

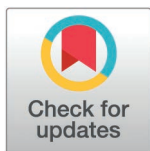
RESEARCH ARTICLE

# Large vision model framework for automated *C. elegans* analysis: From static morphometry to dynamic neural activity

Aurélie Guisnet<sup>\*</sup>, Michael Hendricks<sup>\*</sup>

Department of Biology, McGill University, Montreal, Quebec, Canada

\* [aurelie.guisnet@mail.mcgill.ca](mailto:aurelie.guisnet@mail.mcgill.ca) (AG); [michael.hendricks@mcgill.ca](mailto:michael.hendricks@mcgill.ca) (MH)



## Abstract

Quantitative phenotyping of *Caenorhabditis elegans* is essential across numerous fields, yet data extraction remains a significant analytical bottleneck. Traditional segmentation methods based on pixel-intensity thresholding are highly sensitive to variations in imaging conditions and often fail in the presence of noise, overlaps, or uneven illumination. These failures necessitate meticulous experimental setups, expensive hardware, or extensive manual curation, which reduces throughput and introduces bias. Here, we introduce TWARDIS (Tools for Worm Automated Recognition & Dynamic Imaging System), a modular, Python-based analysis suite that leverages large foundation vision models, specifically the Segment Anything Models (SAM and SAM2) and a fine-tuned vision transformer classifier, to overcome some of these limitations. We demonstrate the versatility of an AI compound system approach across diverse modalities. For static morphological analysis, TWARDIS successfully resolved overlapping worms in noisy images without human intervention, showing a 0.999 correlation with manual segmentation. In behavioral assays (swimming and crawling), the pipeline enabled high-definition postural analysis even in low-resolution, wide-field recordings where the worm occupied only ~0.25% of the field of view, accurately resolving complex postures without frame rejection. Finally, when applied to calcium imaging of semi-restricted animals, TWARDIS provided precise, frame-by-frame segmentation of neural compartments, reducing the artificial signal flattening common in traditional region-of-interest-based approaches and enabling the extraction of biologically accurate, absolute head positions. The system's hardware-scalable architecture and modular design ensure both current accessibility and future improvements without restructuring. By automating some of the most time-consuming aspects of image analysis, TWARDIS removes many critical bottlenecks and tradeoffs in *C. elegans* research, enabling researchers to focus on biological questions rather than technical image processing challenges.

## OPEN ACCESS

**Citation:** Guisnet A, Hendricks M (2026) Large vision model framework for automated *C. elegans* analysis: From static morphometry to dynamic neural activity. PLoS Comput Biol 22(7): e1013441. <https://doi.org/10.1371/journal.pcbi.1013441>

**Editor:** Jian Liu, University of Birmingham, UNITED KINGDOM OF GREAT BRITAIN AND NORTHERN IRELAND

**Received:** August 15, 2025

**Accepted:** June 26, 2026

**Published:** July 22, 2026

**Peer Review History:** PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.pcbi.1013441>

**Copyright:** © 2026 Guisnet, Hendricks. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use,

distribution, and reproduction in any medium, provided the original author and source are credited.

**Data availability statement:** All relevant data and code are within the paper, its [Supporting Information](#) files, and on GitHub at <https://github.com/lillyguisnet/TWARDISv0.1>. The fine-tuned worm classifier model is available on HuggingFace at <https://huggingface.co/lillyguisnet/celegans-classifier-vit-h-14-finetuned>.

**Funding:** This work was supported by funding from the National Science and Engineering Research Council (NSERC) (RGPIN/05117-2014 to MH), the Canadian Institutes of Health Research (CIHR) (PJT-155980 to MH) the Canadian Foundation for Innovation (CFI) (32581 to MH), the Canada Research Chairs Program (950-231541 to MH), and the Fonds de Recherche du Québec – Nature et Technologies (300853 to AG). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** I have read the journal's policy and the authors of this manuscript have the following competing interests: MH is a member of the PLoS ONE Editorial Board. The authors have declared that no other competing interests exist. This does not alter our adherence to PLoS policies on sharing data and materials.

## Author summary

The small roundworm *Caenorhabditis elegans* is widely used by scientists to study fundamental biological questions, such as aging and how the brain works. A crucial part of this research involves analyzing images and videos to measure the worm's shape, movement, and neural activity. However, extracting accurate data is a major challenge. Traditional software tools often fail if the lighting is uneven, the image is noisy, or worms overlap. This forces researchers to spend countless hours manually correcting errors or investing in expensive, specialized equipment. To overcome this bottleneck, we developed TWARDIS (Tools for Worm Automated Recognition & Dynamic Imaging System). Our system utilizes recent advances in Artificial Intelligence, harnessing powerful, generalized AI models to automatically and accurately identify the worms, even in challenging images. We demonstrated that TWARDIS reliably analyzes complex behaviors and neural activity without human intervention. Our approach lessens the traditional trade-off between data quality and equipment cost, enabling high-precision analysis using simple setups. Implemented entirely in Python and compatible with both standard CPUs and GPUs, TWARDIS requires as little as 4 GB of RAM. By automating the most tedious parts of image analysis, TWARDIS allows scientists to focus on biological discovery rather than technical hurdles.

## Introduction

The nematode *Caenorhabditis elegans* has long been established as a leading model organism in biological research, owing to its genetic tractability, conserved biological pathways, short life cycle, and optical transparency. These attributes make it invaluable for studies spanning aging, development, neurobiology, and drug discovery. A cornerstone of *C. elegans* studies is the ability to quantitatively phenotype the organism, which involves, for example, accurately measuring morphology, characterizing complex behaviors, or tracking neural activity. While imaging technology has become increasingly accessible, the extraction of meaningful, high quality data from these images remains a significant time consuming analytical bottleneck.

The first and among the most critical step in virtually all image-based phenotyping pipelines is segmentation: the process of accurately delineating the organism or structures of interest from the background. Traditionally, this step has relied heavily on pixel intensity thresholding and basic feature detection algorithms [1,2]. However, these methods are notoriously fragile and highly sensitive to experimental variability. Imperfections common in routine data acquisition, such as uneven illumination, variations in focus, background debris, or overlapping individuals, frequently cause threshold-based methods to fail, yielding fragmented or erroneous segmentations [3,4].

To compensate for these failures, researchers often invest significant effort in optimizing imaging conditions, utilize expensive high-resolution hardware, compromise on limited feature extraction or resort to extensive manual curation [5]. In dynamic

analyses, such as behavioral tracking or calcium imaging, poor segmentation frequently leads to the rejection of frames or the artificial flattening of signals, resulting in the loss of critical temporal data and potentially masking biological phenomena. Semi-automated tools often require user input to define regions of interest, adjust parameters or manually curate results, which increases processing time and introduces inter-observer variability and bias [2,6,7]. Furthermore, many existing automated tools require specialized computational environments or proprietary licenses (e.g., MATLAB, C++, or complex Docker configurations), limiting their accessibility.

Recent advancements in deep learning have shown promise in automating *C. elegans* analysis [3,8,9]. While powerful, many of these approaches required extensive, expertly annotated datasets tailored to specific imaging conditions and tasks, prohibitively limiting their usage as further fine-tuning is required for performance across different experimental setups [6,8,10]. In addition, others employ deep learning techniques for tasks downstream of segmentation, such as worm identification, phenotype classification, or identity tracking, while still relying on traditional thresholding for the initial mask generation, meaning their segmentation-derived metrics are still affected [3,11–18]. Thus, there still remains a need for a segmentation approach that does not depend on thresholding or require contextual fine-tuning.

A transformative development in computer vision has been the emergence of vision transformer-based foundation models, some of which are trained on extremely diverse, high quality datasets and can perform new tasks with minimal or no fine-tuning. Notably, the Segment Anything Model (SAM) [19] and its successor, SAM2 [20], offer unprecedented generalization capabilities for object segmentation and have demonstrated remarkable robustness to image variability without customization. The success of such foundation models in medical imaging, microscopy, and other biological applications suggests their potential to transform *C. elegans* research as well [21–23].

Here, we introduce TWARDIS (Tools for Worm Automated Recognition & Dynamic Imaging System), a modular, Python-based analysis suite that harnesses the power of these foundation vision models to overcome some long-standing compromises in *C. elegans* phenotyping. TWARDIS provides automated, high-precision pipelines that require minimal user input and is resilient to many imperfect conditions typical of routine laboratory imaging. We demonstrate the versatility of TWARDIS across four modalities of increasing complexity: multi-worm morphological feature extraction from static images; high-definition analysis of swimming behavior in liquid; tracking and postural analysis of crawling worms in low-resolution, wide-field environments; and precise segmentation of neural compartments and head kinematics during calcium imaging.

To contextualize our performance, we compare TWARDIS against traditional thresholding. Since published deep learning-based tools claim to necessitate condition-specific fine-tuning for these tasks, thresholding serves as the most practical baseline for evaluating methods that do not require dataset-specific retraining efforts to use for segmentation. While image quality still dictates the fundamental limits of performance, our approach demonstrates broadened applicability and reduced setup requirements across diverse experimental conditions. Across all applications, the TWARDIS approach also reduces the need for manual intervention, addressing a fundamental tension between the high-throughput nature of *C. elegans* experiments and the labor-intensive nature of image analysis. The system's pure Python implementation and hardware scalability ensures accessibility across users. As foundation models continue to improve, TWARDIS pipelines will inherit these advances without requiring fundamental restructuring, providing a sustainable solution for evolving needs. This work demonstrates how leveraging generalized AI capabilities through a compound system, rather than developing increasingly specialized solutions, can eliminate some longstanding technical bottlenecks, improve community adoption and accelerate biological discovery.

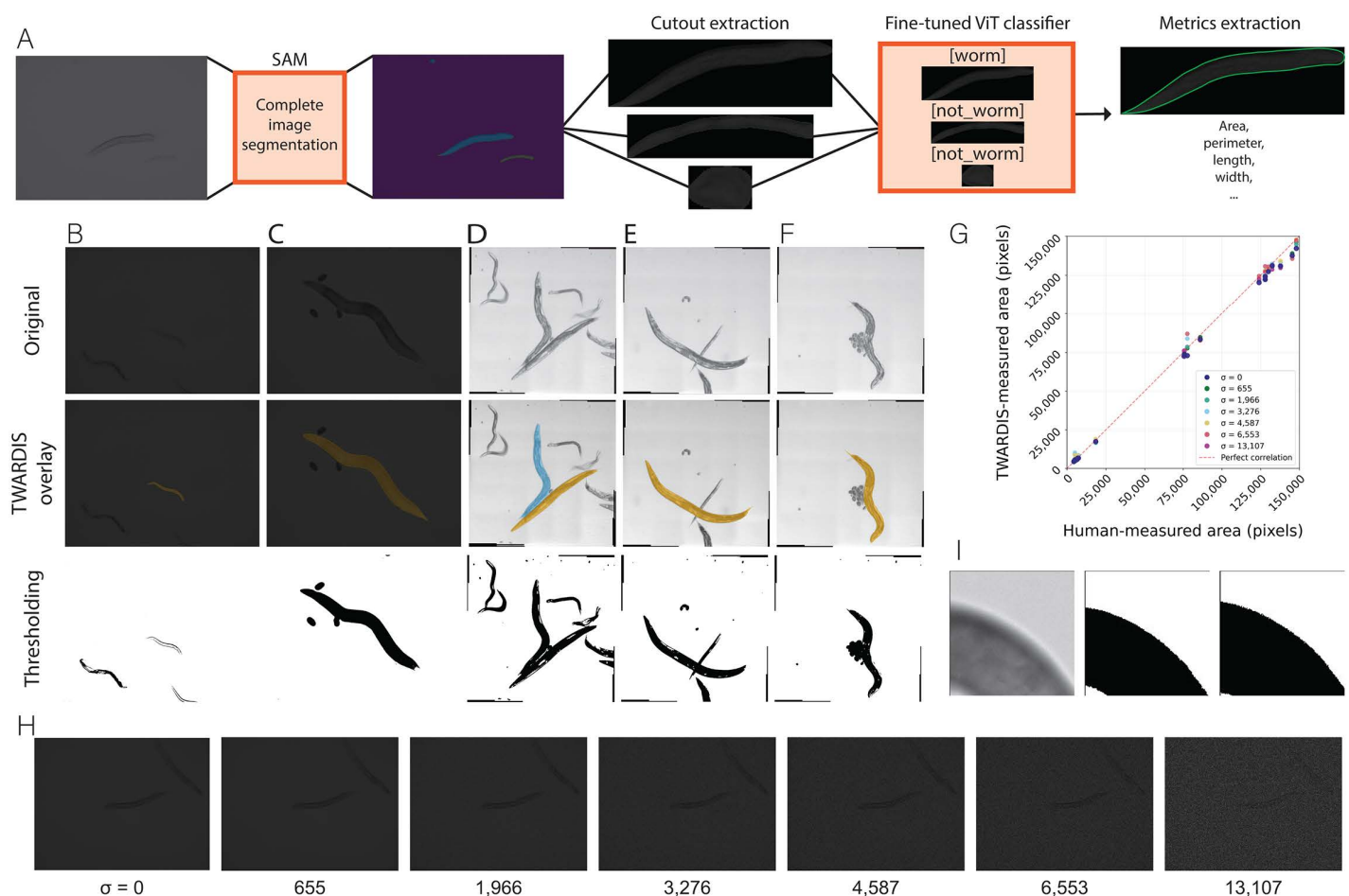
## Results

### Static multi-worm morphological feature extraction

We first integrated the SAM vision model with our fine-tuned worm classifier to automatically extract morphological features from static images of worms. After initial image segmentation of the entire image with the SAM model, we fine-tuned

