WormNet: An Automated Deep Learning Platform for Robust and High-Throughput *C. elegans* Behavior Analysis

Qingyu Shi, Yining Shi, Shiyan Li, Lei Fu, Jinxin Gu, Yixue Qiao, Ian Sandall, Eng Gee Lim, Pengfei Song

***Abstract*—Caenorhabditis elegans is a model organism widely used in genetics, neuroscience, aging, and toxicology, where quantitative behavior analysis is essential for high-throughput screening and phenotype assessment. Automated worm tracking and analysis software based on traditional image processing has been developed to reduce manual workload, but such methods still degrade markedly under low contrast, lighting drift, and dense multi-worm conditions, limiting real-world scalability and robustness. Based on microscope imaging and deep learning, we develop WormNet, an integrated hardware–software platform for automated *C. elegans* behavior analysis with enhanced accuracy and throughput. The system combines a temperature-controlled motorized stage, bright-field microscope, and industrial camera, enabling stable acquisition of multi-worm videos under controlled experimental conditions. On the algorithmic side, a multi-scale backbone with a simplified snake convolution (SimSnake Conv) module is used to accurately detect worms, while a multi-scale cross-modal gated fusion (MCGF) module fuses detection features to refine instance masks. A downstream tracking and skeletonization pipeline then automatically calculates morphology and motility parameters, including body length, width, speed, and bending angle. Through experiments on two representative datasets, WormDrop-Micro and WormDish-Real, the system achieves maximum to 1.90% improvement in AP_{50} and 28.50% improvement in $mAP_{50–95}$ over state-of-the-art (SOTA) detectors, and in terms of segmentation accuracy, the Dice coefficient was improved by up to 1.71%, and mIoU by 3.00%. The WormNet pipeline further achieved a processing speed of 20.45 FPS and supported batch analysis of multi-worm videos containing up to approximately 30 worms per field of view. With this performance, WormNet enables robust, fully automated, and high-throughput *C. elegans* behavior analysis under diverse experimental conditions, making large-scale drug and toxicology screening studies feasible.**

***Note to Practitioners*—In the field of worm behavior and drug screening, many laboratories still rely on manual interpretation or semi-automated software to complete multi-worm behavior analysis, which significantly limits throughput, objectivity, and batch-to-batch consistency. This article aims to introduce the functionality and performance of our WormNet automated platform for high-throughput worm behavior analysis. Using this system, researchers can automatically complete processes such as multi-worm video acquisition, worm detection, instance segmentation, trajectory tracking, and quantification of behavioral and morphological parameters in real-world experimental scenarios. We provide a user-friendly graphical interface. Researchers only need to culture worms according to existing procedures and place samples on the platform. By selecting the field of view and analysis parameters through the interface, the system can automatically output structured data such as body length, body width, speed, and bending angle at the single-worm level for subsequent statistical analysis and dose-response assessment. This platform boasts advantages such as high automation, high analytical throughput, and is expected to liberate researchers from tedious frame-by-frame annotation and manual counting, providing a stable and reliable behavioral quantification tool for neurological function research, toxicological evaluation, and drug screening.**

***Index Terms*—Automated high-throughput analysis, worm behavior analysis, deep learning.**

I. Introduction

CAENORHABDITIS elegans, a free-living small worm, is one of the most important model organisms in neuroscience, genetics, aging, and environmental toxicology [1], [2], [3]. Its limited number of neurons but well-defined connectome, short lifespan, and ease of genetic manipulation make worm behavior a key observational window connecting genes, neural circuits, and overall phenotypes [4], [5], [6], [7]. The crawling and swimming behaviors of *C. elegans* contain rich dynamic characteristics, including speed and bending angle [8]. These quantitative indicators not only reflect neural regulation and muscle function but are also highly sensitive to drug effects, environmental toxin exposure, and lifespan decline [9], [10], [11]. Therefore, in high-throughput drug screening and mechanism-of-action studies, the ability to stably and accurately track and quantify *C. elegans* behavior on large sample sizes and long time scales has significant value for basic research and translational application.

