Brief Investigation

Automated analysis of *C. elegans* behavior by *LabGym*: an open-source, AI-powered platform

L. Amanda Xu[1], Hongjiang Liu[2], Zhaoyu Li[2], X.Z. Shawn Xu[1,3], Bing Ye[1,4*], Yujia Hu[5*], and Elizabeth A. Ronan[6*]

[1] Life Sciences Institute, University of Michigan, Ann Arbor, Michigan, United States of America

[2] Queensland Brain Institute, University of Queensland, St Lucia, QLD 4067, Australia

[3] Department of Molecular and Integrative Physiology, Medical School, University of Michigan, United States of America

[4] Department of Cell and Developmental Biology, Medical School, University of Michigan, Ann Arbor, Michigan, United States of America

[5] Lerner Research Institute, Cleveland Clinic, Cleveland, Ohio, United States of America

[6] Department of Biologic and Materials Science, School of Dentistry, University of Michigan, Ann Arbor, Michigan, United States of America

*Correspondence: bingye@umich.edu (B.Y.); lizronan@umich.edu (E.A.R.); huy7@ccf.org (Y.H.)

ABSTRACT

The genetic tractability, well-mapped circuitry, and diverse behavioral repertoire of the nematode *C. elegans* make it an ideal model for physiological and behavioral studies. A wide range of methods has been developed for analyzing *C. elegans* behaviors, evolving with advances in technology such as videography and computer-assisted analysis. However, unlike organisms with distinct body features such as limbs and wings, the contour of *C. elegans* is rather uniform, posing unique challenges for automated analyses of *C. elegans* behavior. Here, we introduce *LabGym*—an open-source, artificial intelligence (AI)-based platform we recently developed—to the *C. elegans* research community. We trained deep learning models in *LabGym* capable of automatically categorizing and quantifying multiple user-defined parameters of worm locomotion behavior in multi-worm videos with high accuracy. Furthermore, we demonstrated their efficacy in quantifying locomotion changes in aging worms. Our work offers a cost-effective, user-accessible, and comprehensive approach to behavioral analysis in *C. elegans*.

INTRODUCTION

Research analyzing animal behavior has provided fundamental insights into all levels of biology (Altimus et al., 2020; Jorgensen and Mango, 2002; Pereira et al., 2020). Behavior serves as a powerful readout of an organism's physiological state. The model organism *Caenorhabditis elegans* is an established system for behavioral genetic studies, having led to groundbreaking discoveries in neurobiology, aging, and cell biology, among others (de Bono and Maricq, 2005; Herndon et al., 2002; Jorgensen and Mango, 2002; Xiao and Xu, 2021). Although *C. elegans* are relatively simple animals with a cylindrical body shape and a nervous system comprised of a mere 302 neurons, they are capable of exhibiting a rich repertoire of behaviors (de Bono and Maricq, 2005; Jorgensen and Mango, 2002). Worms rely on locomotion behaviors to survive and thrive as they effectively navigate their environments, find food sources, and avoid harmful stimuli (Bargmann, 2006; Croll, 1975; Gong et al., 2019; Li et al., 2011; Pierce-Shimomura et al., 1999; Wang et al., 2016). In the laboratory setting, worm behavior is primarily two-dimensional as worms are cultivated on an agar surface, further simplifying behavioral categorization and increasing their accessibility for ease of behavioral analysis.

Methods for analysis of worm behavior continue to evolve over time with technological advances. Upon the inception of *C. elegans* as a model organism in the 1960s, initial genetic screens relied on labor-intensive manual quantification of worm behaviors (Brenner, 1974; Jorgensen and Mango, 2002). As technology developed over the decades, the implementation of videography, along with the later incorporation of computer-assisted analysis, including the development of commercial software packages such as WormLab, has continued to improve and standardize worm behavior analysis methods (Datta et al., 2019; Feng et al., 2006; Husson, 2012; Pokala and Flavell, 2022; Zheng et al., 2012; Hsu et al., 2009; Broekmans et al., 2016; Huang et al., 2006; Roussel et al., 2014). Nevertheless,

most commercial software for behavioral analysis is not open-source and comes at a financial cost, limiting user accessibility and customizability.

Recently, AI-based machine learning approaches have garnered significant attention for their potential to automate analysis of animal behavior, promising to greatly reduce users' workload and improve reproducibility (Goss et al., 2024; Graving et al., 2019; Hu et al., 2023; Kabra et al., 2013; Mathis et al., 2018). However, the use of machine learning to automate *C. elegans* behavioral analysis has yet to achieve widespread use in the worm research community. Current machine learning methods to automatically analyze *C. elegans* behavior primarily rely on postural estimations or body skeleton analysis (Alonso and Kirkegaard, 2023; Hebert et al., 2021; Weheliye et al., 2025), but extra efforts are needed to transform these numerical outputs into ethologically relevant behaviors. More recently, unsupervised representation learning has been utilized for embedding raw pixels into motion representations (Deserno and Bozek, 2025). These mathematical representations of unsupervised approaches still require additional steps to be translated into interpretable behaviors, which may be difficult for researchers without extensive computational expertise. Furthermore, unlike insects, rodents, and primates with more complex body structures like limbs and/or wings, worms are more uniform in morphology (Altun et al., 2009). This gives rise to major technical challenges in automated *C. elegans* behavioral analysis, such as distinguishing head from tail to identify directional movements, and accurately tracking animals during visually challenging scenarios, including self-overlapping maneuvers like omega bends and physical entanglement of multiple worms. Although certain algorithms and mathematical models have been proposed to tackle these technical difficulties, most of these approaches merely solve some but not others (Broekmans et al., 2016; Huang et al., 2006) or require integration into pre-existing commercial software (e.g., WormLab) to be functional (Roussel et al., 2014).

To address these challenges, here we introduce a user-friendly, AI-based approach for *C. elegans* behavioral analysis utilizing *LabGym*, a software platform that we recently developed for customizable and efficient behavioral categorization and quantification (Goss et al., 2024; Hu et al., 2023). Unlike existing platforms that rely heavily on pose estimation or mathematical representations, which are not directly interpretable in a biological context, *LabGym* leverages advanced deep learning algorithms to holistically evaluate spatiotemporal changes and overall motion patterns of target objects—not only quantifying, but also categorizing, their behaviors. We trained and validated *LabGym* models to analyze acute *C. elegans* movements, including speed, trajectory, forward locomotion, reversals, omega turns, and head-twitching, with high accuracy. Our *LabGym*-based approach reliably tracks individuals even during self-overlap and multi-worm contact, while also distinguishing forward from reverse movement without explicit head-tail assignment. Additionally, we demonstrate *LabGym*'s utility to be tailored to specific experimental needs by analyzing worm behavior changes across aging. Furthermore, the AI-based models within *LabGym* are user-customizable, allowing experimenters to define, detect, and analyze diverse behaviors of interest with high accuracy and efficiency. All models and training datasets are shareable among users, ensuring reproducibility and transparency of behavioral analysis. This work provides the *C. elegans* research community with a powerful, generalizable, and user-friendly tool poised to significantly advance the study of behavior.

RESULTS

The pipeline for training Categorizers in *LabGym*

LabGym is an AI-based, open-source analysis software designed for the identification and quantification of user-defined behaviors across diverse animal species and experimental paradigms (Goss et al., 2024; Hu et al., 2023). To detect and holistically assess user-defined behaviors, *LabGym* employs deep, customizable neural networks trained through a supervised learning approach. A detailed user manual and the source code of *LabGym* are publicly available (https://github.com/umyelab/LabGym).

To classify and quantify user-defined behaviors, *LabGym* uses the Categorizer module, which applies a deep-learning model to categorize the behaviors of target individuals at every frame across user-specified time windows (Goss et al., 2024; Hu et al., 2023). To train a Categorizer, users begin by inputting training videos of individuals exhibiting the behaviors of interest into the Training Module of *LabGym*'s GUI (graphical user interface), which then extracts and identifies individual animals from the videos via one of two methods: background subtraction or the Detector module (Figure 1a). Users have the option of choosing either method. The latest release of *LabGym* incorporates an additional module named the Detector (Goss et al., 2024). Adapted from *Detectron2* (Wu et al., 2019), a deep-learning instance-segmentation package, the Detector is an integrated object-tracking component that utilizes a mask region-based convolutional neural network to generate masks for user-defined objects, efficient at object detection even with background noise and preventing identity switches during social proximity (Goss et al., 2024).