a vision transformer classifier to automatically select worm segments as the SAM model does not have object-specific segmentation capabilities (Fig 1A). We tested this TWARDIS pipeline across diverse worm life stages (L1 larvae to day 12 adults) (Table 1). We successfully segmented images and used our fine-tuned classifier to select worm segments. We obtained perfect classification of segments at all life stages in a single forward pass with no human intervention on our fine-tuning dataset, demonstrating resilience of the pipeline to common imaging challenges present in our dataset such as variability in lighting, focus, noise artifacts and sample preparation. To isolate worms of the target developmental stage in images containing mixed-age populations, a size deviation threshold was applied per imaging day.

Traditional binary thresholding methods, and even more modern approaches, are sensitive to image imperfections: for *C. elegans* images, they often fail on out of focus elements, worm overlaps, sharp pharynx contrast, or noisy images, requiring either meticulous experimental setups for near perfect data acquisition conditions with manual or automated



**Fig 1. Multi-worm morphological feature extraction in static images.** (A) Multi-worm static image feature extraction workflow. Each image is first passed through SAM for a complete segmentation. Cutouts are extracted for each identified mask that is not touching the sides of the image. The cutouts are classified as “worm” or “not worm” by our fine-tuned ViT classifier. Morphological metrics are extracted for each identified worm mask. (B–F) Example comparisons of worm segmentation by thresholding versus TWARDIS age-based extraction. A–B samples from the first dataset. C–E samples from the second dataset. Colors represent distinctly identified worms. (G) Area in pixels of randomly selected worms measured by human manual segmentation versus TWARDIS-based segmentation. Colors represent Gaussian noise levels, where 0 is none (original image). Perfect correlation represented by the dashed line.  $n=21$ . (H) Representative image at the different Gaussian noise levels. (I) Example discrepancy in edge placement when thresholding on a blurry image.

<https://doi.org/10.1371/journal.pcbi.1013441.g001>

**Table 1. Number of analyzed images by worm age and total number of worms analyzed in each dataset.**

| Worm day post-hatching | First dataset (number of images) | Second dataset (number of images) |
|------------------------|----------------------------------|-----------------------------------|
| 1                      | 41                               | 29                                |
| 2                      | 46                               | 31                                |
| 3                      | 44                               | 35                                |
| 4                      | 41                               | 33                                |
| 5                      | 32                               | 49                                |
| 6                      | 33                               | 0                                 |
| 7                      | 0                                | 61                                |
| 9                      | 0                                | 52                                |
| 12                     | 0                                | 53                                |
| Total worms extracted  | 237                              | 438                               |

First dataset: Compound microscope at 2x magnification. Second dataset: Spinning-disk confocal microscope with image tiling. 1–5 worms per image.

<https://doi.org/10.1371/journal.pcbi.1013441.t001>

curation to discard invalid samples [2–4]. Visual comparisons in Fig 1B–F exemplify these advantages: raw images with overlapping worms and noise yield fragmented or erroneous outputs via thresholding, while our implementation produces clean, accurate masks without requiring any manual input, such as clicking on objects or defining regions of interest [6,24]. We tested our worm classifier on a second set of images collected on another recording set up and obtained zero false positives and less than 5% of false negatives showing that over-fitting on the optical profile of our training set was minimal (Fig 1D–F).

From these segmented masks, we extracted key morphological features, including area, perimeter, length, and width along the entire worm length with minimal additional computational effort. When compared to manual segmentation, the correlation with our TWARDIS implementation was 0.999 (Fig 1G, Table 2). No images from our datasets could be correctly segmented by using only traditional thresholding, showing that our method dramatically reduces the need for pre-acquisition care or post-analysis handling time (Fig 1B–F).

To characterize the robustness boundaries of our pipeline against image quality, we applied progressive levels of artificial Gaussian noise ( $\sigma = 655\text{--}13,107$ ) (Fig 1G and 1H, Table 2). At moderate noise levels ( $\sigma \leq 1,966$ ), the pipeline

**Table 2. TWARDIS performance under progressive Gaussian noise degradation. Performance metrics of TWARDIS static-morphometry pipeline compared to the manual segmentation comparison dataset (n=21 worms) with increasing levels of additive Gaussian noise ( $\sigma$ ). Correlation: Pearson correlation coefficient between TWARDIS-measured and manually measured worm area in pixels for detected worms. Mean % agreement  $\pm$  SD: per-worm percent area relative to manual segmentation (< 100%: TWARDIS smaller estimate compared to human; > 100%: TWARDIS higher estimate compared to human). False negative (FN) and false positive (FP) rates reported separately for segmentation step (SAM mask generation) and classification step (ViT classifier assignment).  $\sigma = 0$  corresponds to the original unmodified images.**

| Noise level ( $\sigma$ ) | Correlation | Mean % agreement $\pm$ SD | % FN, segmentation | % FP, segmentation | % FN, classification | % FP, classification |
|--------------------------|-------------|---------------------------|--------------------|--------------------|----------------------|----------------------|
| 0                        | 0.999       | 97.70 $\pm$ 4.78          | 0                  | 0                  | 0                    | 0                    |
| 655                      | 0.999       | 100.3 $\pm$ 5.23          | 0                  | 0                  | 0                    | 0                    |
| 1,966                    | 0.999       | 110.62 $\pm$ 44.24        | 0                  | 0                  | 0                    | 4.76                 |
| 3,276                    | 0.999       | 106.76 $\pm$ 25.31        | 9.52               | 0                  | 4.76                 | 0                    |
| 4,587                    | 0.999       | 106.84 $\pm$ 20.41        | 9.52               | 0                  | 23.81                | 0                    |
| 6,553                    | 0.988       | 100.49 $\pm$ 4.78         | 33.33              | 0                  | 19.05                | 0                    |
| 13,107                   | 0.998       | 97.86 $\pm$ 2.21          | 57.14              | 0                  | 9.52                 | 0                    |

<https://doi.org/10.1371/journal.pcbi.1013441.t002>

maintained high performance: 0% false negatives, <5% false positives by the classification model and > 0.999 segmentation correlation with human segmentation. As noise increased further ( $\sigma = 3,276\text{--}6,553$ ), false negative rates emerged both from segmentation and classification failures. For the worms that were successfully detected, segmentation quality remained high ( $R \geq 0.988$ ). At the most extreme noise level ( $\sigma = 13,107$ ), only 33% of worms were detected, but the correlation with manual segmentation for detected worms remained high ( $R \geq 0.998$ ) with no false positives. Conservative missed detections rather than erroneous measurements is a desirable property for quantitative phenotyping, as it preserves data integrity at the cost of reduced throughput rather than introducing errors.

This fully automated approach not only eliminates user intervention but also mitigates inter-observer variability and bias inherent in manual annotation [25,26]. For instance, multiple researchers outlining the same worm contours, either manually or by thresholding, will inevitably have discrepancies in edge placement which in turn can lead to variability in measured features (Fig 1I). This is also illustrated by the mean percent area agreement between TWARDIS and human segmentation (Table 2) which shows that image quality affects edge placement ambiguity. In contrast, our method ensures a consistent approach across users and datasets which reduces bias. Most notably, our automation makes processing time irrelevant. By being fully automated, parallelizable and both CPU and GPU compatible (SAMs  $\approx 0.5$  GB, fine-tuned classifier  $\approx 2.5$  GB), this TWARDIS workflow can be used to process any amount of data at a speed that scales with available hardware (S1 Fig).

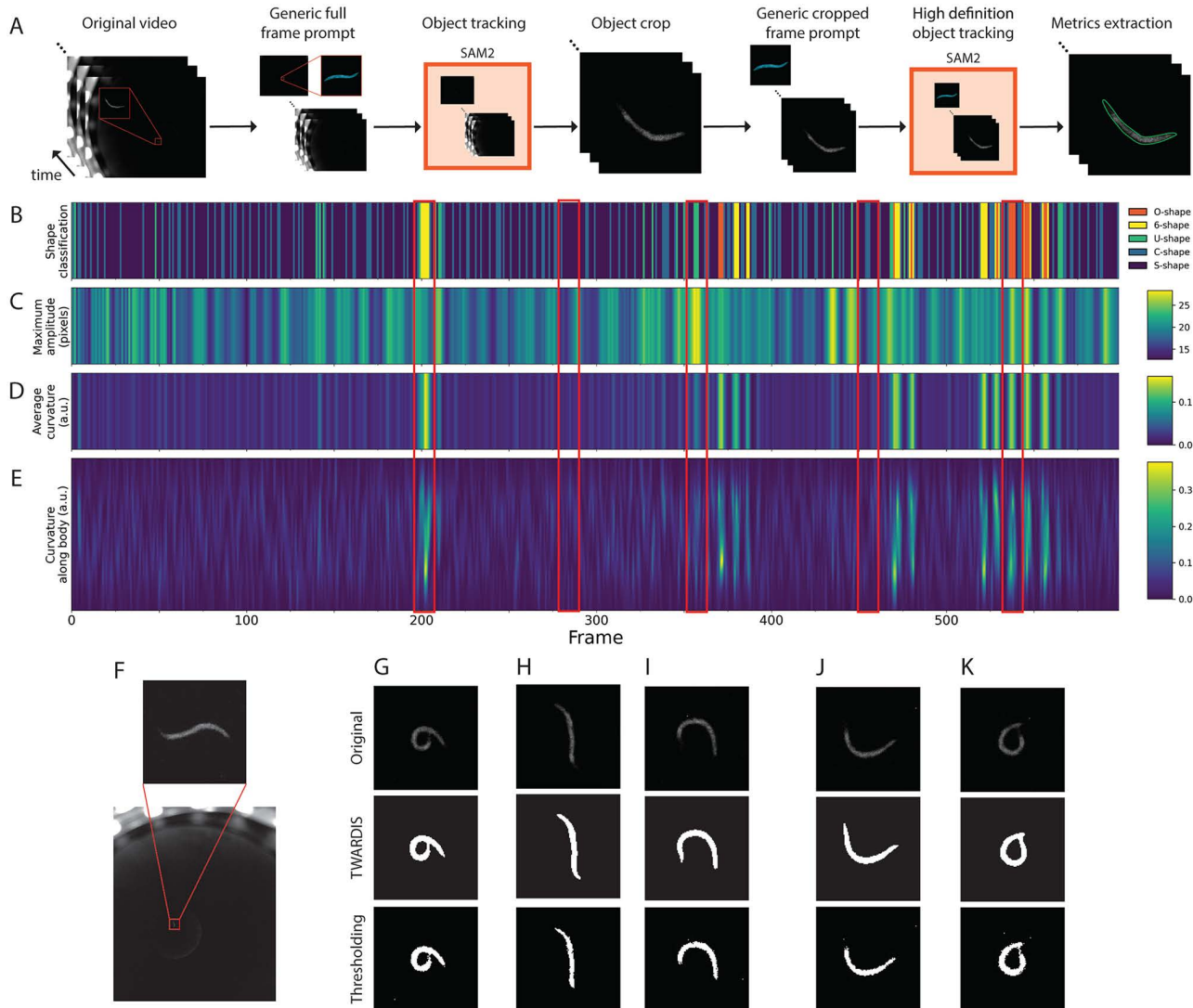
### Quantitative analysis of swimming behavior

We applied our approach to analyze video recordings of *C. elegans* swimming in upside down liquid droplets (Fig 2A). This modality presents new challenges, including self-touching postures (extreme O- or 6-shapes), faint outlines, dynamic noise artifacts from light refraction, background fluctuations and moving extraneous particles. Current swimming analyses rely on thresholding and pixel differences for tracking which is less robust to imperfect conditions resulting in frame rejection or manual intervention to maintain accuracy [7,27]. In contrast, our TWARDIS implementation handled these challenges without any frame exclusions or manual tuning, producing high-definition masks that accurately captured worm shape even in a low-resolution space where the worm occupied only a very small portion of the total field of view ( $\sim 0.25\%$ ) (Fig 2F and S1 Movie). Our approach resulted in more complete body segmentations (Fig 2G–K).