In practical experiments, the analysis of worm behavior still largely relies on manual observation and labeling [12]. For example, in experiments involving chemotaxis, lifespan, and drug or toxin exposure, researchers often need to manually determine under a microscope whether each worm has reached a specific area, record its movement trajectory, count the number of individuals, or classify it according to posture or state. This type of manual analysis is not only time-consuming and labor-intensive, but also easily affected by subjective experience, operator fatigue, and differences between different experimental personnel, leading to limited repeatability and comparability of results [13], [14]. For drug screening experiments involving multiple concentrations, time points, and replicates, manual interpretation often becomes a bottleneck

This work was supported by the General Program of the National Natural Science Foundation of China under Grant No.62573363; in part by the Suzhou Proof-of-Concept Centre for Embodied AI in Future Education Programme.

(Corresponding authors: Eng Gee Lim and Pengfei Song.)

Qingyu Shi, Yining Shi, Eng Gee Lim, and Pengfei Song are with the School of Advanced Technology, Xi'an Jiaotong-Liverpool University, Suzhou, China, and also with the Department of Electrical Engineering and Electronics, University of Liverpool, Liverpool, U.K. (e-mail: Enggee.Lim@xjtlu.edu.cn; Pengfei.Song@xjtlu.edu.cn).

Shiyan Li is also with the School of Advanced Technology, Xi'an Jiaotong-Liverpool University, Suzhou, China.

Lei Fu, Jinxin Gu, and Yixue Qiao are with the Academy of Pharmacy, Xi'an Jiaotong-Liverpool University, Suzhou, China.

Ian Sandall is with the Department of Electrical Engineering and Electronics, University of Liverpool, Liverpool, U.K.

limiting experimental throughput and expanding experimental design space.

To reduce manual workload, early work largely relied on traditional image processing and machine learning to build worm tracking and segmentation software. Typical examples include Tierpsy Multi-Worm Tracker and Multi Animal Tracker, which primarily depend on background subtraction, thresholding, connected component analysis, and morphological operations [15], [16], [17], [18], [19], [20], [21]. These methods perform reasonably well when the background is uniform, the contrast is good, and the worm density is low. However, in real-world scenarios such as illumination drift, uneven fungal mats, and complex agar textures, frequent manual parameter tuning is often required to maintain usable performance, and their robustness and scalability across imaging conditions are limited [22], [23], [24]. In recent years, deep learning (DL), especially convolutional neural networks (CNNs), cross-scale feature networks and attention mechanism have become the mainstream tool for complex behavioral image analysis [25], [26], [27]. Faster R-CNN and its variants can robustly identify worms and their eggs in heterogeneous environments [28]. Mask R-CNN [22], the YOLO series [29], and WormYOLO [30], with their large kernel convolutions, feature pyramids, and attention mechanisms, significantly improve detection and segmentation accuracy compared to earlier Transformer methods in environments with adhesion, self-overlap, and strong noise [31]. Methods such as WormPose have made significant progress in complex pose reconstruction by training CNNs with synthetic images to predict skeletons and centerlines [17]. Furthermore, recent studies have expanded deep learning for worm analysis from basic detection and segmentation to multi-worm tracking, behavioral quantification, and improved robustness across imaging conditions. For example, Liu et al. combined an improved YOLOv8 with ByteTrack to enhance the accuracy and continuity of multi-worm tracking [32]. Escobar-Benavides et al. developed a generalist detection model to improve cross-dataset performance under diverse acquisition conditions [33]. Although these advances demonstrate the increasing capability of deep learning for automated worm analysis, most existing methods remain limited to particular tasks or platforms, and still lack a unified solution for robust multi-worm detection, segmentation, tracking, and phenotypic quantification in complex high-density imaging scenarios.

Therefore, based on microscopic imaging and deep learning, we constructed WormNet, a high-throughput worm analysis platform, as shown in Fig. 1. The main contribution of WormNet lies in its unified framework for *C. elegans* analysis, integrating detection, segmentation, tracking, and behavioral quantification into a single pipeline. Combined with worm-oriented modules for dense and ambiguous scenes, it provides a more robust solution for high-throughput multi-worm analysis. The platform integrates a temperature controlled stage, a microscopic imaging system, an industrial camera, and a GPU workstation for stable multi-worm video acquisition. The proposed algorithm is built on a multi-scale feature backbone, with simplified snake convolution (SimSnake Conv) for detection and multi-scale cross-modal gated fusion (MCGF)