Once individuals are identified, a pair of behavioral data is generated for each individual over consecutive, user-defined time windows. Each data pair is comprised of a pixelated "animation" of the animal's appearance and postural changes as they appear in the inputted video, as well as the corresponding "pattern image", which encodes the animal's movement

pattern from initial to final position using a color gradient from blue to white, then pink (Figure 1b). This dual representation captures both broader locomotor patterns and fine-scale postural changes. Following the generation of behavior data pairs, users then sort the data pairs into user-defined behavior categories (by separating them into individual folders named by behavior type) to create the ground-truth dataset for Categorizer training (Figure 1c). As *LabGym* employs a supervised learning approach, this data-annotation step is essential for providing the ground-truth examples that guide the Categorizer in learning to recognize each target behavior type.

After establishing the ground-truth behavior dataset, it is used to train a Categorizer, which employs deep neural networks (the Animation Analyzer and the Pattern Recognizer) to acquire and process the two pieces of behavioral information at each frame of an analyzed video (Figure 1d; see Methods). The Animation Analyzer extracts spatiotemporal features from the animation, while the Pattern Recognizer interprets overall movement sequences from the pattern image. These two streams of information are then integrated by the Decision Maker, which assigns a behavior category to each video frame along with the event probability—a calculated percentage representing the likelihood that the given behavior was exhibited at that frame (Figure 1d).

Following behavior categorization, the Quantifier module integrates behavior category labels from the Categorizer and positional information of each tracked individual to calculate detailed quantitative measurements of behaviors exhibited at each frame, which are outputted into Excel files organized by individual animal and behavior type (Figure 1e). Additional outputs include raster plots color-coded by behavior category and corresponding probability, to enable temporal visualization of all behaviors exhibited by each individual over the analysis window (Figure 1e). *LabGym* also generates annotated videos of the original behavioral recordings, where each individual is labeled frame-by-frame with its classified behavior and corresponding event probability (Videos S2 and S3).

LabGym **accurately analyzes user-defined** ***C. elegans*** **locomotion behavior**

To test if *LabGym* can be applied to analyze *C. elegans* behavior, we attempted to train Detectors and Categorizers for identifying and quantifying worm locomotion, a prominent *C. elegans* behavior. Worms spend the majority of their time crawling forward, with occasional reversals typically followed by omega bends to reorient their movement (Parida, 2022; Croll, 1975; Pierce-Shimomura et al., 1999). However, unlike other more complex organisms, for example, insects and vertebrates, worms are relatively uniform in morphology (Altun, Zeynep F. et al., 2009). Such morphological simplicity poses challenges for automated analysis of *C. elegans* behaviors, such as distinguishing individual worms during self-overlap (e.g., omega bends) or overlap with other worms (e.g., entanglement), as well as differentiating the head from the tail of worms following direction change (e.g., reversal and post-omega turn. Therefore, besides basic parameters (e.g., trajectory and speed), we primarily focused on analyzing challenging parameters such as forward and reverse crawling, which typically require accurate head/tail distinction, and omega bends. We also attempted to analyze immobility and head-twitching, two behavioral states infrequent in younger adults but increasingly common in aged worms (see below).

As we analyzed individual-worm behaviors in video assays containing multiple worms, we first sought to resolve the challenge of multi-worm entanglement, as worms may collide with each other to make the detection of individual worms difficult. To do so, we trained a Detector named *multiple worm detector 5,* and found that it is capable of tracking individual worms even during multi-worm collisions (Video S3).

We then trained a Categorizer, *crawling categorizer 9.7_v2_12 frames*, which employs the Detector to classify locomotion parameters of N2 worms (Figure 2a; Video S1; Table S1), including forward crawling, reverse crawling, omega bend, immobile, and twitching (Figure

2a). We achieved an overall *LabGym* self-determined accuracy of 0.93. It should be noted that the five selected categories represent only a subset of locomotion behavior states that *LabGym* can analyze, as its Categorizers are fully customizable and allow experimenters to define behavior categories according to their research needs.

Although *LabGym*'s self-estimated accuracy of the Categorizer was 0.93, this value reflects performance in a simulated test environment (Hu et al., 2023). To directly assess the accuracy of the Categorizer in identifying and quantifying *C. elegans* behaviors, we compared the *LabGym*-computed behavioral outputs with human annotation of N2 worm locomotion (Figures 2b-d). Side-by-side raster plots of *LabGym* versus manual analysis outcomes revealed strong alignment between the two methods, evidence of *LabGym*'s categorization accuracy (Figure 2b). We then calculated the percentage of frames in which the behavioral events matched between the two approaches for each worm, and averaged across individuals, yielding an overall accuracy of 92.5% (Figure 2c; see Methods). This indicates that *LabGym* is estimated to successfully analyze 92.5% of behavioral events identified by human vision. To further evaluate the Categorizer's performance, we compared the total durations (in seconds) of each behavioral parameter computed by *LabGym* and manual scoring. The two methods showed only slight, non-significant differences, indicating close alignment of computed metrics across behavior categories (Figure 2d). This provides further validation of the Categorizer's accuracy.

LabGym **captures subtle behavioral changes in aging** ***C. elegans***

Having demonstrated the efficacy of *LabGym* in analyzing *C. elegans* locomotion behavior with high accuracy, we next sought to apply this approach to analyze biologically meaningful behavioral paradigms. One such application is in the context of aging, for which *C. elegans* is a well-established model system. The worm's short lifespan of approximately three weeks, genetic tractability, and well-characterized developmental timeline provide a powerful

platform for investigating age-dependent changes in locomotion behavior (Liu et al., 2013; Son et al., 2019).

Worm locomotion has been shown to decline rapidly with age. The forward crawling speed in exploratory behaviors typically peaks in early adulthood (Day 1-5) and progressively declines over time (Liu et al., 2013). In late age (Day 11 and beyond), functional declines in the motor nervous system and muscle lead to marked reductions in overall locomotion activity, increased immobility (Liu et al., 2013; Herndon et al., 2002; Herndon et al., 2018). Aging studies are longitudinal and require analyses at multiple time points across lifespan, making behavioral analysis laborious. Therefore, automated, AI-assisted detection of age-related behavioral changes would be a valuable tool in *C. elegans* aging studies.

To address this need, we applied our *LabGym* approach to analyze locomotion behavior in aging worms. As worms exhibit morphological changes when they age, notably in color and body appearance (Herndon et al., 2002; Garigan et al., 2002), this can compromise the analysis accuracy of Categorizers trained only on behavior examples from young worms. To address this, we trained an additional Categorizer, *crawling categorizer 11 – locomotion across aging*, using a higher proportion of behavior examples from aging worms. The Categorizer identifies the same five behavioral parameters as in our initial Categorizer and achieved a *LabGym*-reported classification accuracy of 0.97 (Video S1; Table S1). Using this model, we analyzed behavioral videos of aging worms across adult lifespan, repeating every other day until Day 15 adulthood. Older worms were not analyzed, as they began to die and exhibited little to no locomotion activity. For each video, *LabGym* automatically generated trajectory plots color-coded by individual worm, which revealed a decline in exploratory locomotion across age (Figure 3a), consistent with previous findings (Liu et al., 2013). Similar age-related reductions were also apparent in trajectory plots colored by time progression (Figure S1). These consistencies suggest that our Categorizers reliably capture age-associated changes in worm locomotion.

To further evaluate *LabGym*'s accuracy in assessing age-related behavioral changes, we quantified changes in individual behavior metrics across age (Figures 3b–g) and evaluated whether observed trends are consistent with prior research. Forward crawling speed (Figure 3b) and distance (Figure 3c) declined with age, in congruence with observed decreases in worm exploratory locomotion over time (Glenn et al., 2004; Liu et al., 2013). Reversal and omega bends frequencies peaked in mid-adulthood (Figures 3d and 3e), in line with reports of increased reorientation during these ages (Glenn et al., 2004; Wirak et al., 2022). In late-age adults, movement became more erratic and less coordinated, marked by elevated levels of immobility (Figure 3f) and bouts of head-twitching (Figure 3g). This may reflect a physiological shift from locomotion to near-complete immobility at late life, likely driven by motor nervous system decline and muscle degeneration (Glenn et al., 2004; Liu et al., 2013; Guo et al., 2012). In summary, we developed a Categorizer capable of detecting subtle, age-dependent behavioral changes in concordance with prior literature. This demonstrates the utility of *LabGym* as a robust tool for automated behavioral analysis in *C. elegans* aging studies.