From these refined masks, we extracted worm skeletons as described above, enabling the computation of a broad suite of additional behavioral features of interest with high fidelity such as curvature along the body, amplitude (perpendicular distance to the centerline) and wavelength. The superior foundation segmentation quality directly translated to more accurate skeleton extractions, eliminating the need to discard frames due to poor shape representation and allowing for continuous time-lapse metrics and precise timings of behaviors (Fig 2B–E). For example, deep 6-shaped bends can be distinguished as periods of very high curvature and small amplitude, whereas O-shaped bends have both high curvature and high amplitude (Fig 2G and 2K, S1 Movie). Using a single inexpensive camera, standard LED lighting and a simple set up, we achieved detailed feature extraction without the need for costly or careful preparations and removed the need for manual curation.

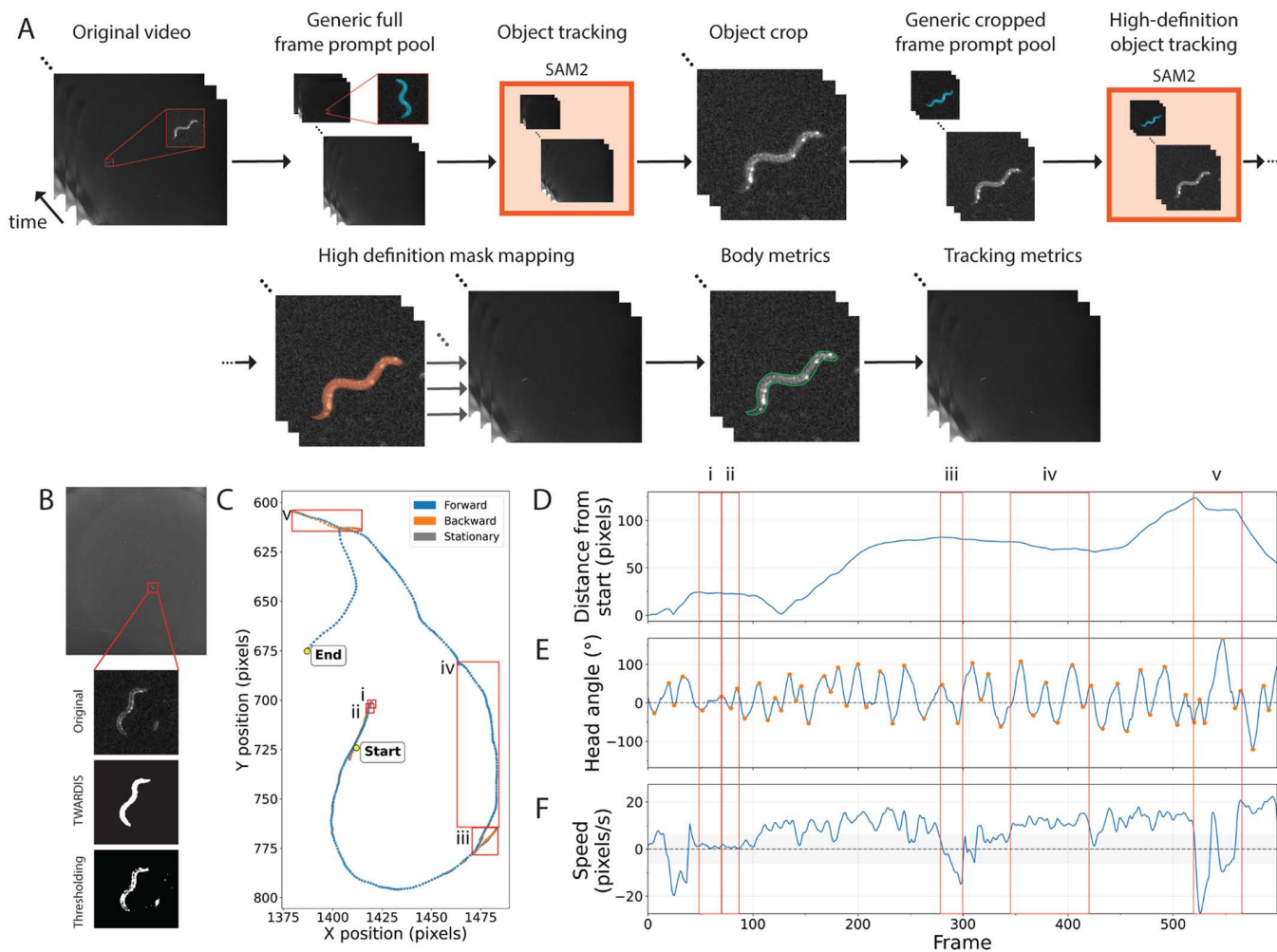
### Single worm high-definition crawling tracking

We next extended our process to crawling videos of *C. elegans* on agar plates in unrestricted environments, where, like in our swimming recordings, worms occupied a small portion of the total field of view (Fig 3A and 3B). These simple recording conditions were also low in hardware cost and preparation time. Again, conventional thresholding methods typically struggle here, producing noisy or incomplete masks in the presence of self-touching postures or background artifacts, often leading to frame rejection, reliance on manual corrections or use of expensive recording equipment [28–30]. For example, out of 114 one-minute recordings of wild-type worms in our conditions, a popular threshold-based worm tracker, Tierpsy, provided features across the entire recording for only 76.3% of the dataset with variable levels of completeness



**Fig 2. Quantitative analysis of swimming behavior.** (A) Swimming behavior analysis workflow. First, a generic prompt frame (a reference image containing user-provided mask(s) of the object(s) of interest, which the model uses to identify and segment the same object(s) in unannotated frames) is added to the video recording. Then, SAM2 is used to track the worm object through the recording. A cropped version of the video is created using the tracked worm. Another size-matched generic prompt frame is added to the cropped video and passed through SAM2 for a high-definition segmentation of the worm. Finally, metrics are extracted for each frame. Red picture-in-picture of worm shown for reference. (B–E) Selected time-series metrics of a 10 frames-per-second one-minute recording of a wild-type worm swimming in a liquid droplet. (B) Per-frame shape classification based on amplitude and curvature. (C) Maximum bending amplitude per frame measured as the maximum perpendicular distance between the points along the worm's body and a straight line connecting the head and tail. (D) Average body curvature per frame from E. (E) Gaussian-weighted curvature intensity of 100 interpolated points along the worm's body per frame. (F) Example frame from recording representing the low space occupied by the worm within the total image field-of-view. (G–K) Original worm cutout with TWARDIS- and threshold-based segmentations for different frames along the recording. (G) Worms in a 6-shape are characterized by a small amplitude and very high curvature. (H) Worms in a mild S- or near straight shape are characterized by a very small amplitude and very small curvature. (I) Worms in a U-shape are characterized by a high amplitude and moderate curvature. (J) Worms in a wide C-shape are characterized by moderate amplitude and moderate curvature. (K) Worms in an O-shape are characterized by a high amplitude and high curvature.

<https://doi.org/10.1371/journal.pcbi.1013441.g002>



**Fig 3. Single-worm tracking.** (A) Crawling behavior analysis workflow. First, a generic prompt frame pool (a curated set of multiple prompt frames, selected to represent a range of positions the object(s) of interest can take, providing the model with diverse reference examples to improve accuracy) is added to the video recording. Then, SAM2 is used to track the worm object through the recording. A cropped version of the video is created using the tracked worm. Another size-matched generic prompt frame pool is added to the cropped video and passed through SAM2 for a high-definition segmentation of the worm. The high-definition segmented worm is mapped back to the original video size. Finally, body metrics and path tracking metrics are extracted for each frame. Red picture-in-picture of worm shown for reference. (B) Example frame from recording representing the low space occupied by the worm within the total image field-of-view with a worm cutout with TWARDIS- and threshold-based segmentations. (C–F) Selected time-series metrics of a 10 frames-per-second one minute recording of a wild-type worm crawling on an agar plate with food. (C) Worm path analysis color-coded by movement classification from (F) and where each point is a frame. Start and end points are labelled. X and Y-axes represent pixel position in the video resolution. (D) Euclidean distance from start point per frame in pixels. (E) Head bending angle relative to body in degrees per frame. 0 indicates a straight head relative to the body. Positive and negative values represent bends on either side, but ventral and dorsal sides are not distinguished. Points along the trace represent the detected peaks and troughs used for head bend metrics. (F) Worm speed by frame. Positive values represent forward motion, negative values represent backward motion, stationary motion is determined by absolute speed less than 0.6 pixel/second (shaded area). (C, Fi) Immobile stationary period. (C, Fii) Stationary period with head movement (C, Fiii) Reversal. (C, Fiv) Continuous forward crawl with a slightly curved path, distinguishable by head bends being deeper on one side. (C, Fv) Fast reversal followed by omega-turn (very deep head bend with sharp path direction change).

<https://doi.org/10.1371/journal.pcbi.1013441.g003>

(mean % of the recording with failed detection  $\pm$  SD:  $2.01 \pm 9.78$ , min: 0%, max: 96.33%,  $n = 114$  one-minute recordings at 10 fps). Our automated approach, however, resolved these challenges without any manual corrections or frame exclusions, achieving clean masks even in our imperfect conditions (Fig 3B).