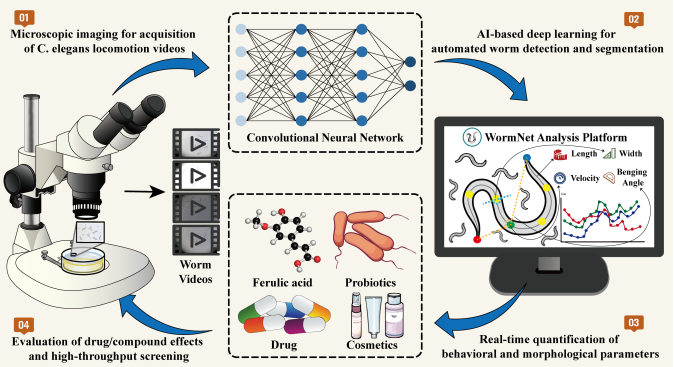

Fig. 1. Schematic of the robust and high-throughput *C. elegans* behavior analysis platform, integrating microscope-based video acquisition, deep learning based worm detection, tracking and segmentation, and automated morphology and motility quantification.

for fusing detection priors with segment anything model 2 (SAM2) encoder features to improve instance segmentation. On two representative real-world datasets, WormDrop-Micro and WormDish-Real, WormNet outperforms state-of-the-art (SOTA) methods by up to 1.90% in the detection metric AP_{50} (average accuracy calculated at IoU = 0.50), and in the more stringent $mAP_{50–95}$ (average detection accuracy across multiple thresholds at IoU range 0.50–0.95), WormNet outperforms by up to 28.50%. In the segmentation task, Dice coefficients (calculating pixel-level overlap between predicted and actual regions) and mIoU (evaluating segmentation performance by the average overlap between predicted and actual regions) are improved by up to 1.71% and 3.00% respectively. WormNet achieves a processing speed of 20.45 FPS and supports batch analysis of multi-worm videos containing up to approximately 30 worms per field of view. These results demonstrate that WormNet can consistently output stable and reliable instance-level detection and segmentation results in low-contrast, complex backgrounds, and densely distributed worm scenes. It also exhibits excellent robustness and scalability in high-throughput, real-time analysis environments, providing an engineering-feasible automated technology path for worm-based drug screening, toxicology research, and behavioral phenotypic analysis.

II. System Setup and Data Acquisition

This section outlines the hardware configuration of the automated worm analysis instrument (Sec. II-A) and the setting of the datasets used to train and evaluate WormNet (Sec. II-B). The system prototype, AI analysis pipeline, and validation workflow are summarized in Fig. 2.

A. System Setup

This study employed a self-developed intelligent observation and analysis instrument to collect and analyze worm behavior, as shown in Fig. 2a. The platform integrates optical system, temperature control system, a high-definition camera (Basler a2A4096, 1200 × 874 resolution), and a GPU host (Geforce RTX 3060). The microscope is equipped with 4×, 10×, and 20× objectives, combined with a 0.35× C-mount