DISCUSSION

In this report, we present the first demonstration of AI-based categorization and analysis of *C. elegans* behavior using *LabGym*. We trained Categorizers in *LabGym* capable of accurately identifying and quantifying worm locomotion, including subtle behavioral changes during aging. These models and the datasets for training them will be publicly available, allowing researchers to use them directly for deep phenotyping worm behavior and for benchmarking their own tools. Beyond the behavior categories and experiments described here, the accessibility and customizability of *LabGym* enable users to train models for analyzing *C. elegans* behaviors of interest across diverse backgrounds and settings, opening new avenues in biological research.

Our *LabGym*-based approach addresses important challenges in *C. elegans* behavioral analysis: head/tail assignment, single-worm entanglement (e.g., omega bends), and multi-worm entanglement. These technical hurdles arise from the morphological similarities between the head and tail, as well as the adoption of atypical overlapping worm body postures (e.g., omega bends) and conformations (e.g., multi-worm entanglement). Despite these challenges, *LabGym* reliably quantifies behavioral parameters with precision, a success we attribute to two main factors. First, *LabGym*'s holistic assessment approach evaluates an animal's overall motion patterns and spatiotemporal changes (Hu et al., 2023; Goss et al., 2024). Multiple parameters, such as crawling speed, body posture, bending patterns, and subtle morphological features, are integrated into the analysis criteria, allowing Categorizers to analyze visually challenging behavioral states, such as distinguishing forward from reverse crawling, as well as omega bends and multi-worm collisions. Second, the deep convolutional neural networks underlying *LabGym* are trained via a supervised learning approach, requiring human inputs to sort behavior examples for network training (Hu et al., 2023; Goss et al., 2024). The human-integrated component of *LabGym* capitalizes on the strengths of human vision and cognition to ensure that training datasets are validated

by experimenters themselves, offering an advantage over unsupervised methods that depend primarily on machine vision. Together, these elements enable *LabGym* to overcome significant challenges in *C. elegans* behavioral analysis.

While we only focused on a subset of locomotion parameters in our analyses to demonstrate the utility of *LabGym*, users can certainly define and analyze additional locomotion parameters using *LabGym*. Beyond locomotion behavior, *LabGym*'s user-customizability can be further applied to more complex locomotion-based behaviors, such as sensory behaviors (e.g., chemotaxis, phototaxis, phonotaxis, and electrotaxis) (Bargmann, 2006; Iliff et al., 2021; Liu et al., 2010; Wang et al., 2023; Ward et al., 2008; Gabel et al., 2007), as well as interactive behaviors like social feeding and mating in multi-worm contexts (de Bono and Bargmann, 1998; Liu and Sternberg, 1995).

Our work highlights not only the feasibility, but also the potential, of *LabGym* for behavior-based applications in *C. elegans* research. The open-source availability and user-friendly nature of the platform lower the technical barrier of its use, which may encourage researchers to customize and apply *LabGym* to their own behavioral studies.

MATERIALS AND METHODS

C. elegans strains cultivation and maintenance

The wildtype *C. elegans* strain N2 was used for all experiments. Worms were maintained at 20°C on nematode growth medium (NGM) with *Escherichia coli* strain OP50. Age-synchronization was performed by transferring L4-stage worms onto nematode growth medium (NGM) plates seeded with OP50 and allowed to develop to Day 1 of adulthood (D1). All experiments were performed on adult worms.

C. elegans behavioral assays

To assess behavior in D1 young adult worms, each trial consisted of transferring 5-10 age-synchronized D1 worms to an NGM testing plate without OP50, using an eyelash to minimize stimulation. Worms were allowed to acclimate for 2 minutes prior to behavior recording. Videography was performed using the WormLab system (MBF Bioscience). For experiments examining behavioral changes with aging, ~300 age-synchronized L4-stage worms were transferred onto NGM maintenance plates containing 50 μM fluoro-2′-deoxyuridine (FuDR) to prevent progeny production. Behavior was recorded on days 1, 5, 7, 9, 11, 13, and 15 of adulthood. At each time point, 10–25 worms were placed on unseeded NGM testing plates and behavior was recorded as described above. Worms were not reused after behavioral recordings.

Computational hardware

The computational procedures in this study were performed on the following computer system: Dell Enterprise Server with Microsoft Windows 11 Education, 3.10 GHz Intel Xeon w3-2435 with 64 GB RAM, NVIDIA RTX 4500 Ada Generation with 24 GB VRAM.

Behavioral video processing

Behavioral videos (video format: .avi) to assess D1 adult behaviors were 2056 x 2056 pixels and 8 frames per second, and downsized to 4 frames per second using the *LabGym* Preprocess Videos submodule in the Preprocessing Module. During analysis, the framesize was further downsized using the Analysis Module of *LabGym*.

For experiments examining age-related behavior changes, videos (video format: .avi) were 2456 x 2052 pixels and 7.5 frames per second, and downsized to 2056 x 1716 pixels and 4 frames per second using the *LabGym* Preprocess Videos submodule in the Preprocessing Module. During analysis, the framesize was further downsized using the Analysis Module of *LabGym*.

Detector training in *LabGym*

Adapted from *Detectron2*, an instance-segmentation package, the Detector is an integrated object tracking component of *LabGym*, efficient at object detection and preventing identity switches during social proximity. In the Generate Behavior Examples submodule of the Training Module, users may choose to use a trained Detector to detect animals and generate corresponding behavior examples for training a Categorizer. In the Analysis Module of *LabGym*, users may also choose to implement a trained Detector to assist the Categorizer in detecting objects and animals of interest during behavioral analysis. More details on the mechanisms of the Detector have been reported previously (Goss et al., 2024).

Like Categorizers, Detectors must be trained. The Detector used to track worms and generate the behavioral data reported in this study, *multiple worm detector 5*, was trained using the Train Detectors submodule embedded within the Training Module of *LabGym*. To train, images were generated from multiple behavioral videos using the “Generate Image

Examples" feature of the Train Detectors submodule, and individual worms from the multi-worm images were annotated and defined as a single class "worm" in *Roboflow* (Dwyer et al., 2025). 501 annotated worm images were then used for training the Detector with an inferencing framesize of 960 and 1,000 training iterations.

Categorizer training in *LabGym*

Two Categorizers trained in *LabGym* were used to analyze the worm behavioral data reported in this study. All Categorizers were trained using the Train Categorizers submodule embedded within the Training Module of *LabGym*. Training examples were augmented via six methods built into *LabGym*'s Train Categorizers submodule: random rotation, horizontal flipping, vertical flipping, random brightening, random dimming, and random deletion. All Categorizers were trained on behavior examples generated in the "Non-interactive" mode, specified to identify behaviors of non-interacting individual worms. Categorizers incorporated both the *Animation Analyzer* (complexity level 2) and *Pattern Recognizer* (complexity level 3) submodules. The training settings that resulted in the highest average macro categorization accuracy were retained: Animation Analyzer (LV2, 16 x 16 x 1) and Pattern Recognizer (LV3, 32 x 32 x 3). The performance metrics of each trained Categorizer, including categorization accuracy, precision, recall, and f1 score, were calculated on the Categorizer's ground-truth testing data (sorted behavior examples) using the Test Categorizers submodule embedded within the Training Module. The descriptions and specific methods of calculating each type of performance metric have been previously reported in detail (Goss et al., 2024; Hu et al., 2023).

Criteria for sorting the behavior examples

Young adult (D1) N2 worm locomotion (Figure 2)

The Categorizer *crawling categorizer 9.7_v2_12 frames* analyzed D1 N2 worm locomotion, and outputs were used to compare *LabGym*'s accuracy with manual analysis. To train the

Categorizer, we sorted “Non-interactive” mode behavior examples (12 frames in duration) into behavior categories. Most behavior examples were extracted from 4 fps (frames per second) videos of D1 N2 worm locomotion, with a small proportion of examples from 4 fps videos of aging N2 worm locomotion using the *multiple worm detector 5* in the Generate Behavior Examples submodule. The criteria and number of behavior examples used to train each behavior category are described as follows (1 example pair consists of one Animation and its corresponding Pattern Image):

Forward crawling (964 example pairs): The worm crawls in the forward direction, where sinusoidal body waves propagate from the tail toward the head. Typically slower and longer in duration than reversals.

Reverse crawling (238 example pairs): The worm crawls in the reverse direction, with typically deeper sinusoidal body waves than forward locomotion that propagate from the head toward the tail. Typically brief and often succeeded by an omega bend.

Omega bend (79 example pairs): The worm exhibits a deep, omega-shaped body bend where the head touches the tail, typically followed by a resumption of forward crawling in a new direction.

Immobile (139 example pairs): The worm ceases active locomotion completely.