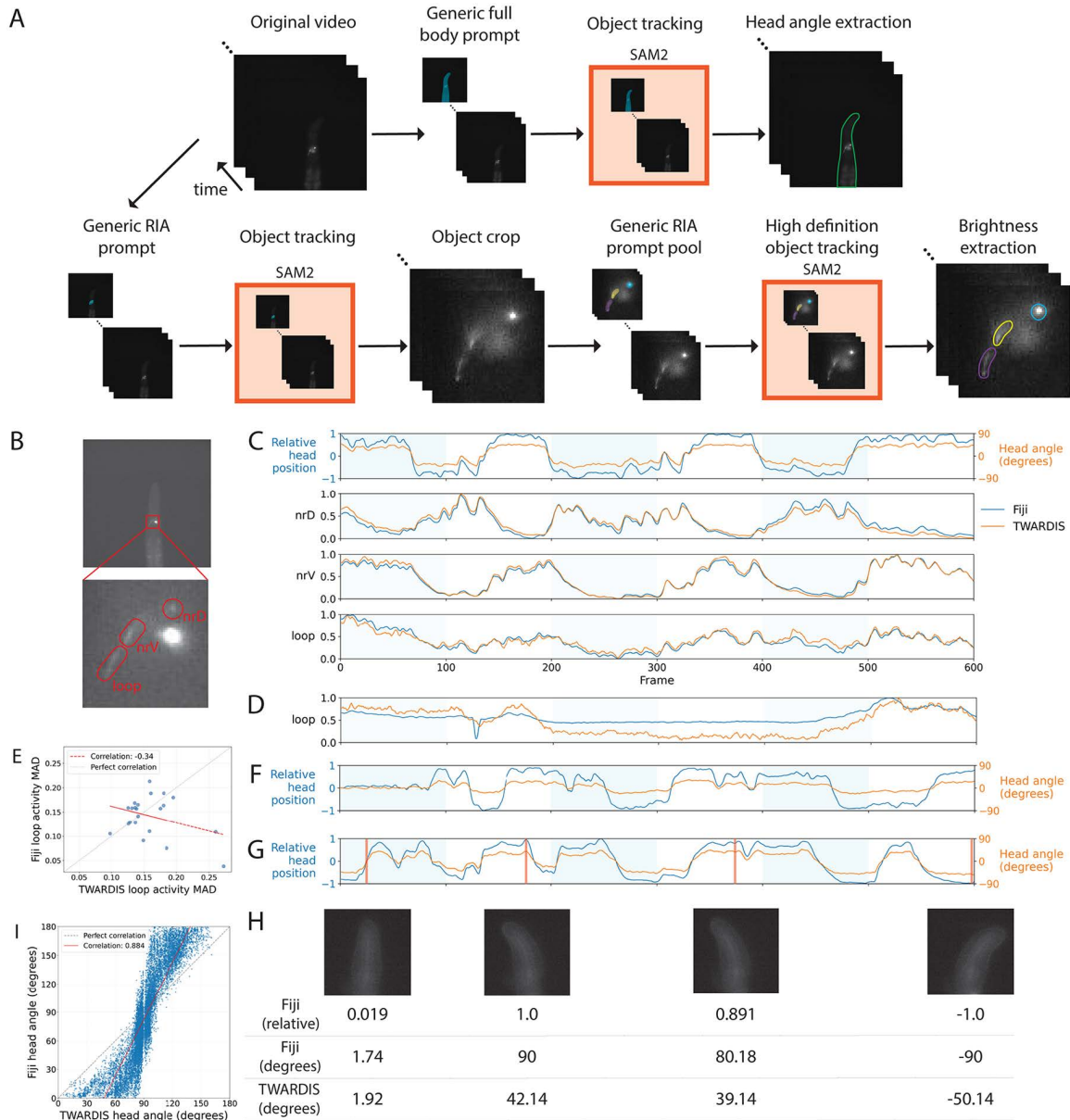
From these high-quality masks, we could extract the same body features as under swimming conditions in addition to a set of path features across frames, including, for instance, head bending angles, speed, and complete path tracking (Fig 3C–F). This enabled reliable body shape analysis alongside path tracking with complete frame inclusion, eliminating the common trade-off between centroid-only tracking more common with low-cost, basic systems versus full posture details which typically require more costly hardware [31]. By delivering high-resolution segmentation and a broad feature set from minimal hardware with no manual corrections, our TWARDIS approach bridges a gap between accessibility, precision and manual analysis time (S2 Movie).

### Segmentation of calcium imaging recordings of semi-restricted worms

Lastly, we applied our strategy to extract fluorescence information and head movements from calcium imaging recordings of semi-restricted *C. elegans* in microfluidic devices. Specifically, we tracked the fluorescence of the three axonal compartments of RIA (nrV, nrD and loop) in a microfluidic device where alternating streams of odors were presented to the animal while the head was free to move (Fig 4A and 4B) [32]. Our automated approach segmented each axonal compartment frame-by-frame, yielding precise masks despite structural change in the neuron through the recording. Using an expertly-annotated prompt pool of approximately 20 images to guide the model's understanding of compartment boundaries, we obtained detailed and precise brightness measurements for calcium dynamics (Fig 4C and S3 Movie). Traditional segmentation tools for this analysis include ImageJ (Fiji), Matlab, or OpenCV software, where plugins or functions use image registration algorithms to align neuronal compartments across frames to extract fluorescence intensity by region-of-interest (ROI) selection [33]. However, these algorithms typically struggle with aligning structures that have inconsistent geometries across frames such as the three axonal compartments of the RIA interneuron. While this can be manually corrected, it significantly increases handling time. On the other hand, if these issues go undetected, errors are introduced in the data: in most cases, a sudden artificial drop in the fluorescence intensity caused by the axonal compartment going momentarily outside of the ROI results in the entire intensity trace to be flattened (Fig 4D). The loop compartment, being particularly dim at baseline, is particularly vulnerable from having an imperfect ROI selection. TWARDIS-based analysis resulted in loop fluorescence intensity traces having similar or greater mean absolute difference than Fiji-analyzed traces demonstrating less artificial flattening with our approach (Fig 4E).

Whole-body segmentation in our approach facilitated accurate head motion tracking via skeletonization and angle computation, providing more biologically relevant metrics. Previous standard approaches estimate head position by using the maximum ellipse line of an ellipse fitted to a binarily thresholded version of the worm's head. These measures are then normalized to the minimum and maximum for each recording to estimate head position on the 180 degrees of possible range. On the other hand, TWARDIS yields the absolute head angle position relative to the midline of the body which provides much more accurate and relevant information for hypothesis testing than recording-specific normalization (Fig 4C, 4F and 4G). In general, the Fiji ellipse-fitting approach estimates deeper head bends than our TWARDIS approach with an overall correlation of 0.884 between the two approaches and of only 0.746 when direction of the head bend is ignored (Figs 4I and S2). These differences are particularly apparent in recordings where the worm's head bends are minimal or restricted on one side where head angle inflation becomes most striking (Fig 4F–H and S3 Movie).

To further compare the two approaches, we computed the percent agreement between Fiji- and TWARDIS-extracted signals for each compartment and head angle (absolute % agreement  $\pm$  SD: nrD:  $96.43 \pm 3.38$ ; nrV:  $96.02 \pm 3.54$ ; loop:  $86.11 \pm 16.40$ ; head angle:  $85.62 \pm 9.66$ ). The divergence between the two methods quantifies the described limitations of the ROI-based approach: particularly, the low agreement and high variability for the loop intensity and head angle is consistent with the signal flattening and head angle inflation artifacts we identified. Similarly, we compared both methods against human-measured head angles for 50 randomly selected frames (S3 Fig). TWARDIS-extracted angles showed substantially closer agreement with human measurements than Fiji-extracted angles (mean absolute difference from human  $\pm$  SD: TWARDIS =  $6.20^\circ \pm 5.56^\circ$ ; Fiji =  $22.98^\circ \pm 16.42^\circ$ ). However, manual head angle measurement is itself subject



**Fig 4. Segmentation of calcium imaging recordings of semi-restricted worms. (A)** RIA calcium imaging analysis workflow. First, a generic prompt frame of the worm body is added to the video recording. SAM2 is used to track the worm object through the recording. The resulting worm masks are used to extract the head angles. Independently, a generic prompt frame of the RIA region is added to the video recording. SAM2 is used to track the RIA region through the original recording. A cropped version of the video is created using the tracked RIA region. A size-matched generic prompt frame pool of the three axonal compartments of RIA is added to the cropped video and passed through SAM2 for a high-definition segmentation of the compartments. The high-definition masks of the axonal compartments are used to extract brightness activity from the original recording. **(B)** Representative frame from calcium imaging recording with brightness enhanced for visualization. Blowup shows the three segmented compartments of the RIA interneuron whose fluorescence intensity through time are presented in (C). **(C)** Head bending and fluorescence time-series values of a 10 frames-per-second one minute recording of a wild-type worm in a microfluidic device expressing GCaMP3.3 in RIA. Blue traces represent Fiji-extracted data and orange TWARDIS-extracted data. Shaded areas represent periods where the positive odor is present. Top panel, Fiji-extracted relative head position (-1, 1) and TWARDIS-extracted head position (degrees relative to body). Second, third and fourth panel, nrD, nrV and loop normalized fluorescence intensity, respectively. **(D)** Illustrative normalized fluorescence intensity of the loop compartment where the Fiji-extracted data is artificially flattened by momentary movement outside of extraction ROI. **(E)** Average absolute difference from mean for Fiji- versus TWARDIS-extracted data of the loop compartment.  $n=22$  recordings. **(F–G)** Illustrative traces where Fiji- and TWARDIS-extracted head positions are notably different. **(H)** Representative head positions from (G) with Fiji- and TWARDIS-extracted head position values. **(I)** Scatterplot of Fiji- versus TWARDIS-estimated head angles. Fiji estimations (-1, 1)

are converted to the (0, 180) range for comparison. 90 represents a perfectly straight head. 0 and 180 degrees represent maximal bending.  $r=0.884$ .  $n=13,486$  frames from 22 recordings.

<https://doi.org/10.1371/journal.pcbi.1013441.g004>

to observer bias and line-placement ambiguity, as discussed for morphological segmentation (Fig 11), and should thus not be treated as an absolute ground truth.

## Discussion

The quantitative analysis of *C. elegans* morphology, behavior, and neural activity is fundamental to its utility as a model organism. However, analytical pipelines in the field have often been constrained by a trade-off between precision, throughput, and accessibility. More precisely, traditional methods frequently require meticulous experimental conditions, expensive hardware, or significant manual intervention [2,5,30]. Furthermore, existing tools can be difficult to deploy, maintain or modify due to dependencies on proprietary software, condition specific re-training or complex programming environments (e.g., MATLAB, Java, C++, or Docker) [27,28,34–37]. This study introduces the first version of TWARDIS (Tools for Worm Automated Recognition & Dynamic Imaging System), a modular analysis suite implemented in Python that leverages the capabilities of foundation large vision models to deliver automated, high-precision phenotyping across diverse imaging modalities addressing some of the key bottlenecks in researcher time spent on repetitive manual image analysis tasks and equipment requirements. We demonstrate robust and accurate segmentation across four distinct imaging modalities: static morphometry, swimming behavior, crawling locomotion, and calcium imaging, all while requiring minimal to no human intervention (Table 3).

The core innovation of TWARDIS lies in addressing a fundamental limitation of existing analysis tools: the segmentation step. Virtually all current segmentation tools rely fundamentally on pixel-intensity thresholding or basic feature detection algorithms, even in more modern tools implementing neural network-based worm tracking [3,13,17,28]. While several recent tools incorporate deep learning techniques for downstream tasks such as shape classification, or identity tracking, these operate on masks already produced by thresholding, meaning the upstream segmentation limitation remains, making these approaches similarly inherently fragile and highly sensitive to variations in lighting, focus, debris, and

**Table 3. Summary of model parameters and post-processing steps by modality. Summary of model variants, parameters, prompt strategies, and main post-processing steps used across the four imaging modalities. Default parameters indicate SAMs generator defaults were used without modification. Main post-processing details are summarized here; see corresponding Methods subsections for full descriptions.**

|                            | Modality                                                                                                                                                                       |                                                           |                                                                                                                                         |                                                                                                                                                                                                                                                                                                                |
|----------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                            | Static                                                                                                                                                                         | Swimming                                                  | Crawling                                                                                                                                | Ca <sup>2+</sup> Imaging                                                                                                                                                                                                                                                                                       |
| Model                      | SAM ViT-H 4b8939                                                                                                                                                               | SAM2 Hiera-L                                              | SAM2 Hiera-L                                                                                                                            | SAM2 Hiera-L                                                                                                                                                                                                                                                                                                   |
| SAM Parameters             | <ul style="list-style-type: none"> <li>points_per_side: 32</li> <li>pred_iou_thresh: 0.95</li> <li>stability_score_thresh: 0.95</li> <li>min_mask_region_area: 1500</li> </ul> | Default                                                   | Default                                                                                                                                 | Default                                                                                                                                                                                                                                                                                                        |
| Prompt strategy            | Auto-grid.                                                                                                                                                                     | Both passes: Single generic prompt frame on entire video. | 1. First pass: Single generic prompt frame on entire video.<br>2. Second pass: Selected generic prompt frame pool on individual frames. | <ul style="list-style-type: none"> <li>Head: Single generic prompt frame on entire video.</li> <li>RIA: <ul style="list-style-type: none"> <li>a. First pass: Single generic prompt frame on entire video.</li> <li>b. Second pass: Selected generic prompt frame pool on entire video.</li> </ul> </li> </ul> |
| Main post-processing steps | Worm selection via custom ViT classifier, overlap resolution.                                                                                                                  | Mask cleaning.                                            | Mask cleaning, full frame size mapping.                                                                                                 | Mask cleaning, background extraction.                                                                                                                                                                                                                                                                          |

<https://doi.org/10.1371/journal.pcbi.1013441.t003>

background noise [3,11–18]. When thresholding-based methods fail, they necessitate manual curation, frame rejection, or complex parameter tuning, creating analytical bottlenecks and introducing potential bias. On the other hand, methods that utilized deep-learning approaches to replace the segmentation step report the need for condition-specific fine-tuning, making their usage prohibitive for many due to technical or dataset size limitations [6,8,10].

