify objects
187 or avoid obstacles in front of them during locomotion³⁰⁻³³. In line with these reports, the

188 time series data of active whisker movement (corresponding to “Whisker-On” label) during
189 “Running” periods were positively correlated while “Whisker-Off” and “Running”
190 correlated negatively (Fig. 3b, c). In addition, the "Running" and "Stop" epochs in the YORU
191 readout are negatively correlated, indicating that mutually exclusive behaviors are labeled
192 exclusively (Fig. 3c). These results further confirmed that YORU’s approach to behavior
193 labeling can efficiently detect typical behaviors of mice.

194 Next, we investigated which brain regions in the cortex correlate with YORU’s readout.
195 The “Running” epoch was highly correlated with the neural activity of medial motor areas,
196 somatosensory areas representing forelimb and hindlimb information, visual areas, and
197 posterior association region (retrosplenial cortex) (Fig. 3d). Several whisker movements
198 associated with sniffing during “Stop” period and “Blinking” behavior were also coupled to
199 distinct macroscopic activity patterns of widespread regions in somatosensory and visual
200 areas. As expected, grooming behavior correlated specifically with neural activity of
201 forelimb somatosensory and motor areas (Fig. 3d, e). These results demonstrate the
202 applicability of its utility and potential for precise and quantitative interpretation of neural
203 activity with various animal behaviors.

204 **Inference speed and system latency of YORU**

205 A closed-loop system that relies on live feedback of animal behaviors requires a low-
206 latency solution for behavior detection and feedback outputs¹². To assess YORU’s potential
207 for application in closed-loop systems for social behavior analysis, we measured the total
208 time required from frame acquisition to behavior estimation using a simple light detection
209 task (Fig. 4a). Possible major factors that could affect the speed of YORU’s behavior
210 detection are (i) the network structure, (ii) image size, and (iii) computing hardware,
211 especially the graphics processing unit (GPU). To demonstrate their impact on analyzing
212 each frame, we measured the single-frame inference latency with (i) five YOLOv5
213 architectures (YOLOv5n, YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x), (ii) two input
214 image sizes (640x480 or 1280x1024 pixels, images were resized before feeding to the neural
215 network), and (iii) a variety of NVIDIA GPUs in Windows PCs (Supplementary Table 3).

216 Inference speed was calculated by comparing the time before (t1) and after (t2) the frame
217 was analyzed by a model (Figs. 5a). We found that the smallest network (YOLOv5n) was
218 the fastest and the largest network (YOLOv5x) had the largest inference latency (Figs. 5b,
219 Supplementary Table3). In addition, the inference speed was faster with a smaller image size
220 and more powerful NVIDIA GPUs (Supplementary Table 3). For example, with NVIDIA
221 RTX 4080 GPU, we achieved inference speed as low as ~ 5.0 ms per frame (~ 200 fps) using
222 YOLOv5s networks (Figs. 5b). These results suggest that the network architecture, input
223 image size, and GPU all influence YORU's inference speed.

224 The design of closed-loop systems adapted to neuroscience experiments requires the
225 simultaneous control of several processes, such as camera capture and hardware
226 manipulation³⁴. In addition, several other factors such as camera type and frame rate, trigger
227 destination type, and PC memory can affect system latency. Therefore, we tested the
228 performance of YORU's closed-loop system from end to end, including the entire process
229 from the camera capturing an image of the LED to the PC detecting whether the LED was lit
230 or not and sending a trigger signal to a data acquisition (DAQ) system upon detecting the
231 “ON” state (Fig. 4a, b). The delay between the timing of the LED turning on and the trigger
232 signal output, both detected by measuring their voltages using DAQ, was as low as 30 ms
233 per event (Fig. 4c). This suggests that the end-to-end system latency of the YORU system is
234 around 30 ms in this setup, which is sufficiently low in most cases to provide real-time
235 feedback in response to animal behavior. Next, we assessed the factors that affected the end-
236 to-end system latency, such as networks, input image size, camera frame rate, and system
237 hardware. The effect of the network differences was almost negligible except YOLOv5l and
238 v5x, which showed larger latency than others (Fig. 4c). On the other hand, the input image
239 size severely affected the latency; the average latency of the smaller image (640 x 480 pixels)
240 was ~30 ms while that of the larger image (1280 x 1024 pixels) was ~75 ms (Fig. 4c).
241 According to the camera fps latency results, even if the camera frames were acquired as fast
242 as the model's inference speed (~200 fps), the system delay would not be significant, due to
243 YORU's multi-processing system (Fig. 4d). In a multi-processing system, the size and
244 processing speed of the random access memory (RAM) affect the processing speed of the

245 system due to the shared memory. Interestingly, in YORU's real-time process, the size of the
246 RAM (between 16 GB and 32 GB) had less effect on the system latencies, suggesting that
247 16GB RAM size is minimally sufficient to operate YORU's real-time process (Figs. 6a). The
248 effect on system latency due to hardware differences (cameras and trigger devices) was small
249 (Figs. 6b, c). These results suggest that inference speed and the input image size are the
250 primary factors affecting end-to-end system latency. In addition, YORU's system processing
251 speed was very fast compared to previously reported tracking-based systems ⁷.

252 **YORU application for real-time optogenetic system**

253 Next, to assess the practical applicability of the YORU closed-loop system, we applied
254 it to event-triggered optogenetic manipulation. During courtship, the male fruit fly extends
255 his wing to serenade the female with a unique sound known as the courtship song. Upon
256 hearing it, female flies gradually increase their receptivity to copulation ³⁵. We hypothesized
257 that if male wing extension behavior is inhibited when a male attempts to extend his wing,
258 copulation rates would be reduced. Using a split-GAL4 strain that specifically drives gene
259 expression in pIP10 neurons, which are descending neurons regulating courtship song
260 production ^{36,37}, we expressed the green light-gated anion channel GtACR1 ³⁸ in male pIP10
261 neurons (Fig. 5a). We then paired individual mutant males with a wild-type female in a
262 chamber and allowed YORU to detect the single-wing extension of a fly. In this system,
263 when YORU detects an object labeled as "wing extension," YORU introduces green
264 photostimulation to the entire chamber (Fig. 5b, c). As a control for photo-stimulation, we
265 used an event-triggered light with a 1-second delay that illuminated the entire chamber (Fig.
266 5c, 'Delayed' group). Male flies expressing GFP in pIP10 neurons were also used as a genetic
267 control group. We then analyzed the wing extension ratio during courtship and the copulation
268 rate during the 30-min observation period (Fig. 5d, e). In the experimental group, males
269 decreased the amount of wing extension during courtship (Fig. 5d), validating the optogenetic
270 inhibition of pIP10-induced behavior. In line with this, the cumulative copulation rate was
271 significantly lower in the experimental group than in the control groups (Fig. 5e). These
272 results confirm the importance of male pIP10 neurons for inducing the wing extension

273 behaviors observed in the previous reports^{36,37,39}, which subsequently leads to male
274 copulation success. They also validate YORU's performance in operating real-time
275 manipulation of neural activity in response to the detection of a specific behavior.

276 **YORU application for individual-focused photo-stimulation**

277 Finally, we applied YORU for individual-selective neural manipulation in response to
278 social behavior between multiple individuals. We included a projector in YORU's closed-
279 loop system to control the light pattern for optogenetic stimulation (Fig. 6a-c). When YORU
280 detects the target behavior, it sends location information to focus the projector light. We
281 utilized the fly courtship assay to test the usability of the system by suppressing female
282 hearing only when the male sings a courtship song, as exhibited by his wing extension. We
283 hypothesized that disrupting hearing in females during male courtship song production would
284 suppress female mating receptivity, leading to a reduced copulation rate. Using the *JO15-2-*
285 *Gal4* strain that selectively labels auditory sensory neurons, we expressed GtACR1³⁸ in
286 female auditory sensory neurons (i.e., JO-A and JO-B neurons) (Fig. 6d, Fig. 7). We then
287 paired these females each with a wild-type male in a chamber and utilized the online
288 capabilities of YORU. In this system, when YORU detects an object labeled "single-wing
289 extension," the YORU-operated projector illuminates the object labeled as "fly", which
290 typically is the female courted by the male (Fig. 6b, c, e). As a control for individual-focused
291 photo-stimulation, we used pattern light (3s On, 4s Off) illuminating the entire chamber (Fig.
292 6e). In the experimental group (*JO15-2>GtACR1* female with event-triggered light
293 condition), the copulation rate was significantly lower than in the control groups (Fig. 6e, f).
294 This result confirms the importance of auditory sensory neurons for females to detect the
295 male's courtship song to enhance copulation receptivity. Again, YORU was able to
296 optogenetically manipulate neural activity using individual-focused illumination, even when
297 multiple individuals were moving in a chamber at the same time. Those experiments
298 validated the usefulness of the YORU system, which can manipulate various devices, such
299 as a projector as well as a DAQ and Arduino, by trigger output. These proof-of-principle

300 experiments demonstrate the usefulness of YORU for the online detection of social behaviors
301 and manipulation of individual-focused neural activity through optogenetics.

302

303 **Discussion**

304 Here, we presented YORU, an animal behavior detection system using a YOLOv5-based
305 object detection algorithm. YORU allowed the detection of social behaviors, as well as
306 single-animal behaviors. Furthermore, by introducing real-time analysis, YORU can operate
307 a closed-loop system with low latencies and high user scalability. We also demonstrated the
308 practical applicability of real-time neuronal manipulation using the fly courtship behavior.
309 In particular, we created an individual-focused illumination system to manipulate the neural
310 activity of selected individuals in response to a specific behavior. The YORU's closed-loop
311 system is thus a powerful approach for social behavior research. On the user side, YORU can
312 be used entirely through its GUI without any programming. Just as body parts tracking-based
313 analysis tools such as DeepLabCut⁵ have revolutionized neuroscience research, YORU will
314 meet the needs of many biologists and stimulate the generation of novel, testable hypotheses.

315 In tracking-based behavioral analyses, behavior is typically defined based on the
316 location of body parts. In addition, clustering analysis (e.g., t-SNE, PCA, UMAP) of body
317 part location coordinates can be used to classify behaviors and detect unknown behavior
318 patterns⁵. However, it is challenging for tracking-based analysis to classify already known
319 behaviors with high accuracy in real-time. For example, tracking-based analysis for
320 recognizing behaviors such as wing extension in *Drosophila* requires recognition of wings
321 and body axes and definition of behaviors by their angles, and accordingly, it would fail to
322 capture behaviors if body parts are not correctly tracked even partially. In addition, to capture
323 the behavior of multiple individuals, we need to know which body part belongs to which
324 animal: Despite various algorithms being proposed to overcome these challenges, capturing
325 the behavior of multiple individuals in real time is still difficult^{6,7,13}. YORU's object
326 detection algorithm can sufficiently compensate for the shortage of body part tracking-based

327 behavior quantification even in multiple animal conditions^{6,7}. YORU's processing time is
328 fast enough for real-time processing, even at camera resolutions sufficient for animal
329 behavior quantification.