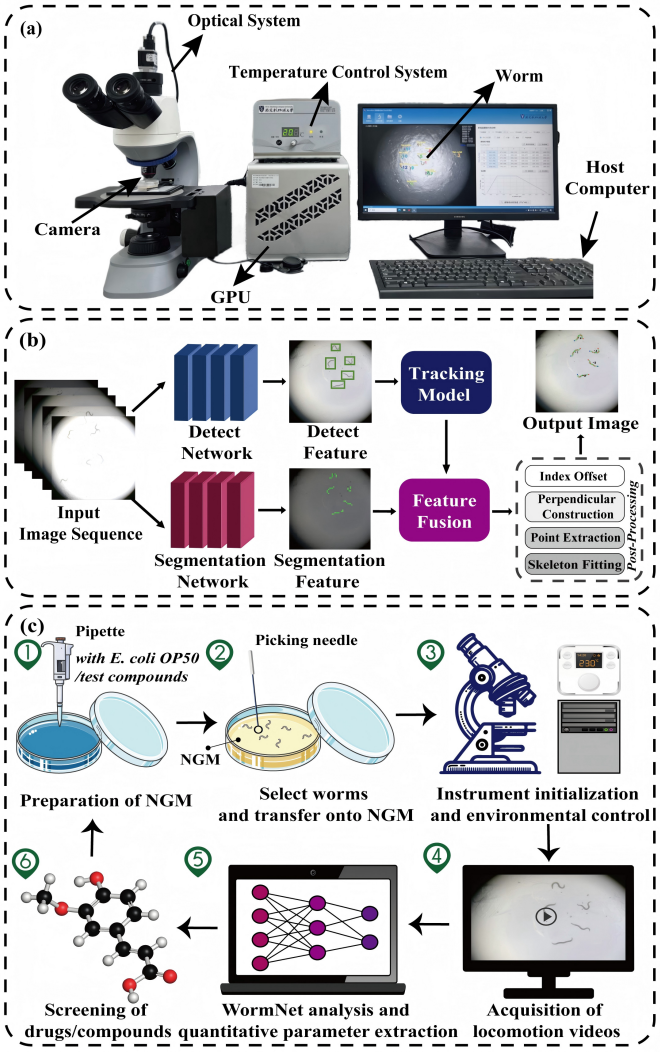

Fig. 2. (a) Prototype of the automated worm analysis system. (b) Principle diagram of the AI-based analysis pipeline for worms. (c)Workflow for *C. elegans* behavior analysis.

interface and an industrial camera to support accurate imaging at different magnifications. The heated stage provides a controllable temperature range from room temperature to 50°C, covering both routine 20°C behavioral assays and temperature-stimulation experiments. A motorized stage enables precise X–Y movement, allowing the operator to bring regions of interest to the center of the field of view via microstepping.

The overall analysis workflow of WormNet is shown in Fig. 2b. The microscopic image is first fed into the detection branch to extract multi-scale features and generate worm candidate boxes, which are subsequently associated by the tracking model, while the same image is simultaneously processed by the instance segmentation branch to produce segmentation features and mask predictions for each candidate. A feature fusion module then jointly optimizes tracking-aware detection and segmentation features, outputting instance-level worm boxes and refined contour masks. Finally, skeleton master-path fitting, key skeleton point extraction, normal-based geometric construction, and index offset are applied to each worm, providing high-quality inputs for subsequent behavioral and motility parameter computation.

The overall workflow of the worm behavior experiment is shown in Fig. 2c and consists of six steps: preparing NGM plates with E. coli OP50 and drug treatments, picking and incubating healthy worms at 20°C, initializing the instrument and setting environmental parameters, continuously acquiring worm movement videos, using WormNet for behavior analysis and parameter quantification, and finally performing statistical analysis to evaluate drug effects.

B. Data Acquisition and Labeling

To evaluate the robustness and generalization of WormNet across different experimental and imaging conditions, we constructed two representative *C. elegans* microscopy datasets: WormDrop-Micro (droplet-based imaging) and WormDish-Real (real petri-dish / plate scenes). All images were acquired using the same microscopic system with different illumination and uniformly resized to 640 × 640 pixels.

For detection, WormDrop-Micro and WormDish-Real contain 1208 and 1613 images, under 3 and 5 illumination conditions, with splits of 1057, 101, 50 and 1399, 143, 71 for training, validation, and testing respectively. For segmentation, they include 262 and 260 annotated images, with train, validation and test splits of 228, 24, 10 and 226, 24, 10 respectively. All annotations were completed in Roboflow and manually verified for consistency.

III. Method

This section presents the key deep learning based methods for automated worm behavior analysis, and details the proposed WormNet framework in Sec. III-A.

A. Overall Deep Learning Framework

As shown in Fig. 3, WormNet is a multi-task parallel network for worm behavior and morphology analysis, consisting of three sub-modules: detection, tracking, and segmentation. The detection branch employs the poly kernel inception network (PKINet) backbone combines multi-scale representation capabilities with global context modeling capabilities [34], with bi-directional feature pyramid network (BiFPN) [35] to fully utilize worm features at different scales. Simplified snake convolution (Sec. III-B) was also proposed to more effectively characterize slender features and improve the detection accuracy of small-scale worms. The tracking branch uses StrongSORT [36] to maintain consistent trajectories across frames [37]. The segmentation branch is built on SAM2 [38] basic architecture and incorporates a multi-scale cross-modal gated fusion module (Sec. III-C) that leverages detection priors to refine contours and produce high-quality instance masks. This design allows segmentation to directly use localization clues, overcoming the limitation of serial pipelines. As a result, WormNet improves boundary localization and reduces missed or blurred masks in dense and ambiguous *C. elegans* scenes. The detection and segmentation loss functions are detailed in Sec. III-D1 and Sec. III-D2, respectively.