By utilizing the foundational Segment Anything Models (SAMs) [19,20], TWARDIS moves beyond many of the current limitations described, demonstrating remarkable resilience to several common imaging challenges. In our analysis of static morphological images, TWARDIS resolved overlapping worms and noisy backgrounds, across all life stages, where thresholding failed (Fig 1B–F). By integrating SAM with a fine-tuned Vision Transformer classifier for automated worm identification in static morphometry, the pipeline generalized effectively across our different recording setups. This automation not only accelerates analysis but fundamentally enhances reproducibility by eliminating the inter-observer variability inherent in manual annotation or threshold adjustment (Fig 1I) [25,26].

A significant advantage of the TWARDIS compound system is its ability to extract high-definition behavioral data from simple, low-cost recording setups. Detailed postural analysis has often been restricted to high-magnification expensive systems, while wide-field, low-resolution cheap recordings typically allow only centroid tracking [5]. TWARDIS bridges this gap, successfully analyzing both swimming and crawling behaviors even when the worm occupied only ~0.25% of the total field of view (Figs 2F and 3B). The superior segmentation provided by the SAM model enabled the resolution of complex postures (Fig 2G–2K). This precision allows for nuanced behavioral classification, distinguishing, for example, a deep 6-shaped bend (high curvature, small amplitude) from an O-shape (high curvature, high amplitude) (Fig 2G and 2K). Crucially, TWARDIS extracted similar features as other tools, but without discarding any frames or requiring manual corrections, preserving the temporal continuity essential for accurate analysis of behavioral sequences and dynamics, such as the tortuosity analysis we presented in [38] which was only possible with complete tracks (Figs 2B–E and 3C–F).

The application of TWARDIS to in vivo calcium imaging further highlights its precision in challenging dynamic contexts. Extracting fluorescence signals from deforming neural compartments in semi-restricted animals, such as the three axonal compartments of the RIA interneuron, is difficult because the structures deform and shift during recording. Traditional ROI-based analysis, such as that carried out in Fiji, often fails to track these changes, leading to misalignment and artificial attenuation of the fluorescence signal when the structure momentarily exits the ROI or requires extensive handling time for corrections (Fig 4D). TWARDIS, guided by prompt engineering, provides adaptive, frame-by-frame segmentation that captures the morphology of the neuronal compartments accurately. This resulted in more faithful representations of calcium dynamics, evidenced by reduced signal flattening frequent in the RIA loop compartment (Fig 4E).

Furthermore, TWARDIS significantly refines the measurement of behaviors during calcium imaging. We demonstrated that the standard method of fitting an ellipse to a thresholded head often inflates the estimation of head bends (Fig 4F–I). In contrast, TWARDIS leverages accurate whole-body segmentation to calculate the absolute head angle relative to the body midline. This provides a more interpretable and biologically relevant metric (Fig 4H), essential for precise correlation of neural activity and behavioral output.

The systematic noise degradation analysis clarifies the practical boundaries of our approach. The pipeline maintained measurement accuracy at noise levels well beyond those encountered in routine imaging, and its failure mode was mostly conservative, decreasing image quality led to missed detections rather than erroneous segmentation. This analysis also makes it explicit that TWARDIS is not immune to image degradation: images of severely low quality will result in incomplete data extraction. This reinforces that our tool relaxes, but does not eliminate, the need for reasonable imaging conditions.

The success of TWARDIS underscores the advantage of employing an AI compound system built upon a powerful, generalized foundation model. Because the Segment Anything Models are pre-trained on a vast dataset of remarkable quality, they provide exceptional, generalized segmentation capabilities without the need for extensive, modality-specific training required by many other deep learning approaches [3,6,8,11,12,17]. We harness this generalized power and use

targeted strategies (a simple classifier and prompt engineering) with deterministic workflows to achieve high performance. This architecture makes the system easily transferable across different imaging settings and also modular, where both the pre-trained models and the deterministic workflows can be changed or improved. As newer versions of SAM and similar models achieve better performance with smaller computational footprints, TWARDIS will inherit these improvements without requiring fundamental restructuring. Requiring only a minimum of 4GB of RAM for the worm classifier, the TWARDIS workflow is compatible with both CPU and GPU architectures and parallelizable, and thus, like other scalable and GPU-powered tools, it can benefit from high-performance hardware. That said, this removal of the human-in-the-loop makes processing time largely irrelevant, as analyses can be run independently of user input (e.g., overnight).

While TWARDIS offers significant advancements, we recognize areas for future development. As of today, the early v0.1 TWARDIS scripts are available in a public repository (<https://github.com/lillyguisnet/TWARDISv0.1>), but we plan to release TWARDIS as a user-friendly, open-source Python package to facilitate widespread adoption and enhance accessibility in the future. We hope this package could also serve as building blocks for future compound system applications. To further lower the barrier to entry, a web-based graphical interface that leverages the user's own local resources, potentially including browser-based GPU access through technologies such as WebGPU, could make the tools accessible to researchers without command-line experience.

The results showed minor overfitting to our training dataset's optical profile and segmentation failures in degraded images (Fig 1 and Table 2), which can be readily addressed by further fine-tuning our classifier and SAMs with worm images available in public datasets with artificial quality degradation to increase diversity. Alternatively, to completely remove the risk of condition-specific biases, the classification step could be replaced by a small, but general, vision-language model accessed through API services or hosted locally. Also, our age-based identification is optimized for uniformly developing worm strains. Mutants with highly variable growth patterns might fail to be correctly filtered.

Currently, the pipeline is optimized for static multi-worm detection and recordings of single-worm swimming, single-worm crawling and RIA fluorescence imaging. Future iterations could aim to incorporate multi-worm tracking, multi-well swimming, diverse neuronal recording conditions, adaptations for severely abnormal mutants, or even freely moving calcium activity as we presented in [39]. Notably, extending the dynamic analyses to multiple worms simultaneously introduces qualitatively distinct challenges, including maintaining identity across frames, resolving prolonged physical overlaps between moving individuals, understanding occluded worm shapes and re-identifying worms after occlusion events, that require dedicated model architectures beyond segmentation alone. Future dynamic multi-worm approaches should leverage recent computer vision approaches developed for similar problems, such as memory-augmented multi-object trackers, specialized pose estimation frameworks, and methods for skeleton extraction and multi-worm overlap resolution [40–43]. Similarly, amodal completion (inferring the full shape of partially occluded objects from visible portions) could enable counting or morphological measurements in densely populated images [44]. However, this approach requires careful validation in a scientific context, as inferred morphology risks introducing systematic measurement errors, particularly for morphologically abnormal mutants. In addition, adaptations for fluorescence tracking in 3D imaging or freely moving worms would require integration with complementary approaches, such as for motion correction, signal demixing and neural mapping, that address challenges distinct from segmentation.

In conclusion, TWARDIS provides a robust, precise, and automated framework for *C. elegans* phenotyping. By capitalizing on the strengths of foundation large vision models, our compound system overcomes inherent limitations of traditional segmentation, enabling high-definition analysis even under imperfect imaging conditions and with low-cost recording equipment. This framework significantly reduces manual labor and improves data quality, paving the way for large-scale reproducible investigations that focus on experimental design and biological interpretation rather than manual image processing.

## Materials and methods

### Static images acquisition and analysis

**Developmental growth, first set up.** Wild-type N2 *C. elegans* were grown at 20°C and were fed OP50 *E. coli* bacteria as described in Guisnet et al. [45]. Worms were timed by placing egg-laying adults on plates for 2 hours. These worms were moved to fresh plates through the experimental days as needed to keep them fed. Every 24 hours after that, worms were placed on 2% agarose pads in a 5  $\mu$ L drop of 1M sodium azide to immobilize them. After 1 minute, the sodium azide was carefully absorbed with a Kimwipe. No coverslip was added to cover the worms as this caused them to be squished and deformed (particularly in width). Starting at day 3, focus was optimized for the head and tail as worms were too thick to be entirely in focus. Worms were imaged within 10 minutes of being immobilized. Grayscale images were taken with an Olympus XM10 CCD camera mounted on an Olympus MVX10 microscope using an MV PLAPO 2XC objective. TIF images were acquired with a 1376 x 1038 resolution and 10 ms exposure with the cellSens software (Evident).

**Developmental growth, second set up.** *C. elegans* (N2 wild-type strain) were grown at 20°C on nematode growth medium (NGM) agar plates seeded with *Escherichia coli* OP50. Synchronized worms were obtained as previously described [46]. Animals were anesthetized in 0.2% (w/v) levamisole prepared in M9 buffer and mounted directly between a glass slide and a coverslip. Imaging was performed at room temperature using a Leica DMI6000B inverted microscope equipped with a 4 $\times$ /0.10 NA objective lens, a spinning-disk confocal head (Yokogawa CSU10), and an EM-CCD camera (Hamamatsu ImagEM). Image tiling was performed using Metamorph software, and subsequent stitching of the tiles was carried out in ImageJ.

**Thresholding analysis.** For both set-ups, thresholding analysis was tested by using the Default and Otsu thresholding options of the Fiji software [33].

**Automated multi-worm morphological feature extraction from images.** After image acquisition of developmental growth animals, we employed the Segment Anything Model (SAM) [19] to generate automatic mask predictions for each image in our dataset. The ViT-H variant of SAM was utilized, with the following custom parameters: points\_per\_side: 32, pred\_iou\_thresh: 0.95, stability\_score\_thresh: 0.95, min\_mask\_region\_area: 1500. These parameters were empirically determined to optimize mask generation for worm-like objects. The SAM model was implemented using PyTorch [47] and Python 3.10.12 and executed on a CUDA-enabled GPU for efficient processing. For each mask predicted by SAM, we extracted a cutout from the original image. Masks touching image edges were excluded to avoid partial worms.

487 generated cutouts from the developmental growth, first set up acquisitions underwent manual classification by *C. elegans* experts. Each cutout was labeled as either “worm” or “not worm” using the LabelBox web tool with academic access [48]. The classified cutouts were randomly split into training (80%) and validation (20%) sets. We employed transfer learning using a Vision Transformer (ViT-H-14) [49] pre-trained on ImageNet [50] as our base model. The classifier was fine-tuned on our worm/notworm dataset. The “worm” classification included partial worms to improve future usability of the model, although these were later filtered out. The base model’s parameters were frozen, and only the final classification layer was retrained. Data augmentation techniques, including random resized crops and horizontal flips, were applied to the training set. The model was trained for 25 epochs using stochastic gradient descent (SGD) optimization with a learning rate of 0.001 and momentum of 0.9. A step learning rate scheduler was implemented, decaying the learning rate by a factor of 0.1 every 7 epochs. Cross-entropy loss was used as the optimization criterion. Training was conducted on a single GPU using PyTorch. The model achieving the highest validation accuracy (100%) was selected as the final model.

Pre-computed mask predictions from SAM were classified as “worm” or “not worm” by the fine-tuned classifier. To address instances of overlapping worm masks (that is, when there were multiple masks produced for the same worm representing subparts of the body), an adjacency matrix was constructed to identify overlapping regions. Connected components analysis was then performed to group overlapping masks. Within each group, only the mask with the largest area

was retained for further analysis, ensuring each worm was represented only once. Touching worms were automatically resolved by SAM.

For each retained worm mask, we extracted the following morphological features: area (sum of pixels within the mask), perimeter (Euclidean number of pixels of the shape's contour), 100-point interpolated medial axis (the longest path along the worm's centerline identified through graph analysis), length (Euclidean pixel count of the medial axis), width (estimated at each point along the medial axis). For age-based identification in images with mixed-age worms, a size deviation threshold relative to all age-matched worms identified was applied per imaging day.

**Noise degradation analysis.** The original 16-bit grayscale images were corrupted with additive Gaussian noise at six increasing intensity levels. Noise was sampled from a zero-mean Gaussian distribution with standard deviations ( $\sigma$ ) of 655, 1,966, 3,276, 4,587, 6,553, and 13,107 pixel intensity units, corresponding to approximately 1%, 3%, 5%, 7%, 10%, and 20% of the full 16-bit range, respectively. Each set of degraded images was then processed using identical parameters as for the original dataset. Performance at each noise level was evaluated by comparison to the expert manual segmentation, using Pearson correlation coefficient and per-worm percent area agreement (TWARDIS area / manual area  $\times$  100). To distinguish between failure at different pipeline stages, false negative and false positive rates were computed separately for the segmentation step (whether SAM could produce a valid non-edge mask for the worm) and the classification step (whether our ViT classifier correctly classified the worm mask).