330 In behavioral neuroscience, optogenetics serves as a powerful approach to cell-type-
331 and spatiotemporal-specific control of neural activity, especially for investigating the causal
332 relationship between neural circuits and behavior^{8,40}. With state-of-the-art genetic tools, it is
333 now possible to express channelrhodopsin only in specific neurons to perform neural activity
334 intervention^{8,38}. In addition, the development of computer science has made it possible to
335 create closed-loop experimental setups and manipulate neural activity during behavior,
336 allowing us to explore the causal relationship between neural circuits and behavior in more
337 detail^{34,41,42}. In the research of the neural bases of social behaviors, optogenetic manipulation
338 of only specific individuals, even in the presence of multiple individuals, can be a powerful
339 approach⁴³. However, the difficulty of capturing the behavior of multiple individuals
340 simultaneously in real-time has made it extremely difficult to conduct online behavior
341 analysis in the presence of multiple individuals. The YORU system allows us to create a
342 closed-loop system that can analyze animal behavior and operate photostimulation for
343 optogenetics in real-time. Furthermore, the YORU system allows us to conduct individual-
344 focused optogenetics experiments with widely available equipment: projectors, cameras, and
345 personal computers.

346 Due to YORU's concept of detecting behaviors by their "snapshot" appearances, it has
347 limitations in detecting behaviors that cannot be defined within a single frame and instead
348 require time-series data for definitions²¹, such as foraging or mating attempts. To capture
349 these behaviors, it is necessary to adapt object detection algorithms to these time-series
350 behaviors, or to perform additional analysis using other tools such as DeepLabCut. Another
351 limitation lies in the hardware operation. Although the system delay in our condition was ~30
352 msec, an additional delay in trigger processing depending on the external hardware should
353 be considered. In our experiments of individual-focused phtostimulation, there was an
354 additional delay on the projector side before the patterned image signal was sent and

355 projected. This time lag may cause fast-moving individuals to move out from the area of the
356 light stimulus. Possible solutions for such cases include incorporating predictive algorithms,
357 or using a high-speed projector such as a low latency gaming projector^{44,45}. Overcoming
358 these two limitations would further widen the potential options for studying dynamic animal
359 social behaviors.

360 Various analysis tools, driven by advances in deep learning, have contributed to
361 biology research. YORU is a user-friendly system that allows all analyses to be performed
362 via a GUI, making it easy for anyone to perform analysis both in offline and online modes.
363 The object detection paradigm has the potential not only for animal behavior quantification,
364 but also for capturing various biological phenomena. In particular, it has been incorporated
365 as a tracking method for humans and fish, and has been applied to the behavioral
366 classification of animals and plants (e.g., *Drosophila* mating behaviors and the stomatal
367 opening and closing of *Arabidopsis*)^{21,22,46,47}. The YORU system can contribute to making
368 the use of object recognition algorithms more widespread in biology and to significantly
369 reduce the labor of biologists.

370

371 Materials and Methods

372 Development of YORU system

373 YORU is written in Python. The GUI of YORU is developed using DearPyGui library,
374 which enabled GUI development in a fast, interactive manner, and plotting of acquisition
375 data in real-time. To use various image acquisition devices such as webcams or other high-
376 performance machine vision cameras, we used OpenCV API for image acquisition. All
377 processing and data streams were performed in a multiprocessing manner with the Python
378 library “multiprocessing”. To label animal behaviors, YORU uses the open-source
379 annotation software “LabelImg” (<https://github.com/HumanSignal/labelImg>). To detect
380 “behavior objects”, YORU uses YOLOv5 packages (<https://github.com/ultralytics/yolov5>).

381 The code for YORU is available at (<https://github.com/Kamikouchi-lab/YORU>) as open-
382 source software. In this study, we used ChatGPT (<https://chatgpt.com>), developed by OpenAI,
383 as a tool to assist in various programming tasks. ChatGPT was used for code generation,
384 debugging, refactoring suggestions, and answering technical questions. While ChatGPT was
385 used to assist in specific tasks, all generated output was reviewed and validated by the
386 researchers to ensure accuracy and reliability.

387 **Datasets**

388 To evaluate the performance of YORU, we prepared different datasets collected under
389 various conditions. The datasets include “Fly - wing extension”, “Ant - trophallaxis”, “Fly -
390 group courtship”, “Ant - group trophallaxis”, “Zebrafish - orientation”, “LED lighting -
391 Small”, and “LED lighting - Large”. The “Zebrafish - orientation” dataset was created using
392 zebrafish videos from a previous study ²⁶, while the other datasets were generated from
393 videos obtained in this study. Each dataset consisted of images (frames manually extracted
394 from videos) and "behavioral object" labels. Each image was then manually labeled using
395 LabelImg. Supplementary Table 5 shows the detailed condition of each dataset.

396 Flies - wing extension

397 This dataset includes a pair of wild-type fruit flies consisting of a male and a female.
398 Fruit flies (*Drosophila melanogaster*, Canton-S strain) were raised on standard yeast-based
399 media on a 12 h light/12 h dark (12 h L/D) cycle. Both sexes of flies were collected within 8
400 h after eclosion to ensure their virgin status. They were maintained at 25°C under a 12 h L/D
401 cycle and transferred to new tubes every 2 to 4 days, except on the day of the experiment.
402 Male flies were kept singly in a plastic tube (1.5 mL, Eppendorf) containing ~200 µL fly
403 food, while females were kept in groups of 10 to 30. Experiments were conducted using
404 males and females 4 to 8 days post-eclosion, with each individual used only once. The video
405 recordings were performed between Zeitgeber Time (ZT) = 1-11 at 25°C and 40-60% relative
406 humidity.

407 Courtship behavior was monitored in a round courtship chamber with a sloped wall
408 (20 mm top diameter, 12 mm bottom diameter, 4 mm height, and 6 mm radius fillet) made
409 of transparent polylactic acid filament by 3D printer (Sermoon D1, Creality 3D Technology
410 Co., Ltd.). The chamber was enclosed with a slide glass and a white acrylic plate as a lid and
411 bottom, respectively. The chamber was illuminated from the bottom by an infrared LED light
412 (ISL-150×150-II94-BT, 940 nm, CCS INC.) to enable recordings in dark conditions. Male
413 and female flies were gently introduced into chambers by aspiration without anesthesia.
414 Videos were captured from the top, with a monochrome CMOS camera (DMK33UX273,
415 The Imaging Source Asia Co., Ltd.) equipped with a 25 mm focal length lens (TC2514-3MP,
416 KenkoTokina Corporation) and a light-absorbing and infrared transmitting filter (IR-82,
417 FUJIFILM), at a resolution of 640 x 480 pixels and 30 fps for 30 min for each pair using IC
418 Capture (The Imaging Source Asia Co., Ltd.). In this dataset, we labeled two behavior object
419 classes, "fly" and "wing_extension", without a female or male identification. The "wing_extension"
420 was labeled on the fly when it extended one of its wings. The "fly"
421 indicates a fly not showing wing extension.

422 Flies - group courtship

423 This dataset includes a group of wild-type fruit flies consisting of four males and four
424 females. We prepared these flies under the same condition described in the "Fly - wing
425 Extension" dataset. Behaviors were monitored in a fly bowl chamber with a sloped wall (60
426 mm in diameter, 3.5 mm in depth)⁴⁸, illuminated by a visible LED light to facilitate
427 recordings. Four male and four female flies were gently introduced into the chamber by
428 aspiration without anesthesia. Videos were captured from the top as described in the "Fly -
429 wing Extension" dataset. In this dataset, we labeled three behavior object classes: "fly",
430 "wing_extension", and "copulation" without a female or male identification. The definitions
431 of "wing_extension" and "fly" were based on those in the "Fly - wing extension" dataset.
432 The "copulation" was defined by genital coupling between a male and a female.

433 Ants - trophallaxis

434 This dataset includes two worker ants. Workers of *Camponotus japonicus* were
435 collected in October 2023 at the Higashiyama Campus of Nagoya University (35°09'15.3" N,
436 136°58'15.8" E). Ant species were identified with the *Encyclopedia of Japanese Ant*⁴⁹. After
437 collection, they were kept individually in Falcon tubes with moistened paper for 24 h at 22°C
438 with 40-60% relative humidity under dark conditions. The video recordings were conducted
439 at 22°C and 40-60% relative humidity.

440 The trophallaxis behavior was monitored in a custom-made chamber (25 mm width,
441 15 mm length, 5 mm depth) made of transparent polylactic acid filament by 3D printer
442 (Sermoon D1, Creality 3D Technology Co., Ltd.). The chamber was enclosed with a glass
443 ceiling. Two worker ants were used for a single video recording: one individual was fed 1 M
444 sucrose solution as much as desired immediately before observation, while the other was not
445 fed anything. These two workers were transferred to a custom-made chamber. Videos were
446 captured from the top, with a color CMOS camera (DFK33UP1300, The Imaging Source
447 Asia Co., Ltd.) equipped with a 50 mm focal length lens (MVL50M23, Thorlabs, Inc.), at a
448 resolution of 1280 x 960 pixels and 30 fps for 30 min for each pair using IC Capture (The
449 Imaging Source Asia Co., Ltd.). In this dataset, we labeled two behavior object classes:
450 “trophallaxis” and “no”. The “trophallaxis” class was defined as a situation when the heads
451 of two individuals were in proximity and the palps were in contact with each other. The “no”
452 class was defined when the heads of the two individuals were in proximity, but the palps
453 were not in contact with each other.

454 Ant - group trophallaxis

455 This dataset includes a group of ants consisting of six worker ants. Among six workers,
456 three individuals were fed 1 M sucrose solution as much as desired immediately before
457 observation, while the other three were not fed anything. These workers were then transferred
458 to the polystyrene chamber (87mm width, 57mm length, 19mm depth) with a glass ceiling,
459 and video recording was started. Videos were recorded from the top with a monochrome
460 CMOS camera (FLIR GS3-U3-15S5; Edmund optics) equipped with a zoom lens (M0814-

461 MP2, CBC Optics Co., Ltd.). Video recordings and behavior object definitions were
462 performed in the same manner as for the “Ant - trophallaxis” dataset.

463 Zebrafish - orientation

464 This dataset includes a pair of wild-type zebrafish. Adult zebrafish, aged four to six
465 months, with the Oregon AB genetic background were used without sex identification. Fish
466 were maintained in a 14/10 h L/D cycle at 28.5°C. The experiments were conducted as
467 described in previous studies^{25,26}. Fish were isolated in individual tanks the day before the
468 experiments, with white paper placed between the tanks to prevent visual contact. The
469 experiments were conducted at ZT = 0-14.

470 The orientation behavior was monitored in two custom-made acrylic tanks (90 mm
471 length, 180 mm width, and 60 mm height). These tanks were separated by a divider made
472 from a polymer dispersed liquid crystal (PDLC) film (Sunice Film) attached to 2 mm thick
473 acrylic sheets. Fish were placed individually into water tanks with a depth of 57 mm for 20
474 min. Following this acclimation period, the fish were recorded for 5 min at 30 fps with the
475 opaque divider condition (invisible condition). The divider was then made transparent to
476 allow the fish to see each other, and the recording continued for an additional 5 min (visible
477 condition). Videos were captured from below, with a monochrome CMOS camera (FLIR
478 FL3-U3-13E4, Edmund optics) at a resolution of 1280 x 1024 pixels and 30 fps. The divider
479 condition (opaque or transparent) was controlled with a DAQ interface (USB-6008; National
480 Instruments Co.) and custom-made software written in LabVIEW (National Instruments Co.).
481 In this dataset, we labeled two behavior object classes: “orientation” and “no_orientation”.
482 The “orientation” class was defined as a situation when two zebrafish showed orientation
483 behavior as defined in a previous study²⁵. The “no_orientation” class was defined when two
484 zebrafish showed no orientation behavior (Fig. 2f).