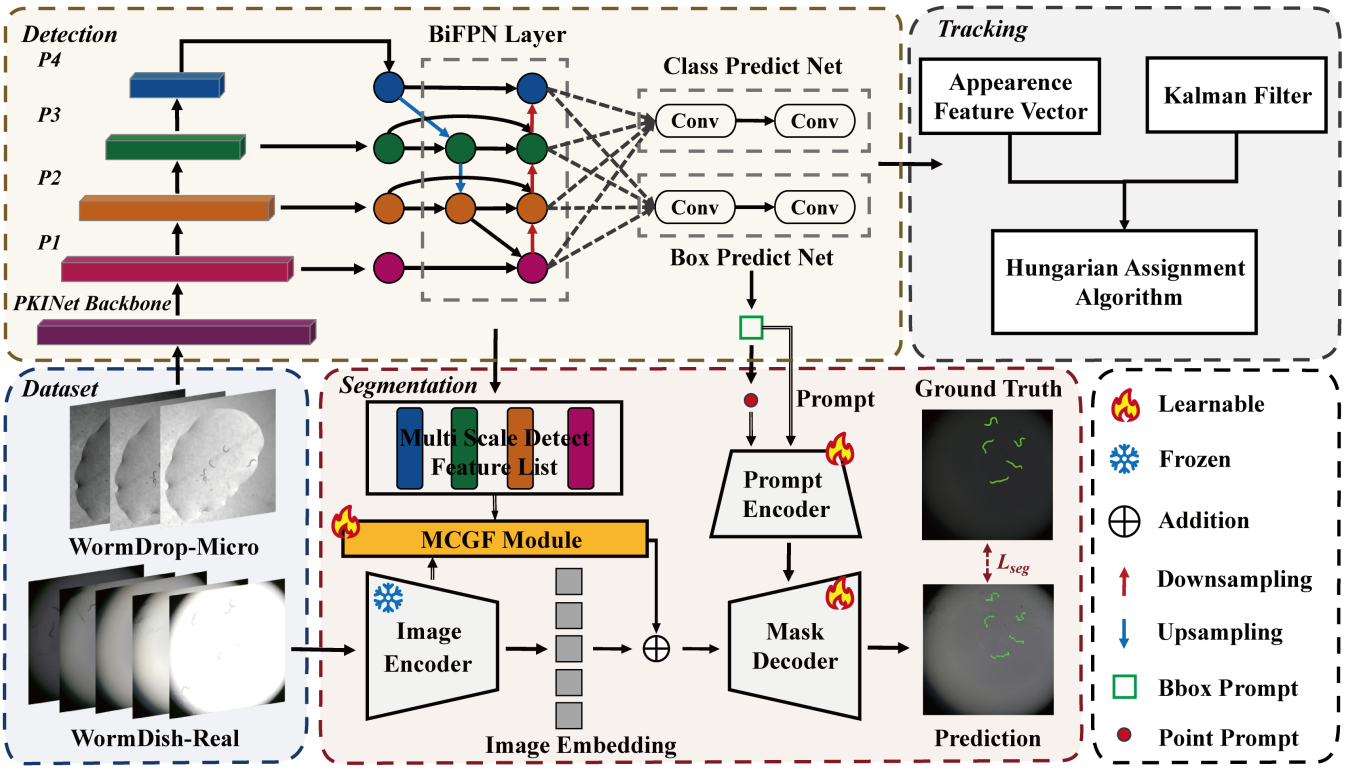

Fig. 3. Overall architecture of the proposed WormNet, where a multi-scale backbone with BiFPN detect worm classes and bounding boxes, and the segmentation branch uses these priors and multi-scale features to refine contours and generate high-quality instance masks.