**Runtime benchmarking.** 21 microscopy images from the dataset were each processed at four spatial resolutions: 100%, 90%, 70%, and 50% of the original pixel dimensions, 1376x1038, 1238x934, 963x726 and 688x519 pixels, respectively, obtained by downscaling with area-based interpolation.

Runtime was measured for the three main pipeline steps: SAM2 automatic mask generation, ViT binary classification of each candidate mask, and (3) morphological metrics extraction for each detected worm. Metrics extraction was further decomposed into six sub-steps: connected component extraction, contour detection and perimeter measurement, medial axis skeletonization, skeleton endpoint and branchpoint cleaning, and width/length measurements along the pruned skeleton (worm spline).

GPU benchmarks were performed on an NVIDIA RTX 3090 (24 GB VRAM) using CUDA with bfloat16 autocast and TF32 precision. A warmup pass was run before timing. CPU benchmarks were performed on the same machine using 4, 8, and 16 threads, controlling parallelism through PyTorch threads option (`torch.set_num_threads`), OpenCV threads (`cv2.setNumThreads`), and environment variables (`MKL_NUM_THREADS`). The skeleton cleaning step was additionally parallelized using a thread pool (`concurrent.futures.ThreadPoolExecutor`). All timing measurements used `time.perf_counter()`. Classification was performed in-memory to avoid disk I/O artifacts in timing measurements.

## Droplet swimming recording and analysis

**Droplet swimming recording set up.** Wild-type N2 *C. elegans* were grown at 20°C and were fed OP50 *E. coli* bacteria as described in Guisnet et al. [45]. A large 10 cm diameter plexiglass dish that was first cleaned with ethanol and a Kimwipe. It was then grounded by bringing the surface in very close proximity to a metal cabinet. This removed static from the plexiglass helping small droplets to stick to the surface. 50  $\mu$ L of filtered NGML was then pipetted onto the center of the dish. Young adult worms were selected from maintenance plates and left to roam off food on an empty agar plate for 2 minutes before being moved to the droplet with a platinum worm pick.

Uncompressed AVI videos were recorded for 30 seconds at 10 fps with a FLIR Blackfly (BFS-U3-51S5M-C) 5.0 MP camera and Spinnaker SDK software (Teledyne) at 2448 x 2048 resolution. The recording dish was held upside down with a custom made rig and recorded from below. An LED light ring of the same size as the plate was positioned centered with the plate, but further up ( $\sim$  8 centimeters), to minimize light refraction artifacts in the droplet and dish plastic.

**Automated quantitative analysis of swimming behavior.** Worm segmentation was performed using the Segment Anything Model 2 (SAM2) [20]. The model was initialized using a pre-trained checkpoint (`sam2_hiera_large.pt`) and GPU

acceleration was employed using CUDA, with TensorFloat-32 compatible hardware. A generic prompt frame was added to all videos and used to automatically guide segmentation on the worm object. Segmentation was propagated backwards from the added prompt frame through the video. To improve segmentation accuracy, a second pass of segmentation was performed on cropped regions of interest. The SAM2 model was applied again to these cropped frames, using a new generic prompt frame matching the crop size. Segmentation was again propagated backwards through the cropped frames. The resulting worm high-definition masks were used for further analysis.

SAM2 masks from the high-definition segmentation were cleaned using morphological operations to remove occasional small particles. Worm skeletons were extracted by mask skeletonization. A custom algorithm was implemented to correct self-touching skeletons, ensuring a single, continuous skeleton for each frame. Skeleton points were smoothed using spline interpolation to 100 points. In addition to those described in the morphological extraction section above, several other metrics were extracted from the skeletons. Gaussian-weighted curvature was calculated along the smoothed skeleton using a window size of 50 points and sigma of 10. Amplitude along the worm's body was measured as the perpendicular distance from skeleton points to the worm's centerline (defined as the straight line from head to tail). Worms were classified by shape. Classification was based on curvature and amplitude values (Table 4). Thresholds for classification were empirically determined. Worms were classified as turned on the z-plane, if their length was smaller by more than 1 standard deviation from the video mean. Wavelength was based on curvature peaks for S-shaped worms, set at 2 worm length for C-shapes, and none for straight or turned worms. Wave number was calculated as the ratio of worm length to wavelength. Spatial frequency analysis was performed using Fast Fourier Transform on the curvature data. Temporal frequencies were calculated from the curvature data. The power spectral density (PSD) of the normalized curvature time series was calculated using Welch's method with a window size of 30 and an overlap of 25. A sliding window analysis was conducted to identify the dominant frequencies within each segment. Peaks in the PSD were detected, and the frequency corresponding to the peak with the maximum power was selected as the dominant frequency. If no peaks were found, the frequency with the maximum power was used. The dominant frequencies were interpolated across the entire time series to ensure a frequency value for each frame.

## Crawling recording and analysis

**Crawling recording set up.** Wild-type N2 *C. elegans* were grown under standard conditions at 20°C and were fed OP50 *E. coli* bacteria as described in Guisnet et al. [45]. Young adult worms were selected from maintenance plates and left to roam off food on an empty agar plate for 2 minutes before being recorded. AVI videos were recorded for 1 minute with the same setup and settings as described for swimming.

**Tierpsy analysis.** The version of Tierpsy available on April 23rd 2026 was installed via Docker and used through GUI by following the instructions presented on the GitHub page (<https://github.com/Tierpsy/tierpsy-tracker>) [51]. Thresholding parameters were selected to optimize for real worm detection, while minimizing false positives from background artifacts, by manually iterating through 5 randomly selected videos. These parameters were then used to carry out the analysis in batch on all 114 recordings. All Tierpsy defaults were maintained, except for the use\_nn\_filter option which was changed to False, since keeping this default option resulted in zero worm detection.

**Table 4. Shape classification.**

|           | Body shape |            |          |          |      |
|-----------|------------|------------|----------|----------|------|
|           | 6          | S          | U        | C        | O    |
| Amplitude | Small      | Very small | High     | Moderate | High |
| Curvature | Very high  | Very small | Moderate | Moderate | High |

<https://doi.org/10.1371/journal.pcbi.1013441.t004>

To quantify Tierspy analysis success, we used the `timeseries_data` from the `featuresN` output file. A frame was qualified having failed to be analyzed by Tierspy if all features were NaN, regardless of how many or which features were provided.

**Automated single worm high-definition tracking.** Worm tracking was performed in a similar way as described for swimming with a few differences for path tracking and addressing more complicated body postures. For the high-definition segmentation, two generic prompt frames with complicated shapes were added on cropped frames and the SAM2 model was reinitialized and applied to each of the cropped frames one by one. Segmentation was propagated through these short cropped sequences to generate the high-definition masks. The resulting cropped masks were then resized and mapped back to the original frame dimensions for further analysis.

Skeleton morphological features were extracted as described in the swimming section with additional extraction of head bends. Due to the low resolution of the worm in the large image space, head and tail differentiation was not possible by traditional method of brightness difference [37]. A temporal tracking algorithm was employed instead, utilizing a 5 frame sliding window to analyze movement patterns. Endpoint grouping was optimized using the Hungarian algorithm, with the more mobile group designated as the head based on cumulative displacement. Error correction was applied for sudden jumps by cross-referencing with recent history and worm movement. Then, head bending magnitude was quantified by calculating the angle between the head endpoint and the main body axis.

For path analysis, the segmentation masks from previous shape analysis were used to obtain the centroid position for every frame using center-of-mass computation. To reduce measurement noise inherent from changes in the worm's shape, centroid trajectories were smoothed using Savitzky-Golay filtering, which preserves the underlying movement patterns while eliminating high-frequency noise. Worm movement was classified into three behavioral states: forward locomotion, backward locomotion, and stationary periods. For each frame, instantaneous velocity vectors were calculated from the change in centroid position between consecutive frames. The orientation of the movement was determined by the vector from the body centroid relative to the head position. When centroid direction was towards the head, movement was classified as forward; centroid direction away from the head was classified as backward movement, while low-magnitude velocities (less than 0.6 pixels per frame) were classified as stationary behavior. To ensure accurate movement classification, some corrections were applied. These arose during stationary and turning periods, where head detection was less accurate. Frames with movement classifications inconsistent with surrounding temporal context were identified and corrected. Continuous periods of each movement type were identified as behavioral bouts. For each bout, both temporal duration (in frames) and spatial extent (total distance traveled) were calculated. Movement metrics were calculated for each analyzed video, including, for example, total path length, maximum displacement from origin, proportion of time spent in each behavioral state, and average bout duration for forward and backward movement.

## Calcium imaging in semi-restricted worms

**Microfluidic imaging of sensory-evoked calcium dynamics in RIA.** For recordings of RIA interneuron calcium dynamics, the strain MMH99 (`pglr-3a::GCaMP3.3`) was used. The worms were grown under standard conditions at 20°C, fed OP50 *E. coli* bacteria and assayed at the early adult stage [52].

Time-lapse fluorescence imaging of the RIA interneuron in living semi-restrained animals was performed in a microfluidic device as described in Hendricks et al. [53]. The aforementioned MMH99 strain selectively expressing GCaMP3.3 in RIA was used to measure calcium dynamics within the three distinct axonal compartments of RIA. Streams of liquid alternating between neutral NGM buffer and the attractive odorant IAA (isoamyl alcohol) were passed in front of the animal's freely moving head to elicit sensory-evoked calcium events [54]. 512 x 512 TIFF stack recordings were acquired for individual animals over 1 minute at 10 frames per seconds and 100 ms exposures on an Olympus IX83 inverted microscope using a 40x U-PlanS Apo (N.A. 1.25) silicone immersion objective, Hamamatsu Orca Flash 4.0LT sCMOS camera, with the cellSens software (Evident). Automatic stimulus stream switches every 10 seconds were controlled by a solenoid pinch valve array (Automate Scientific).

**Manual segmentation of axonal compartments and head bending extraction.** The brightness of the RIA compartments and the worm's head position were manually extracted for comparison using the Fiji software as described in Ouellette et al. [32].

**Automated segmentation of axonal compartments and head bending extraction.** Full-frame videos were processed to extract a region of interest centered on the RIA interneuron. Like for the other modalities, the Segment Anything Model 2 (SAM2) video predictor was initialized using the Hiera Large checkpoint. For each video, a single annotation point was placed on the RIA region in the final frame. This annotation was propagated backwards through all frames. The center of mass for each RIA mask was calculated across all frames to create a fixed crop window of  $110 \times 110$  pixels around the RIA region throughout the video sequence. To simultaneously segment the three anatomical structures within the cropped RIA region (nerve ring dorsal (nrD), nerve ring ventral (nrV), and loop), a prompt management system was developed to enable consistent segmentation across videos. Expertly annotated prompt frames were incorporated into the video sequences to guide the segmentation model. SAM2 was reinitialized with the augmented video sequences, and structure-specific prompts. A comprehensive quality analysis was performed to identify segmentation artifacts, including empty masks, oversized regions, and overlapping structures. Additional prompt frames were added to the prompt pool if refinement was necessary. On rare occasions, sporadic frames with a missing mask were filled by interpolation. Occasional small noise artifacts were removed through connected component analysis. Brightness measurements were extracted from the segmentation masks by extracting the pixel intensities from the corresponding grayscale images.

Background correction was performed by sampling 100 background regions located at least 40 pixels away from any segmented object, as determined by distance transform analysis. The mean background intensity was subtracted from object measurements to obtain corrected brightness values. Since animals were positioned either on their left or right side in the chip, the spatial relationship between the loop structure and nrD was analyzed to adjust head bending angles as dorsal or ventral.

Full-body segmentation was performed on the original full-frame videos to enable postural analysis. A single annotation point was placed on the organism body and propagated through all frames using SAM2. The resulting masks captured the complete body, which served as input for subsequent head angle analysis. Morphological skeletonization was applied to whole-body masks to extract one-pixel-wide centerlines representing the organism's medial axis. The angle between head and body vectors (anterior and posterior portions of the skeleton) was computed for each frame. Temporal smoothing was applied to remove measurement noise while preserving genuine behavioral transitions. A small temporal sliding window smoothing of 3 frames was applied to remove noise from small variations in body segmentation outline. Final head angle sign was adjusted based on the organism's orientation determined in the previous step.