485 LED – ON or OFF

486 This dataset includes a blue LED (OSB5YU3Z74A, OptoSupply). We created “LED
487 lighting - Small” and “LED lighting - Large” datasets; “LED lighting - Small” consists of
488 640x480 pixel videos, and “LED lighting - Large” consists of 1280x1024 pixel videos.
489 Videos were captured with a color CMOS camera (DFK33UP1300, The Imaging Source
490 Asia Co., Ltd) equipped with a 50 mm focal length lens (MVL50M23, Thorlabs, Inc.). In
491 each dataset, we created models based on each YOLOv5 model (YOLOv5n, YOLOv5s,
492 YOLOv5m, YOLOv5l, and YOLOv5x). In these models, there were two classes: “ON” and
493 “OFF”, indicating that LED turned on or off, respectively (Fig. 4a).

494 **Model creation for YORU-based analysis**

495 We created YORU models using the “Training” package.-YORU randomly splits the
496 datasets containing the images and labels into two datasets: 80% for training and 20% for
497 validation. The models were trained with the training and validation datasets for 300 epochs.
498 If the training loss was low enough, training was finished before 300 epochs. For comparison
499 with human manual annotations, we created YORU models based on a YOLOv5s pre-trained
500 model, in which 2000 images were used. Human manual annotations were conducted using
501 BORIS⁵⁰. For model evaluation using “Evaluation” package, we extracted the specified
502 number of images from each dataset and created each model based on YOLOv5 pre-trained
503 models.

504 **Comparison with YORU and human manual annotations in animal videos**

505 To obtain human manual annotation data, we analyzed the behaviors of flies (10
506 videos and ~40 min in total), ants (3 videos and ~90 min in total), and zebrafish (2 videos
507 and ~10 min in total), following the behavior definitions (Supplementary table 5). Four
508 parameters, Accuracy, Precision, Recall, and F1 score, were used to compare the
509 performances between the human manual annotation and YORU. To obtain values for these
510 parameters, each annotation was first classified as follows:

- 511 • True Positive (TP): Correct annotation.
512 • False Positive (FP): Incorrect annotation, such as annotating a non-existing object or a
513 misplaced annotation.
514 • False Negative (FN): Undetected ground-truth.
515 • True Negative (TN): Correct no-annotation.

516 Subsequently, Accuracy, Precision, Recall, and F1 score were calculated as follows:

$$517 \quad \text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{FN} + \text{TN}}$$

$$518 \quad \text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

$$519 \quad \text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

$$520 \quad \text{F1 score} = \frac{2 \times \text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}}$$

521 **Evaluation of YORU models**

522 We evaluated the models using YORU's "Evaluation" package. Several hundred
523 images, which were not used in the model creation, were manually labeled with ground-truth
524 bounding boxes using LabelImg. YORU then predicted bounding boxes on the same images
525 using the models. By comparing the ground-truth and predicted boxes, we calculated the
526 Precision and Recall of the models. In addition, intersection over union (IOU) and average
527 precisions (AP), two of the typical object detection model indexes, were used ²⁸. IOU
528 represents how close the ground-truth box and the predicted box are, based on Jaccard Index
529 that evaluates the overlap between two bounding boxes, and AP represents the accuracy of
530 the behavior classes, reflecting the "Precision" and "Recall" values ²⁸. Since the AP is
531 affected by the IOU threshold, we used two different threshold values to calculate the APs:
532 "AP@50" with a threshold value of IOU = 50% and "AP@75" with a threshold value of IOU

533 = 75%. The AP ranges from 0 to 1, with a value of 1 indicating perfect consistency with the
534 ground-truth labels. As with the PASCAL Visual Object Classes Challenge ⁵¹, model
535 evaluation in YORU adopts the AP using the all-point interpolation method by the Riemann
536 integral to compare the models.

537 **Mice – running in virtual reality and recording neural activity**

538 All mice procedures followed institutional and national guidelines and were approved
539 by the Animal Care and Use Committee of Nagoya University. All efforts were made to
540 reduce the number of animals used and minimize the suffering and pain of the animals. The
541 “Mouse - treadmill” dataset consists of 3 videos of single mouse behaviors in the treadmill
542 environment. Referring to a previous study ²⁹, the videos were collected. C57BL/6J mice
543 were purchased from Nihon SLC. These animals were maintained in a temperature-
544 controlled room (24°C) under a 12 h light/dark cycle with ad libitum access to food and water.
545 Due to the body size needed for adaptation in head-fixed VR, mice aged 11–28 weeks were
546 used in the VR experiments. The VR environment used in this study was referred to in the
547 previous study ²⁹. Briefly, mice were head-fixed and free to run on a one-dimensional
548 treadmill. We recorded mice behavior with a monochrome CMOS camera (DMX33UX174,
549 The Imaging Source) at 30 fps and 640 × 480 pixels resolution.

550 For the “Mouse - treadmill” dataset, we labeled 2171 frames with eight behavior object
551 classes: “Running”, “Stop”, “Whisker-On”, “Whisker-Off”, “Eye-Open”, “Eye-Closed”,
552 “Grooming-On”, and “Grooming-Off” (Fig. 3b). “Eye-Open” and “Eye-Closed” indicate a
553 mouse that opens or closes the eye. “Grooming-On” was defined as the situation where the
554 forelegs of the mouse are touching its head. “Grooming-Off” was defined as all situations
555 that do not satisfy the criteria of “Grooming-On”. “Running” and “Stop” indicate a mouse
556 that runs or stops. “Running” was defined as the situation where the roller was rotating, and
557 “Stop” was defined as the situation where the roller was stationary. “Whisker-On” was
558 defined as the situation where the whisker was oriented in the anterior direction, and
559 “Whisker-Off” was defined as all situations that did not satisfy the criteria of “Whisker-On”.
560 Labels were randomly split into two datasets: creating a model dataset (1975 images) and a

561 test dataset (196 images). Using creating a model dataset, we generated a model based on the
562 YOLOv5s pre-trained model.

563 Surgery for the head plate implantation and virus injection were performed as
564 previously described ²⁹. Briefly, Retro-orbital virus injection, The AAV-PHP.eB-hSyn-
565 jGCaMP7f was injected at 100 μ L into the retro-orbital sinus of mice using a 30-gauge needle.
566 Head plate implantation was performed 14 days after the virus injection. Skin and membrane
567 tissue on the skull were carefully removed, and the surface was covered with clear dental
568 cement (204610402CL, Sun Medical). A custom-made metal head-plate was implanted onto
569 the skull for stable head fixation. During surgery, the eyes were covered with ofloxacin
570 ointment (0.3%) to prevent dry eye and unexpected injuries, and body temperature was
571 maintained with a heating pad. All procedures of surgery were performed under deep
572 anesthesia with a mixture of medetomidine hydrochloride (0.75 mg/kg; Nihon Zenyaku),
573 midazolam (4 mg/kg; Sandoz), and butorphanol tartrate (5 mg/kg; Meiji Seika). After surgery,
574 mice were injected with atipamezole hydrochloride solution (0.75 mg/kg; Meiji Seika) for
575 rapid recovery from the effect of medetomidine hydrochloride.

576 The recording of cortex-wide neural activity was performed as previously described ²⁹.
577 Briefly, we used a tandem lens design (a pair of Plan Apo 1 \times , WD=61.5 mm, Leica). To filter
578 the calcium independent artifacts ⁵², an alternating blue (M470L4, Thorlabs, Inc.) and violet
579 (M405L3, Thorlabs, Inc.) excitation LEDs were used as a light source for excitation,
580 combined with band-pass filters (86-352, Edmund; FBH400-40, Thorlabs, Inc., respectively).
581 Fluorescent emission was passed through a dichroic mirror (FF495-Di03) and filters
582 (FEL0500 and FESH0650, Thorlabs, Inc.) to a scientific CMOS (sCMOS) camera (ORCA-
583 Fusion, Hamamatsu Photonics). The timing of the excitation light was controlled by a global
584 exposure timing signal from the camera, processed with an FPGA-based logic circuit (Analog
585 discovery 2, Digilent), equipped with binary-counter IC (TC4520BP, Toshiba).

586 **Speed benchmarking – inference speed**

587 We measured the inference speed of the YOLOv5 detection using a custom Python
588 code of the YORU's detection function in the "Real-time Process" package. The list of
589 computers used for this analysis is shown in Supplementary Table 6. In the custom Python
590 code, the times before and after the YOLO detection step (t_1 and t_2 , respectively) were
591 logged by the "perf_counter()" function of the "Time" module. Then, we calculated the
592 inference speed of one frame detection as $t_2 - t_1$. For the analyses on the model size and
593 frame size dependency, we used the 640×480 pixels videos for "LED lightning - Small"
594 models and 1280×1024 pixels videos for "LED lightning - Large" models. 50000 frames
595 (60 fps, ~14 min, ~420 events) were used to calculate the inference speed.

596 **Speed benchmarking – real-time system**

597 To estimate the latency of the YORU's "Real-time Process" package, we measured the
598 end-to-end latency of the LED light detection task by running YORU on a Windows desktop
599 PC (CPU, Core i7-13700KF 16core; GPU, NVIDIA RTX 4080; RAM, 32GB or 16GB
600 DDR5). The CMOS camera (DFK 33UP1300, The Imaging Source Asia Co., Ltd or ELP-
601 UBFHD08S-MFV, Autocastle) equipped with a 50 mm focal length lens (MVL50M23,
602 Thorlabs, Inc.) captured an LED and streamed frames. The frames were then processed to
603 detect whether the LED state was on or off using the "LED lighting - Small" or "LED lighting
604 - Large" model. When YORU detected the ON state, it sent a signal to a trigger controller,
605 which then emitted a transistor-transistor logic (TTL) voltage pulse. As the trigger controller,
606 DAQ (USB-6008, National Instruments Co.) or a microcontroller (Arduino Uno, Arduino
607 CC) was used. A recording DAQ (USB-6212, National Instruments Co.) logged the TTL
608 voltage from the trigger controller (Fig. 4b). The voltage data were then processed using a
609 50 Hz low-pass filter ("butterworth_filter()" function of "scipy" package) to block the high-
610 frequency noises. The delay between the timing of LED voltage and that of the trigger TTL
611 was used to estimate the full-system latency, which includes overhead from hardware
612 communication and other software layers.

613 **Optogenetic assays – fly preparation**

614 *D. melanogaster* were raised as described above. Canton-S was used as a wild-type
615 strain. *UAS-GtACR1.d.EYFP (attP2)*³⁸ (RRID: BDSC_32194) and split-GAL4 strain that
616 labels pIP10 neurons specifically (*w*; VT040556-p65.AD; VT040347-GAL4.DBD)³⁶ (RRID:
617 BDSC_87691) were obtained from the Bloomington Drosophila Stock Center. *JO15-2-GAL4*
618⁵³ was a kind gift from Dr. D. F. Eberl (University of Iowa).

619 For optogenetic assays, pIP10 neurons specific *split-GAL4>GtACR1* male or *JO15-*
620 *>GtACR1* female flies were paired with wild-type adult males or females, respectively, as
621 mating partners. Flies used for the behavior assay were collected within 8 h after eclosion to
622 ensure their virgin status. Wild-type male, transgenic male and female flies were kept singly
623 in a plastic tube (1.5 mL, Eppendorf) containing ~200 µL fly food. Wild-type females were
624 kept in groups of 10 to 30. They were transferred to new tubes every 2 to 3 days, but not on
625 the day of the experiment. Males and females, 5 to 8 days after eclosion, were used for
626 experiments and were used only once. All experiments were performed between ZT = 1-11
627 at 25°C and 40-60% relative humidity.