B. Simplified Snake Convolution Module

To more effectively model the morphological features of slender, curved targets (such as worms), WormNet introduces a simplified snake convolution (SimSnake Conv) module in its detection branch as shown in Fig. 4. Traditional 3×3 convolutions tend to favor localized blocky textures and are insufficiently responsive to continuous structures distributed along the body long axis, high-curvature regions, and slender tails, easily resulting in contour breaks and length deviations in low-contrast or localized adhesion. SimSnake Conv introduces deformable snake branches along the horizontal and vertical directions in addition to the standard convolution branches, allowing sampling points to adaptively shift along the potential skeleton direction. This explicitly enhances the sensitivity to long-axis continuity and curved structures without significantly increasing the number of parameters, providing a more stable feature representation for subsequent instance segmentation and skeleton extraction.

Given an input feature map $\mathbf{F}_{in} \in \mathbb{R}^{B\times C_{in}\times H\times W}$, this module includes three branches: a normal 2D branch, a x-type snake branch, and a y-type snake branch. The normal 2D branch uses a standard 3×3 convolutional block for feature extraction:

$$S(x, y) = \text{Conv}_{3\times 3}(\mathbf{F}_{in}) \in \mathbb{R}^{B\times C_{\text{rest}}\times H\times W} \tag{1}$$

The x-type snake branch first generates the offset field for each pixel through an offset convolutional network, and then performs snake-shaped sampling along the horizontal direction (x-direction) using K learnable sampling points.

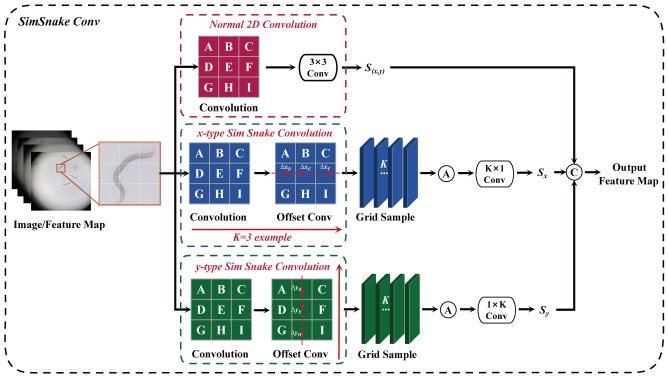

Fig. 4. Architecture of the proposed simplified snake convolution (SimSnake Conv) module, which combines a standard 3×3 convolution branch with horizontal and vertical offset-guided snake branches, to better represent slender, curved structures.

$$S_x = \sigma(\text{GN}(\text{Conv}_{K\times 1}(S_x(\mathbf{F}_{in})))) \in \mathbb{R}^{B\times C_{x-\text{snake}}\times H\times W} \tag{2}$$

where $S_x(\mathbf{F}_{in})$ represents the serpentine sampling operation based on the offset field, GN is group normalization, and $\text{Conv}_{K\times 1}$ is a $K\times 1$ convolutional kernel along the x-direction.

Similarly, the y-type snake branch performs serpentine sampling along the vertical direction:

$$S_y = \sigma(\text{GN}(\text{Conv}_{1\times K}(S_y(\mathbf{F}_{in})))) \in \mathbb{R}^{B\times C_{y-\text{snake}}\times H\times W} \tag{3}$$

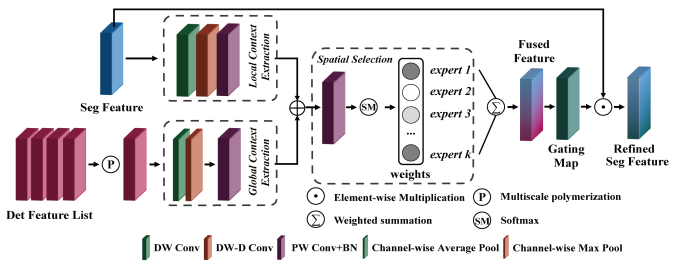

Fig. 5. Schematic of the proposed multi-scale cross-modal gated fusion module, where multi-scale detection features guide gated fusion of segmentation features to improve fine-grained segmentation in dense and occluded scenes.

Similarly, $S_y(\mathbf{F}_{in})$ represents the serpentine sampling based on the offset field in the y-direction, and $\text{Conv}_{1\times K}$ is a $1 \times K$ convolutional kernel.