**Human head angle measurement.** 50 frames were randomly selected across the calcium imaging dataset described above. For each frame, a human expert measured the head angle by manually identifying the head tip and a straight reference line along the body midline using the angle measurement tool in Fiji [33]. The same 50 frames were fetched from both the Fiji ellipse-fitting and the TWARDIS pipeline outputs. Angles were orientation-corrected and translated to match a unified  $0-180^\circ$  convention, where  $90^\circ$  represents a straight head. Absolute differences from the human-measured angle were computed for each automated method.

## Hardware and code

All image analysis, model training and model inference was done on a local workstation with Ubuntu 22.04 and one Nvidia RTX 3090 GPU. To speed up processing, some operations were scaled to 48 parallel cores. For image analysis, the Meta SAM model `sam_vit_h_4b8939` was used. For video analysis, the Meta SAM2 repository was cloned and the model downloaded from the July 28<sup>th</sup> 2024 version (<https://github.com/facebookresearch/sam2/>). All other

data analyses and visualizations were conducted using Python (v3.12.3), uv for environment management and the following openly accessible libraries: h5py, matplotlib, numpy, OpenCV, pandas, PyTorch, scikit-image, scikit-learn, scipy, seaborn and tqdm [47,55–66]. All code is freely accessible at <https://github.com/lillyguisnet/TWARDISv0.1>. Our fine-tuned worm classifier is available on HuggingFace (<https://huggingface.co/lillyguisnet/celegans-classifier-vit-h-14-finetuned>).

## Supporting information

**S1 Fig. Runtime benchmarks across hardware configurations and image resolutions.** 21 images were processed at four spatial resolutions (100%, 90%, 70%, and 50% of the original pixel dimensions; 1376x1038, 1238x934, 963x726 and 688x519 pixels, respectively) to evaluate how input size affects runtime. Performance was measured on a GPU and on CPU with 4, 8, and 16 threads. (A) SAM segmentation time per image. Image size has minimal effect. (B) ViT-classifier classification time per mask. Mask scale has no effect on classification time, as all inputs are resized to 518×518 pixels before inference. (C) Per-worm metrics extraction sub-steps (CPU only). Mask Cleaning (connected component extraction), Area/Perimeter (contour detection), Skeletonization (medial axis transform), Skeleton Cleanup (non-branching spline extraction) and Length/Width (distance measurements along the clean spline). High dependence on image size. Low benefit from parallelization as functions are already preoptimized. Skeleton Cleanup dominates total metrics time, with high variance driven by skeleton complexity. (D) Skeleton Cleanup time as a function of the number of graph nodes (skeleton endpoints and branchpoints), showing quadratic scaling. Each point represents one worm at one image scale; colors indicate thread count. Metrics extraction would benefit most from algorithmic optimization and GPU implementation or multi-image CPU parallelization.

(PDF)

**S2 Fig. Scatterplot of Fiji- versus TWARDIS-estimated head angles with sign ignored.** Fiji estimations (-1, 1) are converted to the (0, 90) range for comparison. 0 represents a perfectly straight head. 90 degrees represent maximal bending.  $r=0.746$ .  $n=13,486$  frames from 22 recordings.

(PDF)

**S3 Fig. Comparison of human, Fiji, and TWARDIS extracted head angle measurements.** Head angles from 50 randomly selected frames measured manually by a human expert observer, Fiji traditional ellipse-fitting, and TWARDIS skeletonization approach, sorted by human-measured angle. Dashed line indicates 90°. Values were corrected for worm orientation (ventral/dorsal - left/right) and translated to the same 0–180° range for comparison.

(PDF)

**S1 Movie. Droplet swimming recording with worm mask and body curvature kymograph.** Background image of original recording with resolution halved. Picture-in-picture in top left corner shows the final worm mask for every frame. Bottom live plot shows Fig 2E, the gaussian-weighted curvature intensity of 100 interpolated points along the worm's body per frame (same legend as Fig 2E).

(MP4)

**S2 Movie. Crawling recording with worm mask, head angle and speed.** Background image of original recording with resolution halved. Picture-in-picture in top left corner shows the final worm mask for every frame. Dots of the worm position are overlaid on the original frames to represent the worm trace through time. Top live plot shows Fig 3E: the head bending angle relative to body in degrees per frame where 0 indicates a straight head relative to the body, and positive and negative values represent bends on either side, but ventral and dorsal sides are not distinguished, and points along the trace represent the detected peaks and troughs used for head bend metrics. Bottom live plot shows Fig 3F: the worm speed by frame, where positive values represent forward motion, negative values

represent backward motion and stationary motion is determined by absolute speed less than 0.6 pixel/second (shaded area). This movement classification is also color-coded as in Fig 3C for the dot overlays and for every frame at the bottom left of the original image.

(MP4)

**S3 Movie. RIA calcium imaging recording with axonal compartment segmentation, head angle, and calcium activity.** Background image of original recording with no modification. Final body segmentation is overlaid on the original frames in purple for every frame. Head angle in degrees extracted from body segmentation is presented for every frame over the head of the worm and in the top-most live plot (same as Fig 4G). Boxed picture-in-picture shows the final masks at every frame for each segmented axonal compartment: nrD (orange), nrV (blue) and loop (green). Resulting normalized fluorescence intensity traces for each compartment are shown in the bottom three live plots.

(MP4)

**S1 Data. Human and TWARDIS-extracted area and perimeter in pixels.**

(CSV)

**S2 Data. Time-series metrics of a 10 frames-per-second one minute recording of a wild-type worm swimming in a liquid droplet presented in Fig 2.**

(CSV)

**S3 Data. Time-series metrics of a 10 frames-per-second one minute recording of a wild-type worm crawling on an agar plate with food presented in Fig 3.**

(CSV)

**S4 Data. Fiji- and TWARDIS-extracted head bending and fluorescence time-series values of 10 frames-per-second one minute recordings of wild-type worms in a microfluidic device expressing GCaMP3.3 in RIA.**

(CSV)

**S5 Data. Normalized fluorescence intensity and average absolute difference from mean for Fiji- and TWARDIS-extracted data of the loop compartment.**

(CSV)

**S6 Data. Fiji- and TWARDIS-estimated head angles centered normalized from 0 to 180 degrees.**

(CSV)

**S7 Data. Gaussian noise degradation extracted values.**

(CSV)

**S8 Data. Human, Fiji, and TWARDIS extracted head angle measurements for 50 randomly selected frames.**

(CSV)

**S9 Data. Runtime by hardware for SAM image segmentation by image size.**

(CSV)

**S10 Data. Runtime by hardware for mask classification by image size.**

(CSV)

**S11 Data. Runtime by hardware for metrics extraction by image size.**

(CSV)

## Acknowledgments

We thank Laeya Baldini and Stephanie C. Weber from McGill University for sharing their developmental growth images used as the second dataset in this publication. Maxime Rivest for guiding early readings on computer vision and troubleshooting workstation issues. Meta FAIR for releasing the SAM models open-weights. All members of the lab for helpful discussions and comments. Strains were provided by the *Caenorhabditis* Genetics Center (CGC), which is funded by NIH Office of Research Infrastructure Programs (P40 OD010440).

## Author contributions

**Conceptualization:** Aurélie Guisnet, Michael Hendricks.

**Data curation:** Aurélie Guisnet.

**Formal analysis:** Aurélie Guisnet.

**Funding acquisition:** Aurélie Guisnet, Michael Hendricks.

**Investigation:** Aurélie Guisnet.

**Methodology:** Aurélie Guisnet.

**Resources:** Aurélie Guisnet.

**Software:** Aurélie Guisnet.

**Supervision:** Michael Hendricks.

**Validation:** Aurélie Guisnet.

**Visualization:** Aurélie Guisnet.

**Writing – original draft:** Aurélie Guisnet, Michael Hendricks.

**Writing – review & editing:** Aurélie Guisnet, Michael Hendricks.

## References

1. Cronin CJ, Mendel JE, Mukhtar S, Kim Y-M, Stirbl RC, Bruck J, et al. An automated system for measuring parameters of nematode sinusoidal movement. *BMC Genet.* 2005;6:5. <https://doi.org/10.1186/1471-2156-6-5> PMID: [15698479](#)
2. Kitto E, Dean J, Leiser S. findWormz is a user-friendly automated fluorescence quantification method for *C. elegans* research. *MicroPubl Biol.* 2025;2025. <https://doi.org/10.17912/micropub.biology.001562> PMID: [40353139](#)
3. Hakim A, Mor Y, Toker IA, Levine A, Neuhoof M, Markovitz Y, et al. WorMachine: machine learning-based phenotypic analysis tool for worms. *BMC Biol.* 2018;16(1):8. <https://doi.org/10.1186/s12915-017-0477-0> PMID: [29338709](#)
4. Moore BT, Jordan JM, Baugh LR. WormSizer: high-throughput analysis of nematode size and shape. *PLoS One.* 2013;8(2):e57142. <https://doi.org/10.1371/journal.pone.0057142> PMID: [23451165](#)
5. Perni M, Casford S, Aprile FA, Nollen EA, Knowles TPJ, Vendruscolo M. Automated behavioral analysis of large *C. elegans* populations using a wide field-of-view tracking platform. *J Vis Exp.* 2018. <https://doi.org/10.3791/58643>
6. Jung S-K. AniLength: GUI-based automatic worm length measurement software using image processing and deep neural network. *SoftwareX.* 2021;15:100795. <https://doi.org/10.1016/j.softx.2021.100795>
7. Restif C, Ibáñez-Ventoso C, Vora MM, Guo S, Metaxas D, Driscoll M. CeleST: computer vision software for quantitative analysis of *C. elegans* swim behavior reveals novel features of locomotion. *PLoS Comput Biol.* 2014;10(7):e1003702. <https://doi.org/10.1371/journal.pcbi.1003702> PMID: [25033081](#)
8. Bates K, Le KN, Lu H. Deep learning for robust and flexible tracking in behavioral studies for *C. elegans*. *PLoS Comput Biol.* 2022;18(4):e1009942. <https://doi.org/10.1371/journal.pcbi.1009942> PMID: [35395006](#)
9. Wang L, Kong S, Pincus Z, Fowlkes C. Celeganser: Automated analysis of nematode morphology and age. *arXiv [cs.CV]*. 2020. Available: <http://arxiv.org/abs/2005.04884>
10. Pan Y, Huang Z, Cai H, Li Z, Zhu J, Wu D, et al. WormCNN-Assisted Establishment and Analysis of Glycation Stress Models in *C. elegans*: Insights into Disease and Healthy Aging. *Int J Mol Sci.* 2024;25(17):9675. <https://doi.org/10.3390/ijms25179675> PMID: [39273622](#)