628 **Optogenetic assays – dissection and immunolabeling**

629 Dissections of the male genitalia and female head were performed as described
630 previously with minor modifications¹¹. Briefly, male genitalia were dissected in phosphate-
631 buffered saline (PBS: Takara Bio Inc., #T900; pH 7.4 at 25°C), kept in 50% VECTASHIELD
632 mounting medium (Vector Laboratories, #H-1000; RRID: AB_2336789) in deionized water
633 for ~5 min, and mounted on glass slides (Matsunami Glass IND., LTD, Osaka, Japan) using
634 VECTASHIELD mounting medium.

635 Immunolabeling of the brains and ventral nerve cord was performed as described
636 previously with minor modifications⁵⁴. Briefly, brains were dissected in PBS (pH 7.4 at
637 25°C), fixed with 4% paraformaldehyde for 60-90 min at 4°C, and subjected to antibody
638 labeling. Brains were kept in 50% glycerol in PBS for ~1 h, 80% glycerol in deionized water
639 for ~30 min, and then mounted. Rabbit polyclonal anti-GFP (Invitrogen, #A11122; RRID:
640 AB_221569; 1:1000 dilution) was used for detecting the mCD8::GFP. Mouse anti-

641 Bruchpilot nc82 (Developmental Studies Hybridoma Bank, #nc82, RRID:AB_2314866;
642 1:20 dilution) was used to visualize neuropils in the brain. Secondary antibodies used in this
643 study were as follows: Alexa Fluor 488-conjugated anti-rabbit IgG (Invitrogen, #A11034;
644 RRID:AB_2576217; 1:300 dilution) and Alexa Fluor 647-conjugated anti-mouse IgG
645 (Invitrogen, #A21236; RRID: AB_2535805; 1:300 dilution).

646 **Optogenetic assays – confocal microscopy and image processing**

647 Serial optical sections were obtained at 0.84 μm intervals with a resolution of 512 × 512
648 pixels using an FV1200 laser-scanning confocal microscope (Olympus) equipped with a
649 silicone-oil-immersion 30× lens (UPLSAPO30XSIR, NA = 1.05; Olympus). Images of
650 neurons were registered to the *Drosophila* brain template⁵⁵ by using groupwise registration⁵⁶
651 with the Computational Morphometry Toolkit (CMTK) registration software. Brain
652 registration, image size, contrast, and brightness were adjusted using Fiji software (version
653 2.14.0; RRID: SCR_002285).

654 **Optogenetic assays – retinal feeding**

655 Transgenic male and female flies were maintained under a dark condition for 4-6 days
656 after eclosion and then transferred to a plastic tube (1.5 mL, Eppendorf) containing ~200 μL
657 of fly food. Plastic tubes that contain males were divided into experimental and control
658 groups. For the experimental group, 2 μL of all-trans-retinal (R2500, Sigma-Aldrich), 25
659 mg/mL dissolved in 99.5% ethanol (14033-80, KANTO KAGAKU), was placed on the food
660 surface. The male and female flies were kept on the food for 1 day and 2 days, respectively,
661 before being used for the assays.

662 **Optogenetic assays – a closed-loop system for event-triggered photostimulation**

663 For the event-triggered optogenetic assay (Fig. 5), we used the YORU system on a
664 desktop Windows PC (the same machine used in the speed benchmarking experiments).
665 Green LED light (M530L4, Thorlabs, Inc.) was used as the light source. The cameras used
666 to record the fly behavior and photostimulations, respectively, are as follows: a monochrome

667 CMSO camera (DMK33UX273, The Imaging Source Asia Co., Ltd.) equipped with a light-
668 absorbing and infrared transmitting filter (IR-82, FUJIFILM); a color CMOS camera (DFK
669 33UP1300, The Imaging Source Asia Co., Ltd). The monochrome CMOS and color CMOS
670 cameras were equipped with a 25 mm focal length lens (TC2514-3MP, KenkoTokina
671 Corporation) and zoom lens (MLM3X-MP, Computar), respectively. The fly behaviors were
672 recorded at a resolution of 640 by 480 pixels and a frame rate of 100 fps for ~40 min for each
673 fly pair using YORU. The recordings of the photostimulation were performed at a resolution
674 of 1280 x 1080 pixels resolution and 30 fps. The chamber was illuminated from the bottom
675 by an infrared LED light (ISL-150×150-II94-BT, 940 nm, CCS INC.) to enable recordings
676 in dark conditions. We used a round courtship chamber with a sloped wall (19 mm top
677 diameter, 12 bottom diameter, 4 mm height, and 6 mm radius fillet) made of the transparent
678 polylactic acid filament produced by a 3D printer (Sermoon D1, Creality 3D Technology Co.,
679 Ltd.). The chamber was enclosed with a slide glass and a white acrylic plate as a lid and
680 bottom, respectively.

681 For analyzing the behaviors, a male and a female fly were introduced into the chamber
682 by gentle aspiration without anesthesia. YORU captured and detected two behavior object
683 classes. When YORU detected a “wing extension”, a green light (light intensity (530 nm):
684 2.0 mW/cm²) illuminated the entire chamber. As a control experiment of the light condition,
685 we used an event-triggered light with a 1-second delay to illuminate the entire chamber (Fig.
686 5c). The light intensity of photostimulation was calibrated with an optical power meter
687 (PM100D, Thorlabs, Inc.).

688

689 Optogenetic assays – a closed-loop system for individual-focused photostimulation

690 For the individual-focused optogenetic assay (Fig. 6), we used the YORU system on
691 a desktop Windows PC (the same as for the speed benchmarking experiments). A projector
692 (HORIZON Pro, XGIMI Technology Co.) was used as a light source to stimulate specific

693 individuals. Camera captures, chambers, and background light setup were configured in the
694 same manner as in the event-triggered photostimulation experiments.

695 Before the experiments, the positions of the camera and projector planes were
696 calibrated using a circle grid pattern (5 circles height x 8 circles width). The
697 “findHomography” function of “OpenCV” package provided the homography matrix that
698 linearly transformed the position from the camera plane to the projector plane.

699 For analyzing the behaviors, a male and a female fly were introduced into the chamber
700 by gentle aspiration without anesthesia, and analyzed by YORU. When YORU detected a
701 “wing extension”, a circle-shaped illumination (~6.5 mm diameter with 1.14 mW/cm² light
702 intensity at 530 nm) was delivered to the individual that was detected as a “fly”. The center
703 coordinates of the predicted bounding boxes were homography transformed using the
704 homography matrix. As a control for the light condition, the projector displayed a patterned
705 light consisting of alternating 3-s of green and 4-s of black”, covering the entire chamber.
706 This was controlled using a custom Python code. This green/black time ratio was determined
707 based on the calculation of courtship duration of wild-type *Drosophila* in courtship assay
708 videos. The light intensity of photostimulation was calibrated with an optical power meter
709 (PM100D, Thorlabs, Inc.).

710 Quantification and statistical analysis

711 Statistical analyses were conducted using Jupyter (Python version 3.9.12) and RStudio
712 (R version 4.3.2). Aligned rank transform one-way analysis of variance (ART one-way
713 ANOVA) was performed to compare the wing extension ratio between conditions in the
714 event-triggered optogenetic assay. All statistical analyses were performed after verifying the
715 equality of variance (Bartlett’s test for three-groups comparisons; F-tests for two-groups
716 comparisons) and normality of the values (Shapiro-Wilk test). Kaplan-Meier curves were
717 generated using R, and a pairwise Log-rank test was performed to compare females’
718 cumulative copulation rates between conditions in the individual-focused optogenetic assay.
719 After ART one-way ANOVA or pairwise Log-rank tests, p values were adjusted using the

720 Benjamini-Hochberg method in the post hoc test. For the ART one-way ANOVA, the
721 ARTool package (version 0.11.1) was used (<https://github.com/mjskay/ARTool/>)^{57,58}. For
722 Kaplan-Meier curves and pairwise Log-rank test, the survival package (version 3.5.7) was
723 used (<https://github.com/therneau/survival>). Statistical significance was set at $p < 0.05$.
724 Boxplots were drawn using the R package ggplot2 (<https://ggplot2.tidyverse.org/>). Boxplots
725 represent the median and interquartile range (the distance between the first and third
726 quartiles), and whiskers denote $1.5 \times$ the interquartile range.

727

728 **Data and code availability**

729 • YORU software and all original code have been deposited at GitHub and are publicly
730 available at <https://github.com/Kamikouchi-lab/YORU>.

731

732 **Declaration of generative AI and AI-assisted technologies in the writing process**

733 During the preparation of this work the authors used DeepL (<https://www.deepl.com>) and
734 ChatGPT (<https://chatgpt.com>) in order to improve language. After using this tool, the
735 authors reviewed and edited the content as needed and take full responsibility for the content
736 of the publication.

737

738 **Acknowledgments**

739 We thank Mayu Yamanouchi for the software logo design; Masahiko Hibi and Shiori
740 Hosaka for collecting animal behavior videos; Ryota Nishimura (the Research Equipment
741 Development Group, Technical Center of Nagoya University) for the chambers used in
742 behavioral experiments, and construction virtual reality environment for mice; Matthew P.
743 Su, Kentaro Noma, Tomoka Nozaki and Haruka Ando for discussions; Kei Ito, Daniel F.

744 Eberl and the Bloomington Drosophila Stock Center (Indiana University, Bloomington, IN,
745 USA) for fly stocks; Developmental Studies Hybridoma Bank for antibodies. This study was
746 financially supported by the MEXT KAKENHI Grant-in-Aid for Transformative Research
747 Areas (A) “iPlasticity” (JP23H04228 to A.K.), Grant-in-Aid for Transformative Research
748 Areas (A) Hierarchical Bio-Navigation (JP22H05650 and JP24H01433 to R.T.), Grants-in-
749 Aid for Scientific Research (C) (JP23K05846 to R.T. and JP23K05845 to T.S.), Grant-in-
750 Aid for Early-Career Scientists (JP20K16464 to R.F.T. and JP21K15137 to R.T.) JST
751 FOREST (JPMJFR2147 to A.K.), the Japan Society for the Promotion of Science (JSPS)
752 Grant-in-Aid for JSPS Fellows (No. JP24KJ1290 to H.M.Y.) Japan.