Twitching (227 example pairs): The worm displays rapid, small-amplitude body sweeps usually localized in the head.

Locomotion of aging N2 worms (Figure 3)

The Categorizer *crawling categorizer 11 – locomotion across aging* was used to analyze behavioral changes of aging N2 worms. To train the Categorizer, we sorted behavior examples (15 frames in duration) into behavior categories. Most of the behavior examples were extracted from 4 fps videos of aging N2 worm locomotion, with a small proportion of examples from 4 fps videos of D1 N2 worm locomotion using the *multiple worm detector 5* in the Generate Behavior Examples submodule. Since this Categorizer analyzed the same

behavior categories as *crawling categorizer 9.7_v2_12 frames*, the same criteria were implemented to sort behavior examples into each category. The criteria and number of behavior examples used to train each behavior category are described as follows (1 example pair consists of one Animation and its corresponding Pattern Image):

Forward crawling (441 example pairs)

Reverse crawling (157 example pairs)

Omega bend (423 example pairs)

Immobile (371 example pairs)

Twitching (901 example pairs)

C. elegans behavioral analysis

Analysis of D1 N2 locomotion using *LabGym* (Figure 2)

The Categorizer *crawling categorizer 9.7_v2_12 frames* analyzed D1 N2 worm locomotion, and outputs were used to compare *LabGym*'s accuracy with manual analysis. For behavior classification, we set the Categorizer at an uncertainty level of 0% and a minimum behavior episode length of 4 frames. We analyzed the first 2-minute window of two behavioral videos of Day 1 N2 worms, proportionally downsized from 2056 x 2056 pixels to a frame size of 1028 x 1028 pixels, with six worms in each video. For each of the six target behaviors the Categorizer is trained to identify (behavioral categories described previously in Methods), quantitative measurements were outputted via *LabGym*'s Quantifier Module. The specific calculation methods of the quantitative measurements have been previously reported in detail (Hu et al., 2023; Goss et al., 2024). Duration was the parameter that was manually scored and used in the benchmark comparison between the output of *LabGym* and manual analysis, as it was one of the few parameters that could also be scored manually (Figure 2d). Raster plots of the behavior events in the 2-minute analysis window were generated to depict the results of *LabGym* analysis and manual scoring (Figure 2b). *LabGym* behavior

plot generation was automatic and derived from an *all_events* Excel sheet that records the behavior event and corresponding event probability for each individual at each analyzed frame (every 0.25 seconds).

Manual analysis of D1 N2 locomotion (Figure 2)

The same 2-minute time window of two behavioral videos of Day 1 N2 worms analyzed by *LabGym* was also analyzed by eye to conduct the benchmark comparison. We acquired the duration of each behavioral event and the corresponding time of occurrence in the videos. To generate the raster plot summarizing behavior events scored from manual analysis, the manually-scored behavior events at each frame (every 0.25 seconds) were recorded (with corresponding event probabilities of >0.90 each) in an Excel sheet, formatted similarly to the *LabGym*-generated *all_events* Excel sheet. This Excel sheet was then used to generate a behavior plot via *LabGym*'s built-in Generate Behavior Plot submodule, embedded within the Analysis Module.

Analysis of locomotion of aging worms using *LabGym* (Figure 3)

The Categorizer *crawling categorizer 11 – locomotion across aging* was used to analyze behavioral changes of aging N2 worms from D1 to D15 adults (Figure 3). For behavior classification, we set the Categorizer at an uncertainty level of 5% and a minimum behavior episode length of 4 frames. We analyzed the full 4.5-minute duration of seven behavioral videos of D1, D5, D7, D9, D11, D13, and D15 adult worms. One video per age was analyzed, and each video was proportionally downsized from a frame size of 2056 x 1716 pixels to 1028 x 858 pixels and contained 9-14 worms. For each of the six target behaviors the Categorizer is trained to identify (behavioral categories described previously in Methods), quantitative measurements were outputted via *LabGym*'s Quantifier Module. Forward crawling average speed, Forward crawling distance, Reverse crawling frequency/count,

Omega bend frequency/count, Immobile duration, and Twitching duration were compared across ages (Figure 3b-g).

Trajectory Images

Images of worm trajectories, color-coded by individual, are automatically generated by *LabGym* (Figure 3a). To output images of worm trajectories color-coded by time instead of by individual animal (Figure S1), we utilized Generate Behavior Examples submodule to generate a single behavior example pair (Animation + Pattern Image) on “Interactive Basic” mode (examples contain all detectable animals/objects in the behavioral video) from each of the seven 4.5-minute aging worm behavioral videos. To ensure the output of only one single behavior example pair that summarized the entire behavioral video, we first reduced the frame rate of each behavioral video from 4 fps to 1 fps via the Preprocessing Module, then inputted the total number of frames in the entire behavioral video when prompted to specify how many frames for each generated behavior example in the Generate Behavior Examples submodule. The Pattern Image from the behavior example pair for each video, now representative of the movement patterns of all worms over the entire duration of the behavioral video, was then used as the trajectory images (Figure S1).

Statistical Analysis

All statistical analyses were performed using GraphPad Prism 10.0 software. Statistical methods, error bars, and p-values are indicated in the figure legends. Two-way ANOVA with Bonferroni multiple comparisons post hoc analysis and Unpaired Student’s t-test were applied to analyze data, specifics further described in the figure legends. For all statistical analyses: ns for $p>0.05$, * for $p<0.05$, ** for $p<0.01$, *** for $p<0.001$, **** for $p<0.0001$.

ACKNOWLEDGEMENTS

We thank the Caenorhabditis Genomics Center (CGC) for strains used in this study.

STUDY FUNDING

This work was supported by grants from NIH (R01 NS128500) and the LSI Innovation Partnership Fund to B.Y. and T32 DC00011 to E.A.R.

AUTHOR CONTRIBUTION

B.Y., Y.H., E.A.R., L.A.X., and X.Z.S.X. conceived the project. L.A.X. labeled datasets, trained the Categorizers and Detectors, and analyzed behavioral videos. H.L., Z.L., and E.A.R. recorded behavioral videos. Y.H. designed and programmed *LabGym*. L.A.X., B.Y., Y.H., and E.A.R. wrote the manuscript with assistance from all other authors.

COMPETING INTERESTS

The authors declare no competing interests.

DATA AND CODE AVAILABILITY

An extended user guide with instructions on how to install and use *LabGym*, as well as the software's open-source code, are publicly available on the *LabGym* GitHub page: https://github.com/umyelab/LabGym. The *LabGym* models reported in this study, the behavioral datasets used to train them, as well as all other data are in the process of being deposited and will be accessible upon publication on Zenodo at doi: 10.5281/zenodo.16890721.

REFERENCES

Alonso A, Kirkegaard JB. 2023. Fast detection of slender bodies in high density microscopy data. *Commun Biol* **6**:1–12. doi:10.1038/s42003-023-05098-1

Altimus CM, Marlin BJ, Charalambakis NE, Colón-Rodríguez A, Glover EJ, Izbicki P, Johnson A, Lourenco MV, Makinson RA, McQuail J, Obeso I, Padilla-Coreano N, Wells MF. 2020. The Next 50 Years of Neuroscience. *J Neurosci* **40**:101–106. doi:10.1523/JNEUROSCI.0744-19.2019

Altun, Zeynep F., Hall, David H., Herndon, Laura A. 2009. Introduction. *WormAtlas*. doi:10.3908/WORMATLAS.1.1

Bargmann CI. 2006. Chemosensation in C. elegansWormBook: The Online Review of C. Elegans Biology [Internet]. WormBook.