11. Weheliye WH, Rodriguez J, Feriani L, Javer A, Uhlmann V, Brown AEX. An improved neural network model enables worm tracking in challenging conditions and increases signal-to-noise ratio in phenotypic screens. *Animal Behavior and Cognition*. bioRxiv. 2024. Available: <https://www.biorxiv.org/content/10.1101/2024.12.20.629717v2.full.pdf>
12. Alonso A, Kirkegaard JB. Fast detection of slender bodies in high density microscopy data. *Commun Biol*. 2023;6(1):754. <https://doi.org/10.1038/s42003-023-05098-1> PMID: 37468539
13. Barlow IL, Feriani L, Minga E, McDermott-Rouse A, O'Brien TJ, Liu Z, et al. Megapixel camera arrays enable high-resolution animal tracking in multiwell plates. *Commun Biol*. 2022;5(1):253. <https://doi.org/10.1038/s42003-022-03206-1> PMID: 35322206
14. Akpu CH, Wei H, Hong X. Deep learning in automated worm identification and tracking for *C. elegans* mating behaviour analysis. *Lecture Notes in Computer Science*. Cham: Springer Nature Switzerland; 2025. p. 113–28.
15. Ding M, Liu J, Luo Y, Tang J. A bilayer segmentation-recombination network for accurate segmentation of overlapping *C. elegans*. *Appl Soft Comput*. 2025;180:113459. <https://doi.org/10.1016/j.asoc.2025.113459>
16. Hebert L, Ahamed T, Costa AC, O'Shaugnessy L, Stephens GJ. WormPose: Image synthesis and convolutional networks for pose estimation in *C. elegans*. 2020. 2020.07.09.193755 p. <https://doi.org/10.1101/2020.07.09.193755>
17. Dong B, Chen W. A high precision method of segmenting complex postures in *Caenorhabditis elegans* and deep phenotyping to analyze lifespan. *Sci Rep*. 2025;15(1):8870. <https://doi.org/10.1038/s41598-025-93533-0> PMID: 40087519
18. Banerjee SC, Khan KA, Sharma R. Deep-Worm-Tracker: Deep Learning Methods for Accurate Detection and Tracking for Behavioral Studies in *C. elegans*. *Animal Behavior and Cognition*. bioRxiv. 2022. Available: <https://www.biorxiv.org/content/10.1101/2022.08.18.504475v1.full.pdf>
19. Kirillov A, Mintun E, Ravi N, Mao H, Rolland C, Gustafson L, et al. Segment Anything. *arXiv [cs.CV]*. 2023. Available: <http://arxiv.org/abs/2304.02643>
20. Ravi N, Gabeur V, Hu Y-T, Hu R, Ryali C, Ma T, et al. SAM 2: Segment Anything in Images and Videos. *arXiv preprint arXiv:2408.00714*. 2024. Available: <https://arxiv.org/abs/2408.00714>
21. Archit A, Freckmann L, Nair S, Khalid N, Hilt P, Rajashekar V, et al. Segment Anything for Microscopy. *Nat Methods*. 2025;22(3):579–91. <https://doi.org/10.1038/s41592-024-02580-4> PMID: 39939717
22. Zhu J, Qi Y, Wu J. Medical SAM 2: Segment medical images as video via Segment Anything Model 2. *arXiv [cs.CV]*. 2024. Available: <http://arxiv.org/abs/2408.00874>
23. Ma J, He Y, Li F, Han L, You C, Wang B. Segment anything in medical images. *Nat Commun*. 2024;15:654.
24. Labocha MK, Jung S-K, Aleman-Meza B, Liu Z, Zhong W. WormGender - Open-Source Software for Automatic *Caenorhabditis elegans* Sex Ratio Measurement. *PLoS One*. 2015;10(9):e0139724. <https://doi.org/10.1371/journal.pone.0139724> PMID: 26421844
25. Bonnard E, Liu J, Zjadic N, Alvarez L, Scholz M. Automatically tracking feeding behavior in populations of foraging worms. *bioRxiv*. 2022. 2022.01.20.477072 p. <https://doi.org/10.1101/2022.01.20.477072>
26. Buckingham SD, Sattelle DB. Fast, automated measurement of nematode swimming (thrashing) without morphometry. *BMC Neurosci*. 2009;10:84. <https://doi.org/10.1186/1471-2202-10-84> PMID: 19619274
27. Jung SK. AniWellTracker: Image analysis of small animal locomotion in multiwell plates. *Appl Sci*. 2023;13:2274.
28. Vedantham K, Niu L, Ma R, Connelly L, Nagella A, Wang SJ, et al. Track-A-Worm 2.0: A Software Suite for Quantifying Properties of *C. elegans* Locomotion, Bending, Sleep, and Action Potentials. *Neuroscience*. bioRxiv. 2024. Available: <https://www.biorxiv.org/content/10.1101/2024.09.12.612524v1.full.pdf>
29. Van Camp BT, Zapata QN, Curran SP. WormRACER: Robust Analysis by Computer-Enhanced Recording. *Geroscience*. 2025;47(3):5377–87. <https://doi.org/10.1007/s11357-025-01631-8> PMID: 40140154
30. Yemini EI, Brown AEX. Tracking single *C. elegans* using a USB microscope on a motorized stage. *arXiv [q-bio.QM]*. 2014. Available: <http://arxiv.org/abs/1407.6889>
31. Perni M, Challa PK, Kirkegaard JB, Limbicker R, Koopman M, Hardenberg MC, et al. Massively parallel *C. elegans* tracking provides multi-dimensional fingerprints for phenotypic discovery. *J Neurosci Methods*. 2018;306:57–67. <https://doi.org/10.1016/j.jneumeth.2018.02.005> PMID: 29452179
32. Ouellette M-H, Desrochers MJ, Gheta I, Ramos R, Hendricks M. A Gate-and-Switch Model for Head Orientation Behaviors in *Caenorhabditis elegans*. *eNeuro*. 2018;5(6):ENEURO.0121-18.2018. <https://doi.org/10.1523/ENEURO.0121-18.2018> PMID: 30627635
33. Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, et al. Fiji: an open-source platform for biological-image analysis. *Nat Methods*. 2012;9(7):676–82. <https://doi.org/10.1038/nmeth.2019> PMID: 22743772
34. Mouchiroud L, Sorrentino V, Williams EG, Cornaglia M, Frochoux MV, Lin T, et al. The Movement Tracker: A Flexible System for Automated Movement Analysis in Invertebrate Model Organisms. *Curr Protoc Neurosci*. 2016;77:8.37.1-8.37.21. <https://doi.org/10.1002/cpns.17> PMID: 27696358
35. Oswal N, Martin OMF, Stroustrup S, Bruckner MAM, Stroustrup N. A hierarchical process model links behavioral aging and lifespan in *C. elegans*. *PLoS Comput Biol*. 2022;18(9):e1010415. <https://doi.org/10.1371/journal.pcbi.1010415> PMID: 36178967
36. Jung S-K, Aleman-Meza B, Riepe C, Zhong W. QuantWorm: a comprehensive software package for *Caenorhabditis elegans* phenotypic assays. *PLoS One*. 2014;9(1):e84830. <https://doi.org/10.1371/journal.pone.0084830> PMID: 24416295
37. Yemini E, Jucikas T, Grundy LJ, Brown AEX, Schafer WR. A database of *Caenorhabditis elegans* behavioral phenotypes. *Nat Methods*. 2013;10(9):877–9. <https://doi.org/10.1038/nmeth.2560> PMID: 23852451

38. Guisnet A, Halaby N, Rivest M, Romero Quineche B, Hendricks M. The impact of rearing environment on *C. elegans*: phenotypic, transcriptomic and intergenerational responses to 3D enriched habitats. *Biology Open*. 2026;15(2):null. <https://doi.org/10.1242/bio.062282>
39. Hendricks M, Wittekindt SN, Owens H, Wittekindt L, Guisnet A. A novel epifluorescence microscope design and software package to record naturalistic behaviour and cell activity in freely moving *Caenorhabditis elegans*. *bioRxiv*. 2025. <https://www.nature.com/articles/s41467-026-72709-w#citeas>
40. Ao J, Jiang Y, Ke Q, Ehinger KA. Open-World Amodal Appearance Completion. *arXiv [cs.CV]*. 2024. Available: <http://arxiv.org/abs/2411.13019>
41. Cai J, Xu M, Li W, Xiong Y, Xia W, Tu Z, et al. MeMOT: Multi-object tracking with memory. *arXiv [cs.CV]*. 2022. Available: <http://arxiv.org/abs/2203.16761>
42. Lee T, Wen B, Kang M, Kang G, Kweon IS, Yoon K-J. Any6D: Model-free 6D pose estimation of novel objects. *arXiv [cs.CV]*. 2025. Available: <http://arxiv.org/abs/2503.18673>
43. Salisu S, Danyaro KU, Nasser M, Hayder IM, Younis HA. Review of models for estimating 3D human pose using deep learning. *PeerJ Comput Sci*. 2025;11:e2574. <https://doi.org/10.7717/peerj-cs.2574> PMID: 40062308
44. Bhattacharjee SS, Campbell D, Shome R. Believing is Seeing: Unobserved Object Detection using Generative Models. *arXiv [cs.CV]*. 2024. Available: <http://arxiv.org/abs/2410.05869>
45. Guisnet A, Maitra M, Pradhan S, Hendricks M. Three-Dimensional Fruit Tissue Habitats for Culturing *Caenorhabditis elegans*. *Curr Protoc*. 2021;1(11):e288. <https://doi.org/10.1002/cpz1.288> PMID: 34767311
46. Zellag RM, Zhao Y, Poupart V, Singh R, Labbé JC, Gerhold AR. CentTracker: a trainable, machine-learning-based tool for large-scale analyses of *Caenorhabditis elegans* germline stem cell mitosis. *Mol Biol Cell*. 2021;32:915–30.
47. Ansel J, Yang E, He H, Gimelshein N, Jain A, Voznesensky M, et al. PyTorch 2: Faster machine learning through dynamic python bytecode transformation and graph compilation. *Proceedings of the 29th ACM International Conference on Architectural Support for Programming Languages and Operating Systems, Volume 2*. New York (NY): ACM; 2024. pp. 929–947.
48. Labelbox. [cited 5 Sep 2024]. Available: <https://labelbox.com>
49. Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, et al. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. *arXiv [cs.CV]*. 2020. Available: <http://arxiv.org/abs/2010.11929>
50. Deng J, Dong W, Socher R, Li L-J, Li K, Fei-Fei L. ImageNet: A large-scale hierarchical image database. *2009 IEEE Conference on Computer Vision and Pattern Recognition*. 2009. p. 248–255.
51. Javer A, Ripoll-Sánchez L, Brown AE. Powerful and interpretable behavioural features for quantitative phenotyping of *Caenorhabditis elegans*. *Philosophical Transactions of the Royal Society B: Biological Sciences*. 2018;373(1758):20170375. <https://doi.org/https://www.biorxiv.org/content/10.1098/rstb.2017.0375>
52. Brenner S. The genetics of *Caenorhabditis elegans*. *Genetics*. 1974;77(1):71–94. <https://doi.org/10.1093/genetics/77.1.71> PMID: 4366476
53. Hendricks M, Ha H, Maffey N, Zhang Y. Compartmentalized calcium dynamics in a *C. elegans* interneuron encode head movement. *Nature*. 2012;487(7405):99–103. <https://doi.org/10.1038/nature11081> PMID: 22722842
54. San-Miguel A, Lu H. Microfluidics as a tool for *C. elegans* research. *WormBook*. 2013;:1–19. <https://doi.org/10.1895/wormbook.1.162.1> PMID: 24065448
55. McKinney W. Data Structures for Statistical Computing in Python. In: van der Walt S, Millman J, editors. *Proceedings of the 9th Python in Science Conference*. 2010. p. 56–61.
56. Hunter JD. Matplotlib: A 2D Graphics Environment. *Comput Sci Eng*. 2007;9(3):90–5. <https://doi.org/10.1109/mcse.2007.55>
57. Harris CR, Millman KJ, van der Walt SJ, Gommers R, Virtanen P, Cournapeau D, et al. Array programming with NumPy. *Nature*. 2020;585(7825):357–62. <https://doi.org/10.1038/s41586-020-2649-2> PMID: 32939066
58. HDF5 for Python — h5py 3.14.0 documentation. [cited 11 Aug 2025]. Available: <https://docs.h5py.org/en/stable/>
59. Bradski G. The OpenCV Library. *Dr Dobb's Journal of Software Tools*. 2000.
60. da Costa-Luis C, Larroque SK, Altendorf K, Mary H, richardsheridan, Korobov M, et al. tqdm: A fast, Extensible Progress Bar for Python and CLI. *Zenodo*. 2024. <https://doi.org/10.5281/ZENODO.14231923>
61. Waskom M. seaborn: statistical data visualization. *JOSS*. 2021;6(60):3021. <https://doi.org/10.21105/joss.03021>
62. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat Methods*. 2020;17(3):261–72. <https://doi.org/10.1038/s41592-019-0686-2> PMID: 32015543
63. The python language reference. In: *Python documentation [Internet]*. [cited 11 Aug 2025]. Available: <https://docs.python.org/3/reference/>
64. Astral Docs. 2025 [cited 12 Aug 2025]. Available: <https://docs.astral.sh/uv/>
65. van der Walt S, Schönberger JL, Nunez-Iglesias J, Boulogne F, Warner JD, Yager N, et al. scikit-image: image processing in Python. *PeerJ*. 2014;2:e453. <https://doi.org/10.7717/peerj.453> PMID: 25024921
66. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O. Scikit-learn: Machine Learning in Python. *J Mach Learning Res*. 2011;12:2825–30.