753

754 **References**

- 755 1. Silk, M. J., Finn, K. R., Porter, M. A. & Pinter-Wollman, N. Can Multilayer Networks
756 Advance Animal Behavior Research? *Trends Ecol. Evol.* **33**, 376–378 (2018).
- 757 2. Han, J. *et al.* Analysis of social interactions in group-housed animals using dyadic
758 linear models. *Appl. Anim. Behav. Sci.* **256**, 105747 (2022).
- 759 3. Batista, G., Levine, J. D. & Silva, A. Editorial: The Neuroethology of Social Behavior.
760 *Front. Neural Circuits* **16**, 897273 (2022).
- 761 4. Bordes, J., Miranda, L., Müller-Myhsok, B. & Schmidt, M. V. Advancing social
762 behavioral neuroscience by integrating ethology and comparative psychology methods
763 through machine learning. *Neurosci. Biobehav. Rev.* **151**, 105243 (2023).
- 764 5. Mathis, A. *et al.* DeepLabCut: markerless pose estimation of user-defined body parts
765 with deep learning. *Nat. Neurosci.* **21**, 1281–1289 (2018).
- 766 6. Lauer, J. *et al.* Multi-animal pose estimation, identification and tracking with
767 DeepLabCut. *Nat. Methods* **19**, 496–504 (2022).
- 768 7. Pereira, T. D. *et al.* SLEAP: A deep learning system for multi-animal pose tracking.
769 *Nat. Methods* **19**, 486–495 (2022).
- 770 8. Bernstein, J. G., Garrity, P. A. & Boyden, E. S. Optogenetics and thermogenetics:
771 technologies for controlling the activity of targeted cells within intact neural circuits. *Curr.*
772 *Opin. Neurobiol.* **22**, 61–71 (2012).
- 773 9. Poth, K. M., Texakalidis, P. & Boulis, N. M. Chemogenetics: Beyond Lesions and
774 Electrodes. *Neurosurgery* **89**, 185 (2021).

- 775 10. Feno, L., Yizhar, O. & Deisseroth, K. The Development and Application of
776 Optogenetics. *Annu. Rev. Neurosci.* **34**, 389–412 (2011).
- 777 11. Yamanouchi, H. M., Tanaka, R. & Kamikouchi, A. Piezo-mediated
778 mechanosensation contributes to stabilizing copulation posture and reproductive success in
779 Drosophila males. *iScience* **26**, 106617 (2023).
- 780 12. Kane, G. A., Lopes, G., Saunders, J. L., Mathis, A. & Mathis, M. W. Real-time, low-
781 latency closed-loop feedback using markerless posture tracking. *eLife* **9**, 1–29 (2020).
- 782 13. Biderman, D. *et al.* Lightning Pose: improved animal pose estimation via semi-
783 supervised learning, Bayesian ensembling and cloud-native open-source tools. *Nat. Methods*
784 **21**, 1316–1328 (2024).
- 785 14. Villella, A. & Hall, J. C. Chapter 3 Neurogenetics of courtship and mating in
786 Drosophila. *Adv. Genet.* **62**, 67–184 (2008).
- 787 15. Burns-Cusato, M., Scordalakes, E. M. & Rissman, E. F. Of mice and missing data:
788 what we know (and need to learn) about male sexual behavior. *Physiol. Behav.* **83**, 217–232
789 (2004).
- 790 16. McGill, T. E. Sexual Behavior in Three Inbred Strains of Mice. *Behaviour* **19**, 341–
791 350 (1962).
- 792 17. Redmon, J., Divvala, S., Girshick, R. & Farhadi, A. You Only Look Once: Unified,
793 Real-Time Object Detection. *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. CVPR* (2016)
794 doi:10.1145/3243394.3243692.
- 795 18. Jocher, G. YOLOv5 by Ultralytics. <https://doi.org/10.5281/zenodo.3908559> (2020).
- 796 19. Du, J. Understanding of Object Detection Based on CNN Family and YOLO. *J. Phys.*
797 *Conf. Ser.* **1004**, 012029 (2018).
- 798 20. A. Aziz, Z., Naseradeen Abdulqader, D., Sallow, A. B. & Khalid Omer, H. Python
799 Parallel Processing and Multiprocessing: A Rivew. *Acad. J. Nawroz Univ.* **10**, 345–354
800 (2021).
- 801 21. Amino, K. & Matsuo, T. *Automated Behavior Analysis Using a YOLO-Based Object*
802 *Detection System. Behavioral Neurogenetics* vol. 181 (2022).
- 803 22. Yamanouchi, H. M., Tanaka, R. & Kamikouchi, A. Event-triggered feedback system
804 using YOLO for optogenetic manipulation of neural activity. *2023 IEEE Int. Conf. Pervasive*
805 *Comput. Commun. Workshop Affil. Events PerCom Workshop BiRD 2023* 312–315 (2023)
806 doi:10.1109/PerComWorkshops56833.2023.10150245.
- 807 23. Negroni, M. A. & LeBoeuf, A. C. Metabolic division of labor in social insects. *Curr.*
808 *Opin. Insect Sci.* **59**, 101085 (2023).
- 809 24. Piqueret, B. & d’Ettorre, P. Communication in Ant Societies. in *The Cambridge*
810 *Handbook of Animal Cognition* (eds. Kaufman, A. B., Kaufman, J. C. & Call, J.) 36–55
811 (Cambridge University Press, Cambridge, 2021). doi:10.1017/9781108564113.004.

- 812 25. Stednitz, S. J. *et al.* Forebrain Control of Behaviorally Driven Social Orienting in
813 Zebrafish. *Curr. Biol.* **28**, 2445-2451.e3 (2018).
- 814 26. Hosaka, S., Hosokawa, M., Hibi, M. & Shimizu, T. The zebrafish cerebellar neural
815 circuits are involved in orienting behavior. *eNeuro* (2024) doi:10.1523/ENEURO.0141-
816 24.2024.
- 817 27. Mathis, M. W. & Mathis, A. Deep learning tools for the measurement of animal
818 behavior in neuroscience. *Curr. Opin. Neurobiol.* **60**, 1–11 (2020).
- 819 28. Padilla, R., Passos, W. L., Dias, T. L. B., Netto, S. L. & Da Silva, E. A. B. A
820 comparative analysis of object detection metrics with a companion open-source toolkit.
821 *Electron. Switz.* **10**, 1–28 (2021).
- 822 29. Takeuchi, R. F. *et al.* Posteromedial cortical networks encode visuomotor prediction
823 errors. 2022.08.16.504075 Preprint at <https://doi.org/10.1101/2022.08.16.504075> (2024).
- 824 30. Petersen, C. C. H. Sensorimotor processing in the rodent barrel cortex. *Nat. Rev.
825 Neurosci.* **20**, 533–546 (2019).
- 826 31. Warren, J. A. *et al.* The HIV-1 latent reservoir is largely sensitive to circulating T
827 cells. *eLife* **9**, e57246 (2020).
- 828 32. Vincent, S. B. *The Function of the Vibrissae in the Behavior of the White Rat.* (Holt,
829 1912).
- 830 33. Sofroniew, N. J., Cohen, J. D., Lee, A. K. & Svoboda, K. Natural Whisker-Guided
831 Behavior by Head-Fixed Mice in Tactile Virtual Reality. *J. Neurosci.* **34**, 9537–9550 (2014).
- 832 34. Lopes, G. *et al.* Bonsai: An event-based framework for processing and controlling
833 data streams. *Front. Neuroinformatics* **9**, 1–14 (2015).
- 834 35. Ishimoto, H. & Kamikouchi, A. Molecular and neural mechanisms regulating sexual
835 motivation of virgin female Drosophila. *Cell. Mol. Life Sci.* **78**, 4805–4819 (2021).
- 836 36. Ding, Y. *et al.* Neural Evolution of Context-Dependent Fly Song. *Curr. Biol.* **29**,
837 1089-1099.e7 (2019).
- 838 37. von Philipsborn, A. C. *et al.* Neuronal Control of *Drosophila* Courtship Song. *Neuron*
839 **69**, 509–522 (2011).
- 840 38. Mohammad, F. *et al.* Optogenetic inhibition of behavior with anion
841 channelrhodopsins. *Nat. Methods* **14**, 271–274 (2017).
- 842 39. Lillvis, J. L. *et al.* Nested neural circuits generate distinct acoustic signals during
843 *Drosophila* courtship. *Curr. Biol.* **34**, 808-824.e6 (2024).
- 844 40. Krakauer, J. W., Ghazanfar, A. A., Gomez-Marin, A., MacIver, M. A. & Poeppel, D.
845 Neuroscience Needs Behavior: Correcting a Reductionist Bias. *Neuron* **93**, 480–490 (2017).
- 846 41. Lopes, G. & Monteiro, P. New Open-Source Tools: Using Bonsai for Behavioral
847 Tracking and Closed-Loop Experiments. *Front. Behav. Neurosci.* **15**, (2021).

- 848 42. Buccino, A. P. *et al.* Open source modules for tracking animal behavior and closed-
849 loop stimulation based on Open Ephys and Bonsai. *J. Neural Eng.* **15**, 055002 (2018).
- 850 43. Wu, M.-C. *et al.* Optogenetic control of selective neural activity in multiple freely
851 moving Drosophila adults. *Proc. Natl. Acad. Sci.* **111**, 5367–5372 (2014).
- 852 44. Watanabe, Y. *et al.* High-speed 8-bit image projector at 1,000 fps with 3 ms delay. in
853 vol. 3 1421–1422 (2015).
- 854 45. Kagami, S. & Hashimoto, K. Interactive Stickies: Low-latency projection mapping
855 for dynamic interaction with projected images on a movable surface. in *ACM SIGGRAPH*
856 *2020 Emerging Technologies* 1–2 (Association for Computing Machinery, New York, NY,
857 USA, 2020). doi:10.1145/3388534.3407291.
- 858 46. Casado-García, A. *et al.* LabelStoma: A tool for stomata detection based on the
859 YOLO algorithm. *Comput. Electron. Agric.* **178**, 105751 (2020).
- 860 47. Walter, T. & Couzin, I. D. Trex, a fast multi-animal tracking system with markerless
861 identification, and 2d estimation of posture and visual fields. *eLife* **10**, 1–73 (2021).
- 862 48. Simon, J. C. & Dickinson, M. H. A New Chamber for Studying the Behavior of
863 Drosophila. *PLOS ONE* **5**, e8793 (2010).
- 864 49. Terayama, M., Kubota, S. & Eguchi, K. *Encyclopedia of Japanese Ants*.
- 865 50. Friard, O. & Gamba, M. BORIS: a free, versatile open-source event-logging software
866 for video/audio coding and live observations. *Methods Ecol. Evol.* **7**, 1325–1330 (2016).
- 867 51. Everingham, M. *et al.* The Pascal Visual Object Classes Challenge: A Retrospective.
868 *Int. J. Comput. Vis.* **111**, 98–136 (2015).
- 869 52. Allen, W. E. *et al.* Global Representations of Goal-Directed Behavior in Distinct Cell
870 Types of Mouse Neocortex. *Neuron* **94**, 891–907.e6 (2017).
- 871 53. Dolezal, D. M., Joiner, M. A. & Eberl, D. F. Two distinct functions of Lim1 in the
872 Drosophila antenna. *MicroPublication Biol.* (2024) doi:10.17912/micropub.biology.001229.
- 873 54. Yamada, X. D. *et al.* GABAergic local interneurons shape female fruit fly response
874 to mating songs. *J. Neurosci.* **38**, 4329–4347 (2018).
- 875 55. Bogovic, J. A. *et al.* An unbiased template of the Drosophila brain and ventral nerve
876 cord. *PLOS ONE* **15**, e0236495 (2020).
- 877 56. Avants, B. & Gee, J. C. Geodesic estimation for large deformation anatomical shape
878 averaging and interpolation. *NeuroImage* **23**, S139–S150 (2004).
- 879 57. Wobbrock, J. O., Findlater, L., Gergle, D. & Higgins, J. J. The aligned rank transform
880 for nonparametric factorial analyses using only anova procedures. 143–146 (2011)
881 doi:10.1145/1978942.1978963.
- 882 58. Kay, M., Elkin, L. A., Higgins, J. J. & Wobbrock, J. O. ARTTool: Aligned rank
883 transform for nonparametric factorial anovas. R package version 0.11.1,
884 https://github.com/mjskay/ARTTool. 594511 (2021) doi:10.5281/zenodo.594511.