Brenner S. 1974. The genetics of Caenorhabditis elegans. *Genetics* **77**:71–94. doi:10.1093/genetics/77.1.71

Broekmans OD, Rodgers JB, Ryu WS, Stephens GJ. 2016. Resolving coiled shapes reveals new reorientation behaviors in C. elegans. *eLife* **5**:e17227. doi:10.7554/eLife.17227

Croll NA. 1975. Components and patterns in the behaviour of the nematode Caenorhabditis elegans. *Journal of Zoology* **176**:159–176. doi:10.1111/j.1469-7998.1975.tb03191.x

Datta SR, Anderson DJ, Branson K, Perona P, Leifer A. 2019. Computational Neuroethology: A Call to Action. *Neuron* **104**:11–24. doi:10.1016/j.neuron.2019.09.038

de Bono M, Bargmann CI. 1998. Natural variation in a neuropeptide Y receptor homolog modifies social behavior and food response in C. elegans. *Cell* **94**:679–689. doi:10.1016/s0092-8674(00)81609-8

de Bono M, Maricq AV. 2005. Neuronal substrates of complex behaviors in C. elegans. *Annu Rev Neurosci* **28**:451–501. doi:10.1146/annurev.neuro.27.070203.144259

Deserno M, Bozek K. 2025. Unsupervised Representation Learning of C. elegans Poses and Behavior Sequences From Microscope Video Recordings. *eLife* **14**. doi:10.7554/eLife.106593.1

Feng Z, Li W, Ward A, Piggott BJ, Larkspur ER, Sternberg PW, Xu XZS. 2006. A C. elegans model of nicotine-dependent behavior: regulation by TRP-family channels. *Cell* **127**:621–633. doi:10.1016/j.cell.2006.09.035

Gabel CV, Gabel H, Pavlichin D, Kao A, Clark DA, Samuel ADT. 2007. Neural Circuits Mediate Electrosensory Behavior in Caenorhabditis elegans. *J Neurosci* **27**:7586–7596. doi:10.1523/JNEUROSCI.0775-07.2007

Garigan D, Hsu A-L, Fraser AG, Kamath RS, Ahringer J, Kenyon C. 2002. Genetic analysis of tissue aging in Caenorhabditis elegans: a role for heat-shock factor and bacterial proliferation. *Genetics* **161**:1101–1112. doi:10.1093/genetics/161.3.1101

Glenn CF, Chow DK, David L, Cooke CA, Gami MS, Iser WB, Hanselman KB, Goldberg IG, Wolkow CA. 2004. Behavioral deficits during early stages of aging in Caenorhabditis elegans result from locomotory deficits possibly linked to muscle frailty. *J Gerontol A Biol Sci Med Sci* **59**:1251–1260. doi:10.1093/gerona/59.12.1251

Gong J, Liu Jinzhi, Ronan EA, He F, Cai W, Fatima M, Zhang W, Lee H, Li Z, Kim G-H, Pipe KP, Duan B, Liu Jianfeng, Xu XZS. 2019. A Cold-Sensing Receptor Encoded by a Glutamate Receptor Gene. *Cell* **178**:1375-1386.e11. doi:10.1016/j.cell.2019.07.034

Goss K, Bueno-Junior LS, Stangis K, Ardoin T, Carmon H, Zhou J, Satapathy R, Baker I, Jones-Tinsley CE, Lim MM, Watson BO, Sueur C, Ferrario CR, Murphy GG, Ye B, Hu Y. 2024. Quantifying social roles in multi-animal videos using subject-aware deep-learning. *bioRxiv* 2024.07.07.602350. doi:10.1101/2024.07.07.602350

Graving JM, Chae D, Naik H, Li L, Koger B, Costelloe BR, Couzin ID. 2019. DeepPoseKit, a software toolkit for fast and robust animal pose estimation using deep learning. *eLife* **8**:e47994. doi:10.7554/eLife.47994

Guo X, Navetta A, Gualberto DG, García LR. 2012. Behavioral decay in aging male C. elegans correlates with increased cell excitability. *Neurobiol Aging* **33**:1483.e5–23. doi:10.1016/j.neurobiolaging.2011.12.016

Hebert L, Ahamed T, Costa AC, O'Shaughnessy L, Stephens GJ. 2021. WormPose: Image synthesis and convolutional networks for pose estimation in C. elegans. *PLOS Computational Biology* **17**:e1008914. doi:10.1371/journal.pcbi.1008914

Herndon LA, Schmeissner PJ, Dudaronek JM, Brown PA, Listner KM, Sakano Y, Paupard MC, Hall DH, Driscoll M. 2002. Stochastic and genetic factors influence tissue-specific decline in ageing C. elegans. *Nature* **419**:808–814. doi:10.1038/nature01135

Herndon, Laura A., Wolkow, Cathy, Hall, David H. 2018. Introduction to Aging in C. elegans In *WormAtlas*. doi:10.3908/WORMATLAS.8.4

Hsu A-L, Feng Z, Hsieh M-Y, Xu XZS. 2009. Identification by machine vision of the rate of motor activity decline as a lifespan predictor in C. elegans. *Neurobiol Aging* **30**:1498–1503. doi:10.1016/j.neurobiolaging.2007.12.007

Hu Y, Ferrario CR, Maitland AD, Ionides RB, Ghimire A, Watson B, Iwasaki K, White H, Xi Y, Zhou J, Ye B. 2023. LabGym: Quantification of user-defined animal behaviors using learning-based holistic assessment. *Cell Rep Methods* **3**:100415. doi:10.1016/j.crmeth.2023.100415

Huang K-M, Cosman P, Schafer WR. 2006. Machine vision based detection of omega bends and reversals in *C. elegans*. *Journal of Neuroscience Methods* **158**:323–336. doi:10.1016/j.jneumeth.2006.06.007

Husson SJ. 2012. Keeping track of worm trackers. *WormBook* 1–17. doi:10.1895/wormbook.1.156.1

Iliff AJ, Wang C, Ronan EA, Hake AE, Guo Y, Li X, Zhang X, Zheng M, Liu J, Grosh K, Duncan RK, Xu XZS. 2021. The nematode C. elegans senses airborne sound. *Neuron* **109**:3633-3646.e7. doi:10.1016/j.neuron.2021.08.035

Jorgensen EM, Mango SE. 2002. The art and design of genetic screens: Caenorhabditis elegans. *Nat Rev Genet* **3**:356–369. doi:10.1038/nrg794

Kabra M, Robie AA, Rivera-Alba M, Branson S, Branson K. 2013. JAABA: interactive machine learning for automatic annotation of animal behavior. *Nat Methods* **10**:64–67. doi:10.1038/nmeth.2281

Li W, Kang L, Piggott BJ, Feng Z, Xu XZS. 2011. The neural circuits and sensory channels mediating harsh touch sensation in Caenorhabditis elegans. *Nat Commun* **2**:315. doi:10.1038/ncomms1308

Liu J, Ward A, Gao J, Dong Y, Nishio N, Inada H, Kang L, Yu Y, Ma D, Xu T, Mori I, Xie Z, Xu XZS. 2010. C. elegans phototransduction requires a G protein-dependent cGMP pathway and a taste receptor homolog. *Nat Neurosci* **13**:715–722. doi:10.1038/nn.2540

Liu J, Zhang B, Lei H, Feng Z, Liu Jianfeng, Hsu A-L, Xu XZS. 2013. Functional aging in the nervous system contributes to age-dependent motor activity decline in C. elegans. *Cell Metab* **18**:392–402. doi:10.1016/j.cmet.2013.08.007

Liu KS, Sternberg PW. 1995. Sensory regulation of male mating behavior in Caenorhabditis elegans. *Neuron* **14**:79–89. doi:10.1016/0896-6273(95)90242-2

Mathis A, Mamidanna P, Cury KM, Abe T, Murthy VN, Mathis MW, Bethge M. 2018. DeepLabCut: markerless pose estimation of user-defined body parts with deep learning. *Nat Neurosci* **21**:1281–1289. doi:10.1038/s41593-018-0209-y

Parida L. 2022. The locomotory characteristics of *Caenorhabditis elegans* in various external environments: A review. *Applied Animal Behaviour Science* **255**:105741. doi:10.1016/j.applanim.2022.105741

Pereira TD, Shaevitz JW, Murthy M. 2020. Quantifying behavior to understand the brain. *Nat Neurosci* **23**:1537–1549. doi:10.1038/s41593-020-00734-z

Pierce-Shimomura JT, Morse TM, Lockery SR. 1999. The fundamental role of pirouettes in Caenorhabditis elegans chemotaxis. *J Neurosci* **19**:9557–9569. doi:10.1523/JNEUROSCI.19-21-09557.1999

Pokala N, Flavell SW. 2022. Recording and Quantifying C. elegans Behavior. *Methods Mol Biol* **2468**:357–373. doi:10.1007/978-1-0716-2181-3_20

Roussel N, Sprenger J, Tappan SJ, Glaser JR. 2014. Robust tracking and quantification of C. elegans body shape and locomotion through coiling, entanglement, and omega bends. *Worm* **3**:e982437. doi:10.4161/21624054.2014.982437

Son HG, Altintas O, Kim EJE, Kwon S, Lee SV. 2019. Age-dependent changes and biomarkers of aging in Caenorhabditis elegans. *Aging Cell* **18**:e12853. doi:10.1111/acel.12853