The offset field is predicted by 3×3 convolution and decomposed into horizontal displacement Δx and vertical displacement Δy, constrained to $[-1, 1]$ by $\tanh$, and used for interpolation to obtain $S_x(\mathbf{F}_{in})$ and $S_y(\mathbf{F}_{in})$. Finally, the three branches are cascaded and fused along the channel dimension to obtain the output features:

$$\mathbf{F}_{out} = \text{Concat}(S(x, y), S_x, S_y) \tag{4}$$

C. Multi-Scale Cross-Modal Gated Fusion Module

The multi-scale cross-modal gated fusion (MCGF) module based on the SAM2 image encoder is shown in Fig. 5. This module aims to enable WormNet not only to perform single-modal image segmentation, but also to collaboratively utilize multi-scale prior information (target location, scale, and density information) obtained from the detection branch. After aligning the multi-scale features of the detection branch with the features encoded by SAM2 into the same space, adaptive weighted fusion is performed. The gating mechanism emphasizes reliable regions related to the contour and suppresses the influence of background and artifacts, thereby improving the robustness and contour accuracy of instance segmentation in high-density, occluded, and complex background scenes.

In this module, the first inputs are feature maps $X \in \mathbb{R}^{B\times C\times H\times W}$ from the SAM2 encoder (i.e., fine-grained features extracted by SAM2) and a set of feature maps $G_{\text{list}} = \{G_1, G_2, \ldots, G_m\}$ from the multi-scale detector, where each feature $G_i \in \mathbb{R}^{B\times C\times H_i\times W_i}$ represents a different scale feature. To ensure compatibility of multi-scale features, the multi-scale detected features first need to be spatially aligned, and a global guiding feature G_{agg} is generated by averaging, as follows:

$$G_{\text{agg}} = \frac{1}{m}\sum_{i=1}^{m} G_i \tag{5}$$

where m is the number of feature scales, and G_{agg} is the global guiding feature obtained by averaging.

Local features are derived from the output features of the SAM2 encoder, while global features are derived from the aggregated detection features. Local contextual information is extracted using depthwise convolution. The global guiding feature G_{agg} is processed using global average pooling and max pooling to obtain the global feature G_{global}, representing the semantic information of the global context. The local and global features are summed and batch normalized to obtain the fused feature F_{fusion}:

$$F_{\text{fusion}} = \text{BN}(X_{\text{local}} + G_{\text{global}}) \tag{6}$$

where BN represents the batch normalization operation. At this point, the fused feature F_{fusion} contains comprehensive information from both local context and global guidance.

Pixel-wise soft weights on multiple paths are predicted using a lightweight routing header. First, the fused feature F_{fusion} is processed using a 1×1 convolution to obtain the logit output for each expert. Then, a softmax operation is performed on the logit at each position to calculate the pixel-wise weight ω_k. Based on the predicted weights ω_k, the features E_k of each expert are weighted and summed to obtain the final fused feature F_{fused}, as shown in the following formula:

$$F_{\text{fused}} = \sum_{k=0}^{K-1} \omega_k E_k \tag{7}$$

where $E_0 = X$ is the SAM2 feature, and $E_1, E_2, \ldots, E_k$ are detection features at different scales. The weighted fused feature F_{fused} is obtained by soft-selecting features from different experts.

Finally, the fused feature F_{fused} is processed through a gating mechanism (consisting of a 1×1 convolution and batch normalization) to generate a gating map G. This gating map is then multiplied element-wise with the original SAM2 feature X to obtain the final output feature Y.

D. Loss Function

1) Detection Loss

The detection loss combines class loss, bounding box regression loss, and objectness loss in a weighted manner to jointly optimize recognition and localization performance.

Class Loss: Cross-entropy is used for classification.

$$L_{\text{cls}} = -\sum_i [y_i \log(y) + (1 - y_i)\log(1 - y)] \tag{8}$$

where y_i is the ground truth class label, and y is the predicted class probability.