885 **Supplemental file legends**

886 **Supplementary Table 1: Summary of model accuracy compared to human annotations**

887 **Supplementary Table 2: Summary of model Precision and Recall in each pretrained**
888 **model and in each training image number**

889 **Supplementary Table 3: Summary of model accuracy of “mouse – treadmill” model**

890 **Supplementary Table 4: Summary of inference speed of one frame detection**

891 **Supplementary Table 5: Summary of dataset conditions**

892 **Supplementary Table 6: Lists of PC hardware information for benchmarking tests**

Figure 1

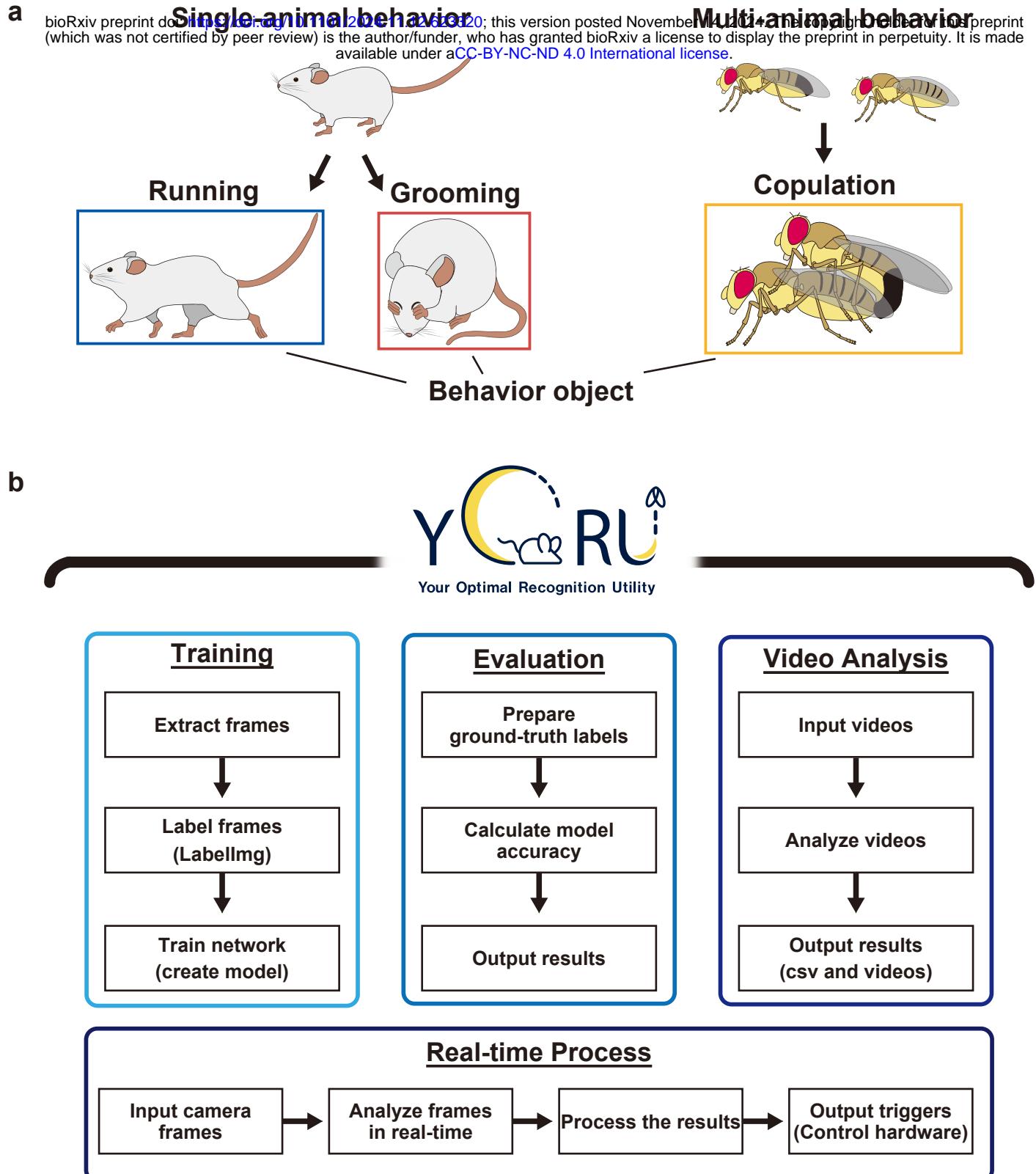


Fig. 1: YORU detects animal behaviors as a behavior object.

a, Illustrations of behavior objects. YORU can adapt to single-animal (Left) and multi-animal (Right) behaviors, including social behaviors. Animal behaviors (Top) can be classified as behavior objects (Bottom).

b, Diagram of YORU. YORU includes four packages: Training, Evaluation, Video Analysis, and Real-time Process.

Figure 2

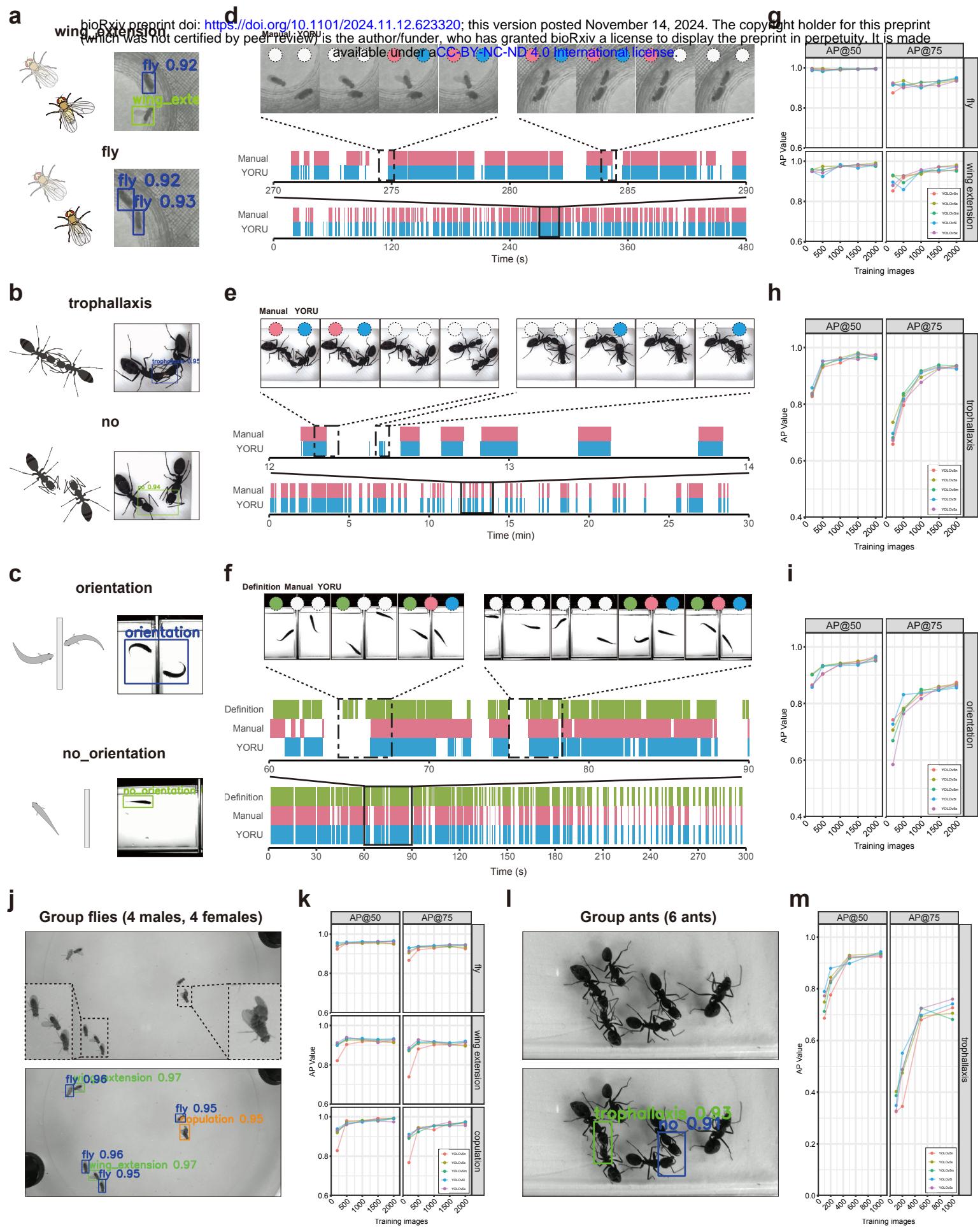


Fig. 2: Detection of animal behaviors by YORU

- bioRxiv preprint doi: <https://doi.org/10.1101/2024.11.12.623320>; this version posted November 14, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.
- a**, “Fly - wing extension” dataset. Two behavior classes were defined: “wing_extension”, “early_extending” one of its wings, and “fly” ; a fly that does not belong to the “wing extension” class.
- b**, “Ant - trophallaxis” dataset. Two behavior classes were defined: “trophallaxis” ; two ants showing nutrition exchange, and “no” ; the situation of no nutrition exchange but the heads of the two ants are close together.
- c**, “Zebrafish - orientation” dataset. Two behavior classes were defined: “orientation” ; two zebrafish showing orientation behavior, and “no_orientation” ; zebrafish showing no orientation behavior.
- d-f**, Ethogram of the fly wing extension (**d**), ant trophallaxis (**e**), and zebrafish orientation (**f**) behaviors. Top panels show the example view of each behavior. The circles show detections by manual analysis (pink), YORU analysis (blue), or analysis using a previous tracking method (green). The horizontal axes in the ethogram represent the observation period. The colored area of the ethogram shows the occurrence of the behavior detected with YORU analysis (blue), manual analysis using BORIS (pink), and the previous tracking method (green).
- g-i**, The AP values of “Fly - wing extension” (**g**), “Ant - trophallaxis” (**h**), and “Zebrafish - orientation” (**i**) models. The values at IOU=50% (AP@50) (Left) and IOU=75% (AP@75) (Right) are shown. The horizontal and vertical axes show the number of images used for training and the AP values, respectively. Each colored line shows the AP values of different pre-trained models (Also in **k, m**).
- j**, “Fly - group courtship” datasets. Three behavior classes were defined: “copulation” ; a pair of flies showing copulation, “wing_extension” ; a fly extending one of its wings, and “fly” ; a fly that does not belong to the “copulation” and “wing extension” classes.
- k**, The AP values for “Fly - group courtship” models.
- l**, “Ant - group trophallaxis” dataset. Two behavior classes were defined: “trophallaxis” ; two ants showing nutrition exchange, and “no_trophallaxis” ; the heads of two ants are close together without exchanging nutrition.