Wang C, Ronan EA, Kim S-K, Kitsopoulos P, Iliff AJ, Grosh K, Kim G-H, Liu J, Xu XZS. 2023. Sensing of sound pressure gradients by C. elegans drives phonotaxis behavior. *Curr Biol* **33**:3985-3991.e4. doi:10.1016/j.cub.2023.08.005

Wang X, Li G, Liu Jie, Liu Jianfeng, Xu XZS. 2016. TMC-1 Mediates Alkaline Sensation in C. elegans through Nociceptive Neurons. *Neuron* **91**:146–154. doi:10.1016/j.neuron.2016.05.023

Ward A, Liu J, Feng Z, Xu XZS. 2008. Light-sensitive neurons and channels mediate phototaxis in C. elegans. *Nat Neurosci* **11**:916–922. doi:10.1038/nn.2155

Weheliye WH, Rodriguez J, Feriani L, Javer A, Uhlmann V, Brown AE. 2025. An improved neural network model enables worm tracking in challenging conditions and increases signal-to-noise ratio in phenotypic screens. doi:10.1101/2024.12.20.629717

Wirak GS, Florman J, Alkema MJ, Connor CW, Gabel CV. 2022. Age-associated changes to neuronal dynamics involve a disruption of excitatory/inhibitory balance in C. elegans. *eLife* **11**:e72135. doi:10.7554/eLife.72135

Wu Y, Kirillov A, Massa F, Lo W-Y, Girshick R. 2019. Detectron2.

Xiao R, Xu XZS. 2021. Temperature Sensation: From Molecular Thermosensors to Neural Circuits and Coding Principles. *Annu Rev Physiol* **83**:205–230. doi:10.1146/annurev-physiol-031220-095215

Zheng M, Cao P, Yang J, Xu XZS, Feng Z. 2012. Calcium imaging of multiple neurons in freely behaving C. elegans. *J Neurosci Methods* **206**:78–82. doi:10.1016/j.jneumeth.2012.01.002

Figure 1

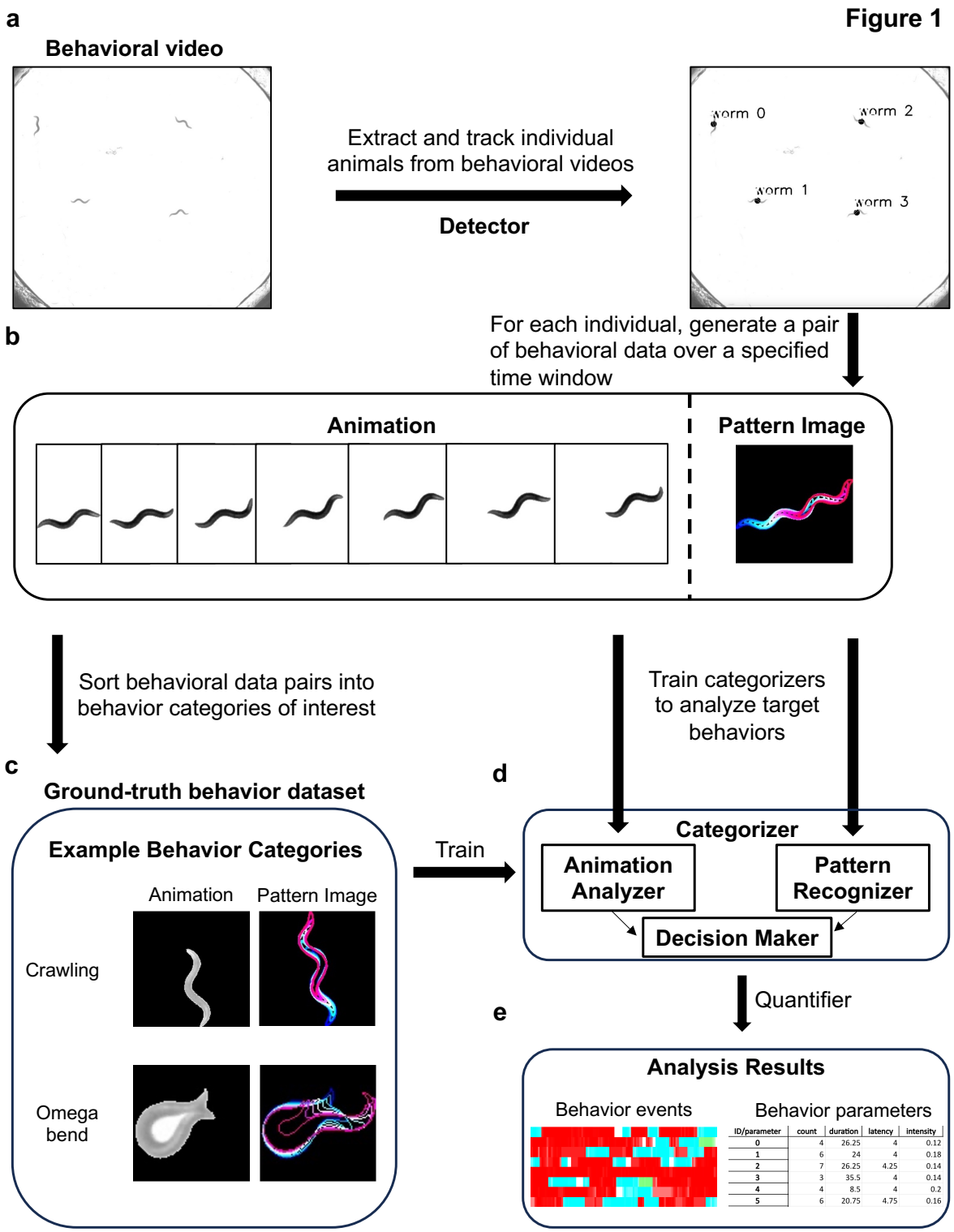

a
Behavioral video
Extract and track individual animals from behavioral videos
Detector
worm 0
worm 2
worm 1
worm 3
b
For each individual, generate a pair of behavioral data over a specified time window
Animation
Pattern Image
Sort behavioral data pairs into behavior categories of interest
Train categorizers to analyze target behaviors
c
Ground-truth behavior dataset
Example Behavior Categories
Animation
Pattern Image
Crawling
Omega bend
Train
d
Categorizer
Animation Analyzer
Pattern Recognizer
Decision Maker
Quantifier
e
Analysis Results
Behavior events
Behavior parameters
ID/parameter count duration latency intensity
0 4 26.25 4 0.12
1 6 24 4 0.18
2 7 26.25 4.25 0.14
3 3 35.5 4 0.14
4 4 8.5 4 0.2
5 6 20.75 4.75 0.16

Figure 1. The pipeline for training deep learning models in *LabGym* for *C. elegans* behavioral analysis

(a) Tracked individual worms are extracted from each frame of a behavioral video using the Detector module.

(b) At each frame of a behavioral video, *LabGym* outputs a pair of behavioral data for each worm over a specified time window: an Animation of the behavioral movement and its corresponding Pattern Image.

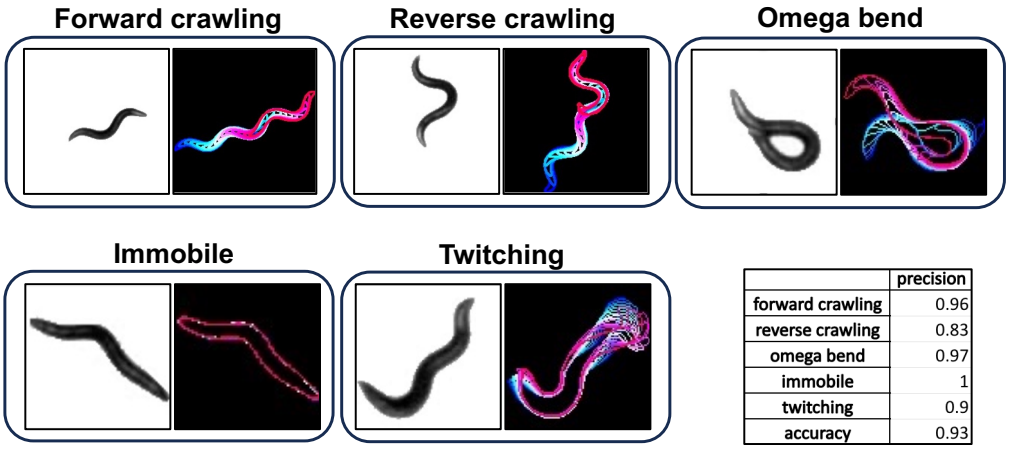
(c) To train the Categorizer, specify behaviors of interest by sorting the generated pairs of behavioral data into user-defined behavioral categories. The Categorizer will utilize both the animation and pattern image to identify and quantify the target behaviors.