Figure 3

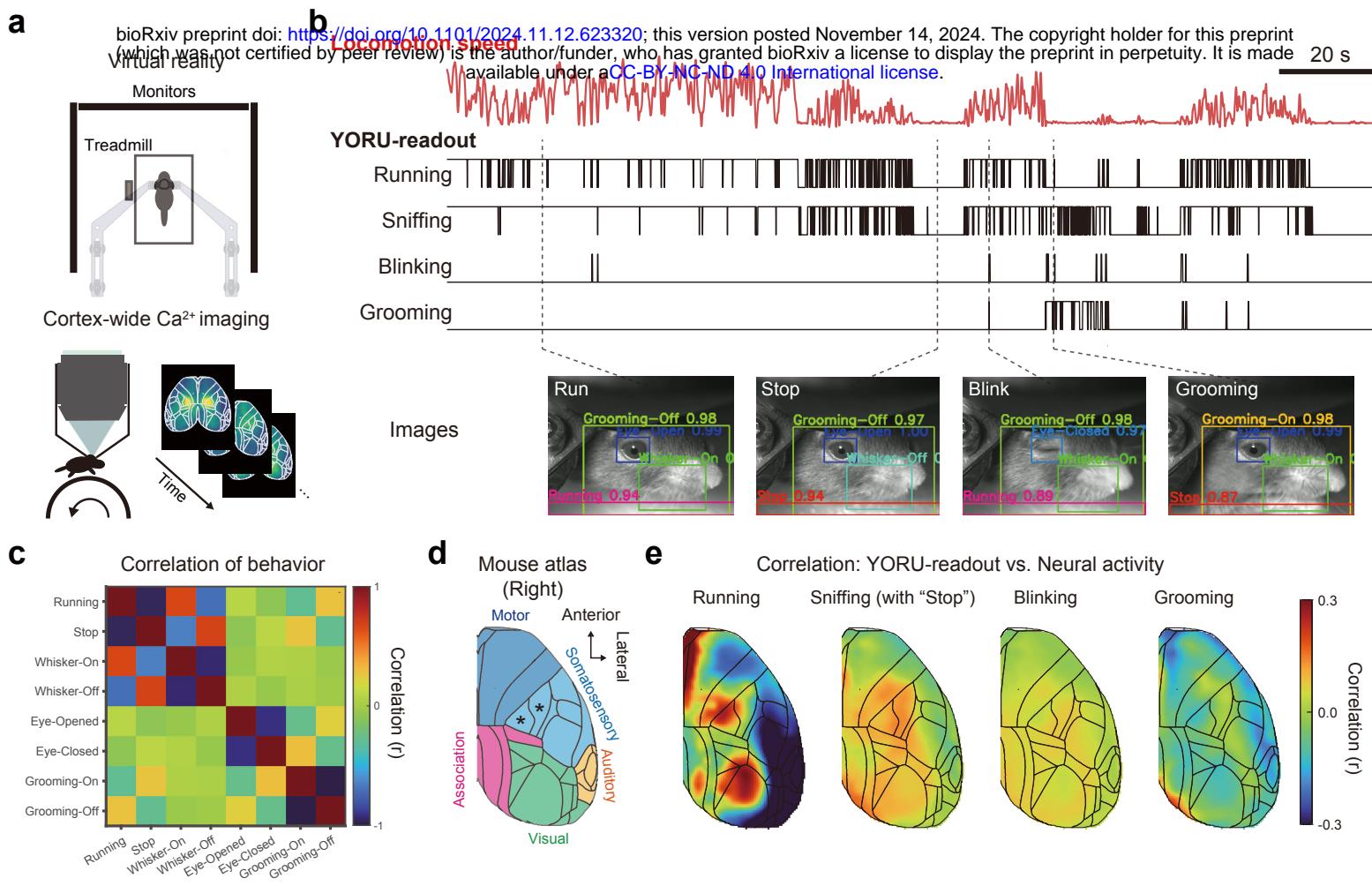


Fig. 3: YORU uncovers the relationship between behavioral readouts and neural activity interpretation.

a, The setup for virtual reality (VR) and cortex-wide imaging in mice.

b, Time-series of locomotion speed (red, calculated from rotary encoder signals of VR) and YORU readout (black). Representative images are shown at the bottom.

c, Correlation matrix between each behavior. Correlation indices are derived from the YORU readout time series data.

d, Top view of the Allen Common Coordinate atlas of the dorsal cortex. Five rough divisions are auditory areas (yellow), association areas (magenta), somatosensory areas (cyan), and motor areas (blue). Black asterisks show the somatosensory areas representing forelimb and hindlimb information.

e, Pseudo-colormap of Spearman's correlation coefficient (YORU-readout vs. Neural activity of each pixel).

Figure 4

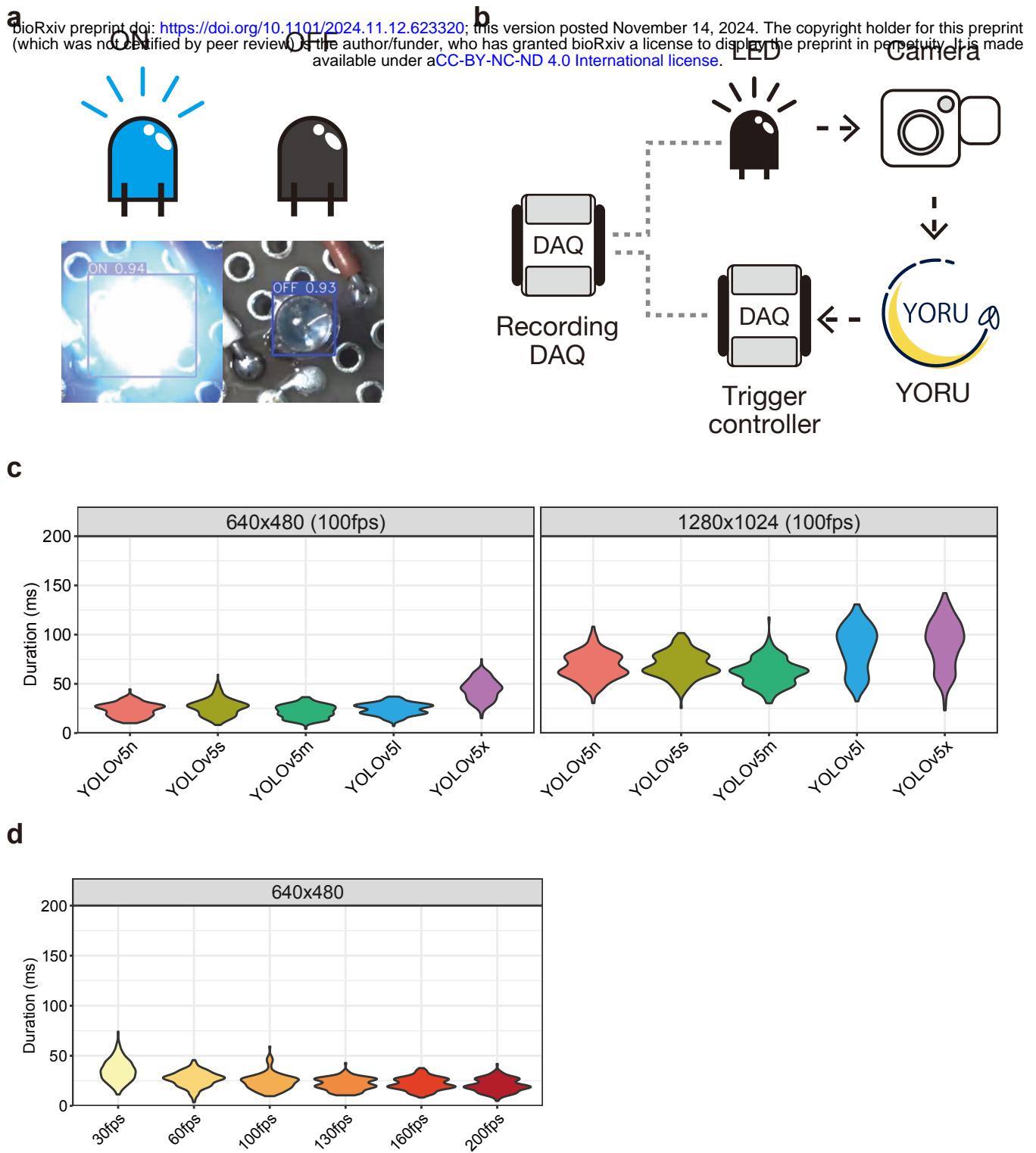


Fig. 4: Validation of YORU's operation speed.

a, “LED lighting” dataset. Two state classes were defined: “ON” and “OFF”, indicating that LED light was turned on and off, respectively.

b, Schematic of system latency measurements. The camera captures the LED light, YORU detects a frame, and the trigger-controller DAQ outputs the TTL voltage based on detection by YORU. The recording DAQ logged the TTL pulse from the trigger controller DAQ and the LED voltage.

c, The system latency of “LED lighting - Small” (Left) and “LED lighting - Large” (Right) models. The system latency of each model was calculated using camera images with resolutions of 640x480 pixels and 1240x1024 pixels, respectively.

d, The system latency at different camera frame rates. The “LED lighting - Small” model was used.

c,d, Violin plot represents the probability density of individual data points within the range of possible values.

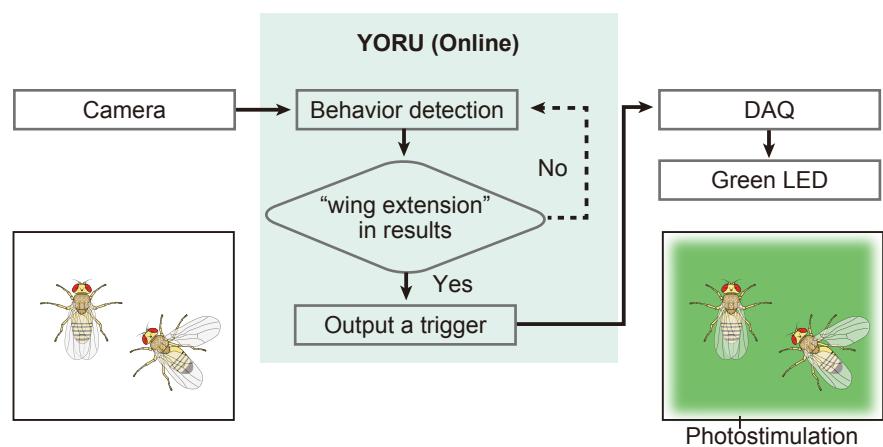
Figure 5

a

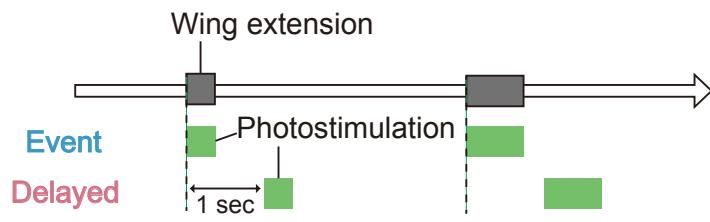


b

bioRxiv preprint doi: <https://doi.org/10.1101/2024.11.12.623320>; this version posted November 14, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

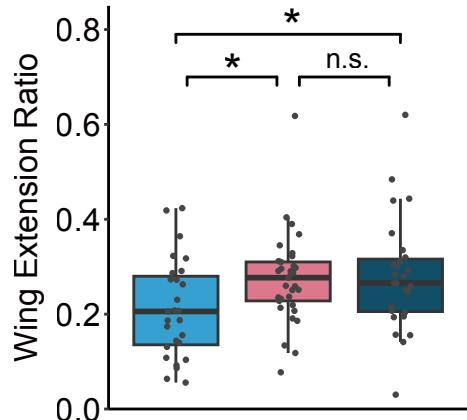


c



	Genotype	Light condition	Sample size
■	piP10 neurons > <i>GtACR1</i>	Event	27
■	piP10 neurons > <i>GtACR1</i>	Delayed	33
■	piP10 neurons > <i>GFP</i>	Event	27

d



e

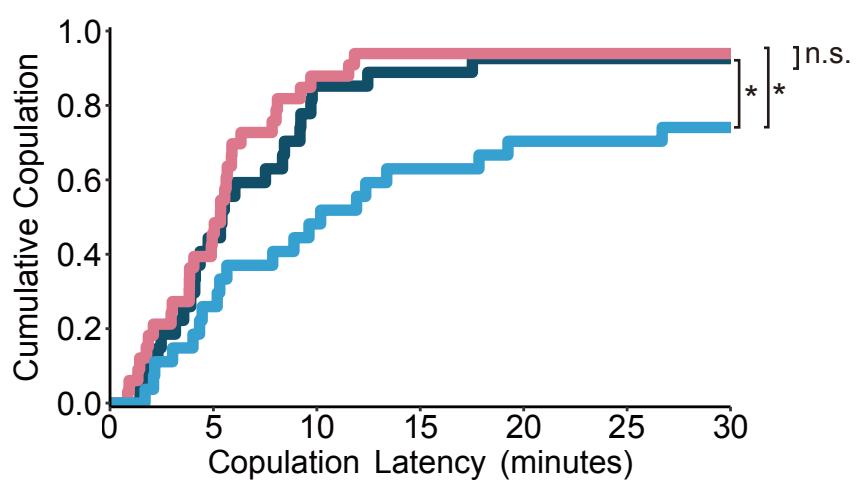


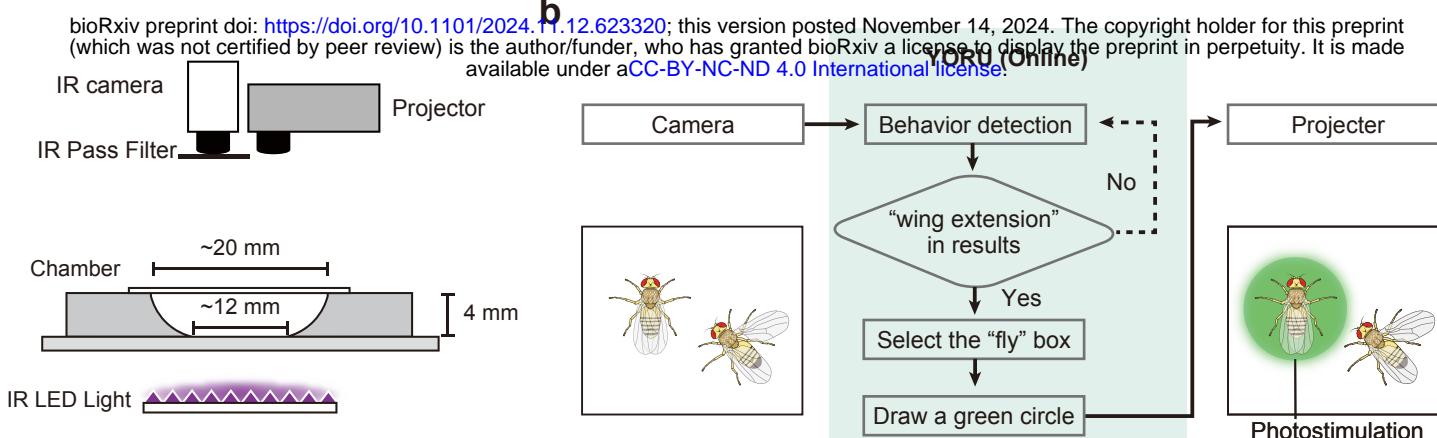
Fig. 5: Neural manipulation in response to male wing extension using YORU.

bioRxiv preprint doi: <https://doi.org/10.1101/2024.11.12.623320>; this version posted November 14, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

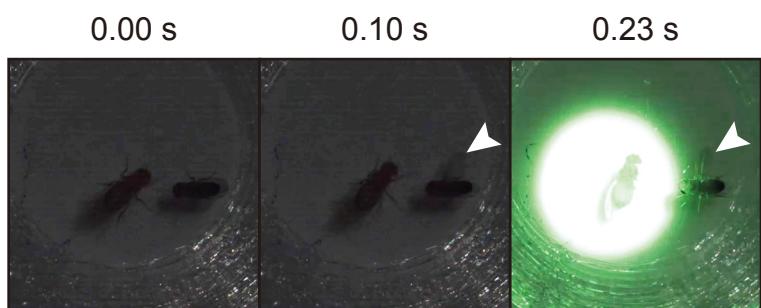
- a**, Microscopy image of pIP10 split-GAL4 expression pattern in a representative male brain and ventral nerve cord. Scale bar, 100 μ m. Signals of the GFP marker (green) and counter-labeling with the nc82 antibody (magenta) are shown.
- b**, Schematic of the YORU' s closed-loop conditions. YORU analyzes camera frames. Then, if YORU detects “wing extension” , it sends signals to the trigger controller (DAQ) and operates the LED light.
- c**, Schematic of light conditions. As the experimental photostimulation condition, YORU introduces green photostimulation to the entire chamber when it detects a fly showing wing extension (Event). As a photostimulation control, we used an event-triggered light with a 1-second delay that illuminates the entire chamber (Delayed). pIP10 specific split-GAL4>UAS-GtACR1 (pIP10 neurons > GtACR1) males were used for these two groups. In addition, pIP10 split-GAL4>20XUAS-IVS-mCD8::GFP (pIP10 neurons > GFP) males were used as a genetic control. The following color code was used: experimental group (blue), photostimulation control group (pink), and genetic control group (dark blue) (Also in **d, e**).
- d**, Ratio of time spent displaying wing extension before copulation. The aligned rank transform one-way analysis of variance (ART one-way ANOVA) test corrected with the Benjamini-Hochberg method was used for statistical analysis. Boxplots display the medians (horizontal white line in each box) with 25th and 75th percentiles and whiskers denote 1.5x the interquartile range. Each point indicates individual data.
- e**, Cumulative copulation rate of pIP10 split-GAL4>GtACR1 males. Pairwise comparisons using Log-Rank test corrected with the Benjamini-Hochberg method were used for statistical analysis.
- d,e**, Not significant (n.s.), $p > 0.05$; *, $p < 0.05$.

Figure 6

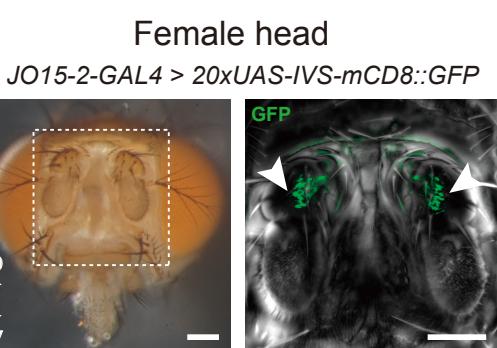
a



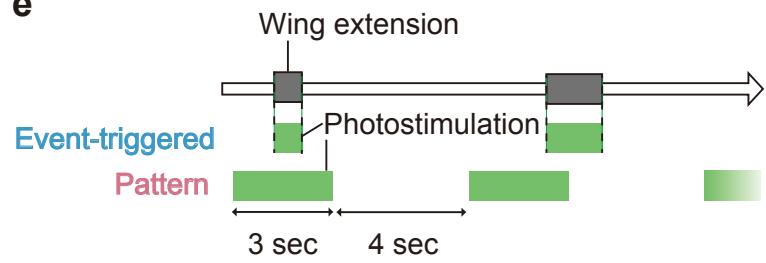
c



d



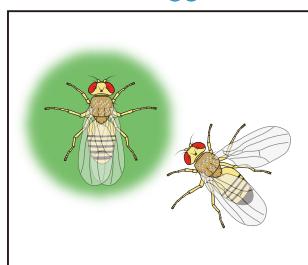
e



f

	Genotype	Light condition	Sample size
■	JO15>GtACR1	Event-triggered	28
■	JO15>GtACR1	Pattern	34
■	+>GtACR1	Event-triggered	32

Event-triggered



Pattern

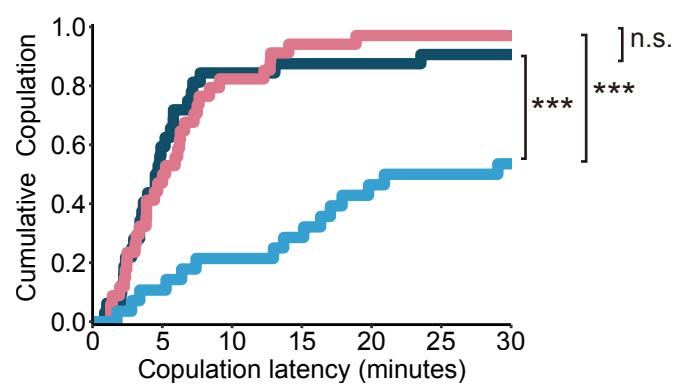
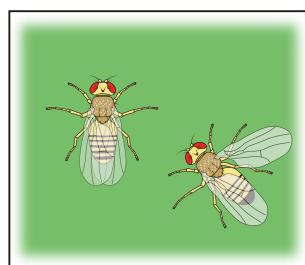


Fig. 6: Individual-specific neural manipulation by YORU.

bioRxiv preprint doi: <https://doi.org/10.1101/2024.11.12.623320>; this version posted November 14, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

- a,** Experimental setup of individual-focused photostimulation. An infrared (IR) camera is used to observe flies. IR LED light is used for IR camera recording. IR pass filter allows only IR light to be captured by the IR camera, preventing visible light noise. The projector is used for introducing individual-focused photostimulation.
- b,** Schematic of the closed-loop system for individual-focused photostimulation assay. YORU analyzes camera frames. If YORU detects “wing extension”, it draws a green circle on the “fly” bounding box and sends the image to the projector. The projector introduces green-circled photostimulation to the fly.
- c,** A representative situation during individual-focused photostimulation. White arrowheads show the fly displaying wing extension.
- d,** *JO15-2-GAL4* expression in female antenna. GFP markers driven by *JO15-2-GAL4* (*JO15-2-GAL4>20XUAS-IVS-mCD8::GFP*) are detected in Johnstons Organs. White arrowheads show JO neurons. D and V indicate the dorsal and ventral sides, respectively. Scale bar, 100 μ m.
- e,** Schematic of light conditions. In the experimental photostimulation condition, when YORU detects a fly showing wing extension, YORU introduces individual-focused photostimulation to the other fly (Event-triggered). In the photostimulation control, we used pattern light (3s On, 4s Off) to illuminate the entire chamber irrespective of the wing extension event (Pattern).
- f,** Cumulative copulation rate of *JO15-2>GtACR1* females. We used *JO15-2-GAL4>UAS-GtACR1* (*JO15-2>GtACR1*) females for the experimental group. As a genetic control, we used *+>UAS- GtACR1* (*+>GtACR1*) females. The following color code was used: experimental group (blue), photostimulation control group (pink), and genetic control group (dark blue). Pairwise comparisons using Log-Rank test corrected with the Benjamini-Hochberg method used for statistical analysis. Not significant (n.s.), p > 0.05; ***, p < 0.001.