(d) A trained Categorizer analyzes both the Animation (via the "Animation Analyzer") and the Pattern Image (via the "Pattern Recognizer") of tracked worms to categorize their individual behaviors over a specified time window.

(e) After behavioral categorization by the Categorizer, the Quantifier evaluates categorization data and animal position to compute quantitative metrics of behaviors exhibited by each worm over any user-specified duration of an analyzed behavioral video.

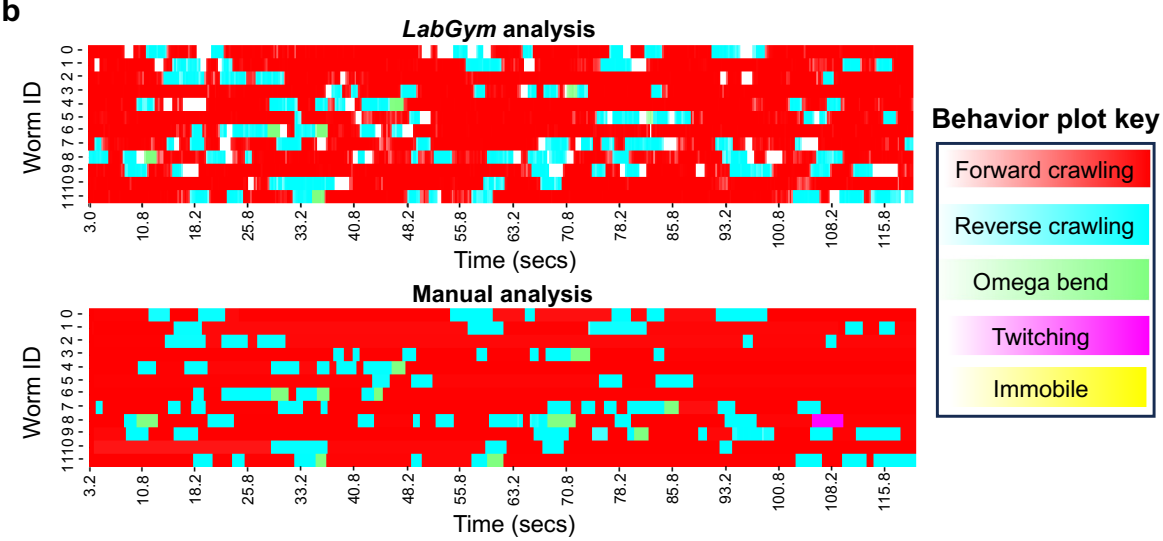
Figure 2

a

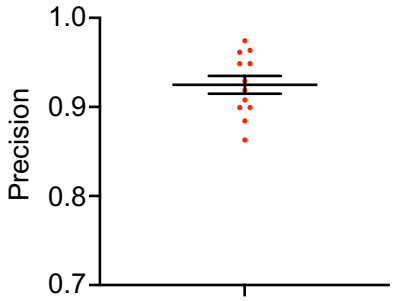


	precision
forward crawling	0.96
reverse crawling	0.83
omega bend	0.97
immobile	1
twitching	0.9
accuracy	0.93

b



c



d

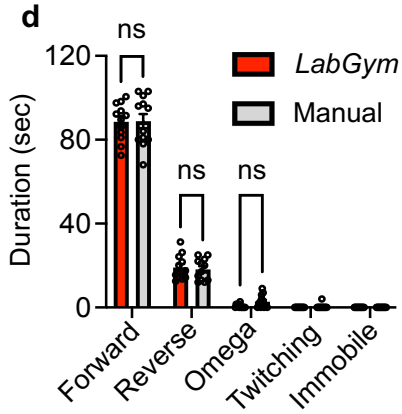


Figure 2. *LabGym* analyzes worm behaviors with precision

(a) Representative behavior example pairs used to train the Categorizer. The categorization precision of each target behavior and the overall categorization accuracy (internally computed by *LabGym*) are summarized in the table.

(b) Raster plots comparing the categorization of behaviors of wildtype N2 worms (n = 12) by *LabGym* with manual analysis (human annotation) per frame (every 0.25 seconds) over a 2-minute time window. X-axis: time (in seconds); Y-axis: each row corresponds to an individual worm. Behavior plot key: color bars indicate behavior categories and their corresponding colors in the raster plot. White spaces indicate uncategorized behavioral events. The length of a colored behavioral event in the raster plot correlates with event duration (in seconds), and the color intensity of a behavioral event correlates with its *LabGym*-computed probability (from 0 to 1.0).

(c) The true accuracy of the *LabGym* Categorizer, determined by the percentage of matching behavioral events between *LabGym* and manual analysis for each individual worm in the raster plot. Scatter dot plot, n = 12. Mean $\pm$ SEM (0.925 $\pm$ 0.008).

(d) The precision of event duration analysis by the *LabGym* Categorizer, determined by comparing behavior event durations calculated from *LabGym* versus manual analysis. Two-way ANOVA with Bonferroni. Grouped bar graph, n = 12 per behavior category.

Figure 3

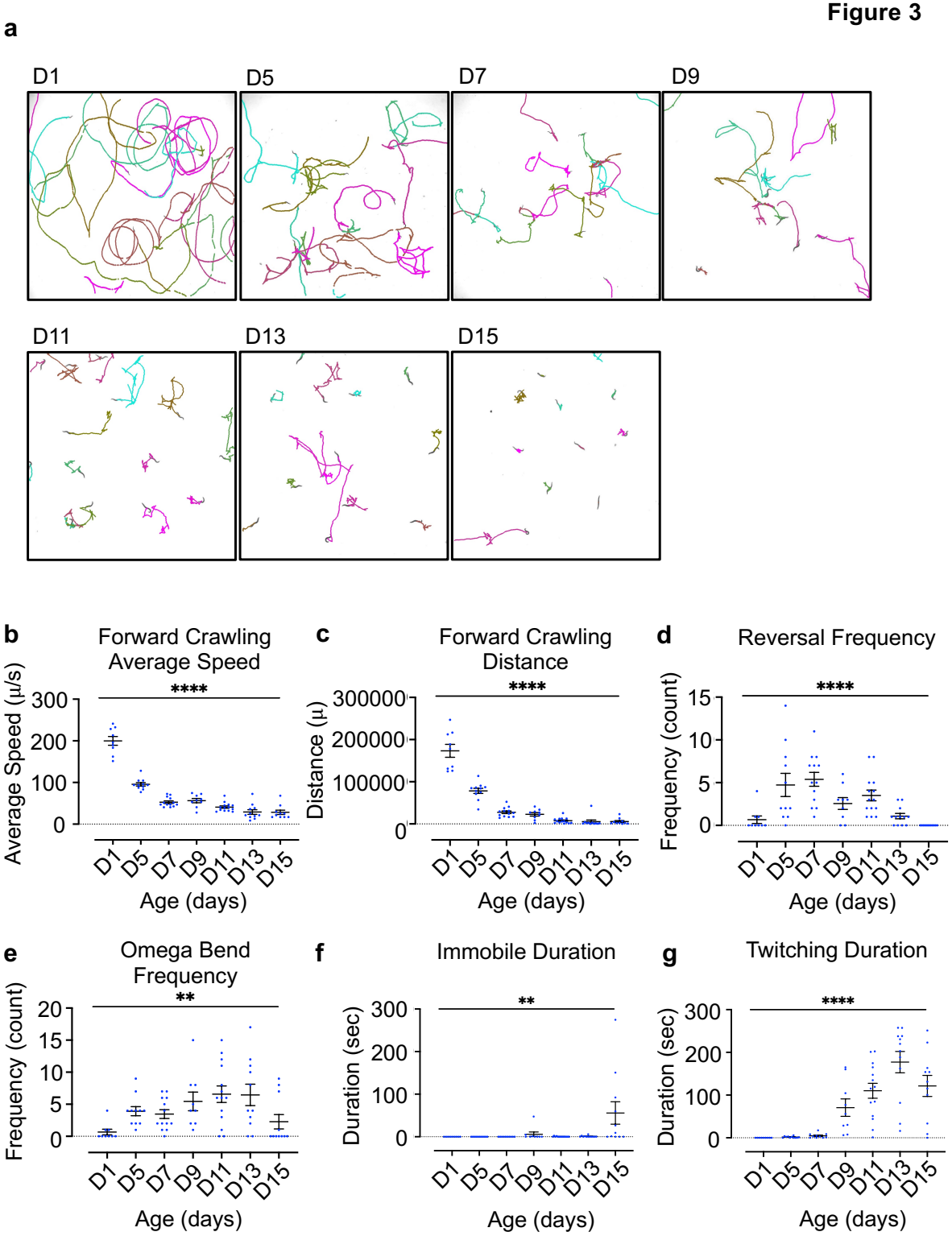

a
D1
D5
D7
D9
D11
D13
D15
b
Forward Crawling Average Speed
Average Speed (µ/s)
300
200
100
0
****
D1 D5 D7 D9 D11 D13 D15
Age (days)
c
Forward Crawling Distance
Distance (µ)
300000
200000
100000
0
****
D1 D5 D7 D9 D11 D13 D15
Age (days)
d
Reversal Frequency
Frequency (count)
15
10
5
0
****
D1 D5 D7 D9 D11 D13 D15
Age (days)
e
Omega Bend Frequency
Frequency (count)
20
15
10
5
0
**
D1 D5 D7 D9 D11 D13 D15
Age (days)
f
Immobile Duration
Duration (sec)
300
200
100
0
**
D1 D5 D7 D9 D11 D13 D15
Age (days)
g
Twitching Duration
Duration (sec)
300
200
100
0
****
D1 D5 D7 D9 D11 D13 D15
Age (days)

Figure 3. *LabGym* captures locomotion changes in aging worms

(a) Trajectory of tracked N2 worms from Day 1 (D1) to Day 15 (D15) adulthood. Each colored trace represents an individual worm. Worms were tracked over a 4.5-minute interval in each image.

(b) - (g) Parameters of behavioral changes exhibited by aging worms from Day 1 (D1) to Day 15 (D15) adulthood, as calculated by *LabGym*. One-Way ANOVA. Scatter dot plot, n $\geq$ 9 per age. **$p<0.01$ and ****$p<0.0001$.

Figure S1

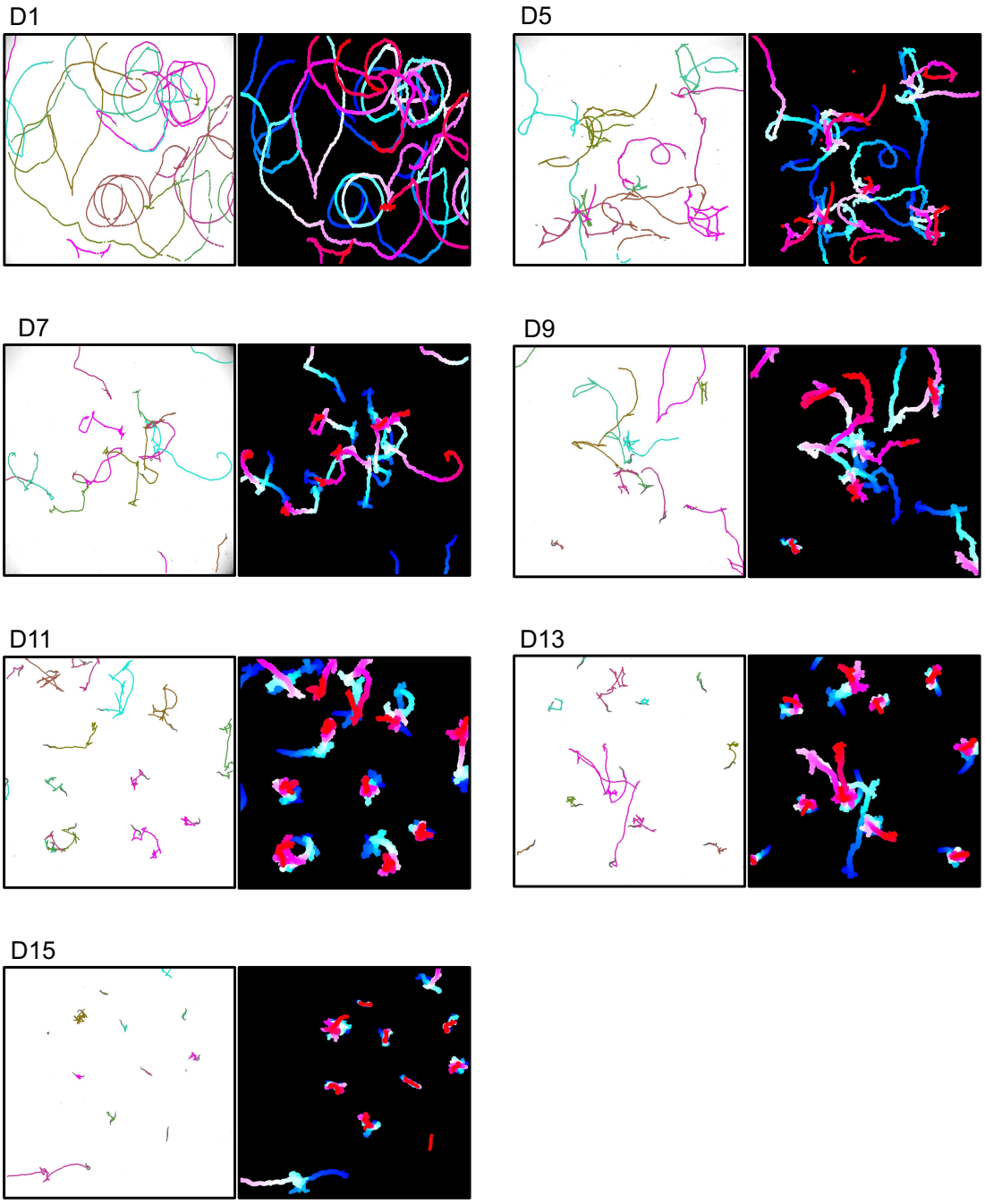


Figure S1. Locomotion trajectories of aging worms, color-coded by individual and by time

For each age, a pair of trajectory images was generated to depict individual worm movement paths, color-coded by individual (left image) or time (right image). For trajectory images color-coded by individual, different colors indicate different individuals. For trajectory images color-coded by time, color gradients indicate temporal progression of movement (earlier timepoints in cool tones, later timepoints in warm tones). These visualizations highlight age-dependent changes in the activity patterns and movement trajectories of worms across adult lifespan.

	crawling categorizer 9.7_v2_12 frames		
	overall accuracy: 0.93		
Animation Analyzer	16 x 16 x 1, level 2		
Pattern Recognizer	32 x 32 x 3, level 3		
	precision	**recall**	**f1**
forward crawling	0.96	1	0.98
reverse crawling	0.83	0.9	0.86
omega bend	0.97	0.9	0.93
immobile	1	0.72	0.84
twitching	0.9	0.83	0.86

	crawling categorizer 11 – locomotion across aging		
	overall accuracy: 0.97		
Animation Analyzer	16 x 16 x 1, level 2		
Pattern Recognizer	32 x 32 x 3, level 3		
	precision	**recall**	**f1**
forward crawling	0.94	1	0.97
reverse crawling	0.98	0.92	0.95
omega bend	0.99	0.98	0.98
immobile	1	0.93	0.96
twitching	0.97	0.98	0.98

Table S1. Summary Table of trained Categorizer metrics

Table summarizing the metrics of each trained categorizer (overall accuracy, Animation Analyzer, Pattern Recognizer, and every behavior category's precision, recall, and f1 score).

Video S1. Representative behavior examples used for Categorizer training

Representative behavior examples used to train the Categorizers. The sorting criteria and total number of example pairs for training each behavior category are detailed in Methods.

Video S2. Annotated videos of N2 worm locomotion by *LabGym*

Left: Behavioral video of six Day 1 adult N2 worms with frame-by-frame annotation of computed behavior category and associated probability. 4x speed, 16 frames per second.

Right: Behavioral video of six Day 1 adult N2 worms with frame-by-frame annotation of computed behavior category and associated probability. 4x speed, 16 frames per second.

Video S3. *LabGym* maintains detection accuracy during multi-worm collision

A video example of a multi-worm collision event where *LabGym* maintains analysis accuracy before, during, and after the collision of two forward crawling worms without identity switching. Although *LabGym* is capable of differentiating individual worms when they physically contact each other in most scenarios, it must be noted that the Categorizer may miscategorize behaviors during and/or a few frames post-collision, but can typically correct itself relatively quickly following entanglement. Given the Categorizer's overall accuracy of ~92.5%, such miscategorizations during collisions account for much of the remaining ~7.5% inaccuracy. Retraining the Categorizer with more behavior examples involving multi-worm collisions is likely to improve its accuracy.