Bounding Box Regression Loss: IoU loss is adopted to optimize box alignment.

$$L_{\text{box}} = 1 - \text{IoU}(B, B_i) \tag{9}$$

where B is the predicted bounding box, and B_i is the ground truth bounding box. IoU is defined as the ratio of the intersection of the predicted bounding box and the ground truth bounding box to their union:

$$\text{IoU} = \frac{B \cap B_i}{B \cup B_i} \tag{10}$$

Objectness Loss: Binary cross-entropy is used to estimate whether a box contains an object.

$$L_{\text{obj}} = -\sum_i [t_i \log(t) + (1 - t_i)\log(1 - t)] \tag{11}$$

where t_i is the true object confidence (1 if the object exists, 0 otherwise), and t is the predicted object confidence.

The final total loss function L_{det} is the weighted sum of the individual losses:

$$L_{\text{det}} = \lambda_{\text{cls}} L_{\text{cls}} + \lambda_{\text{box}} L_{\text{box}} + \lambda_{\text{obj}} L_{\text{obj}} \tag{12}$$

where λ_{cls}, λ_{box}, and λ_{obj} are the weight coefficients for each loss term.

2) Segmentation Loss

The segmentation loss function uses a combination of cross-entropy loss and IoU loss to optimize segmentation performance.

Cross-entropy loss: Used to calculate the difference between the ground truth mask and the predicted mask.

IoU loss: Used to optimize the overlap of segmented regions.

$$L_{\text{IoU}} = 1 - \frac{\text{Area}_{\text{intersection}}}{\text{Area}_{\text{union}}} \tag{13}$$

where $\text{Area}_{\text{intersection}}$ and $\text{Area}_{\text{union}}$ are the area of the intersection and union region between the predicted mask and the ground truth mask respectively.

The final segmentation loss function is:

$$L_{\text{seg\_all}} = L_{\text{seg}} + \lambda_{\text{IoU}} L_{\text{IoU}} \tag{14}$$

where L_{seg} is cross-entropy loss, λ_{IoU} is the weight coefficient of the IoU loss.

IV. Experimental Results

This section evaluates the performance of WormNet. Sec. IV-A describes the implementation details, Sec. IV-B and Sec. IV-C report detection and segmentation results, Sec. IV-D is ablation study, and Sec. IV-E, Sec. IV-F present motion and morphology quantification and ferulic acid behavior analysis case.

A. Implementation Details

1) Evaluation Metrics

For worm detection, we report $\text{mAP}_{50\text{–}95}$ (IoU range 0.50–0.95) and AP_{50} (IoU = 0.50), together with inference FPS, GFLOPs, and parameters to jointly assess accuracy and efficiency. For worm segmentation, Dice and mIoU are used to measure mask quality. All results are averaged over three independent training or testing runs.

2) Other Configurations

Detection task: Based on the Ultralytics framework with input size 640×640, batch size 16, use SGD optimizer, and cosine-annealed learning rate from 0.01 to 1×10^{-4} over 200 epochs. During tracking inference, a pre-trained best multi-scale detector is used, and the confidence and NMS IoU thresholds both set to 0.5.

Segmentation task: Randomly sampled foreground points and detection bounding boxes are used as prompts. Feature extraction uses the best detection weights, while only the fusion module, SAM2 prompt encoder, and mask decoder are updated. Training adopts AdamW with initial learning rate 1×10^{-4} and weight decay 4×10^{-5}, using a combination of binary cross-entropy and IoU-based regression loss, with up to 100,000 iterations.

All experiments run on a workstation with a GeForce RTX 3090 GPU, and the code was implemented using PyTorch, MMdetection [39], and MMsegmentation [40].

B. Detection Results and Analysis

To evaluate the performance of WormNet in worm detection, we compared it with five mainstream detection frameworks, namely TridentNet [41], FCOS [42](with focal loss [43]), Faster R-CNN [44], Cascade R-CNN [45] and YOLOX [46]. The results of their detection accuracy and computational performance are summarized in Tab. I.

On the WormDrop-Micro dataset, WormNet achieves excellent performance with AP_{50} and $\text{mAP}_{50\text{–}95}$ at 99.50% and 95.40%, respectively. Compared to Faster R-CNN (97.60%) and YOLOX (97.40%), it improves AP_{50} by approximately 1.90% and 2.10%, compared to Cascade R-CNN (81.00%), which has the best $\text{mAP}_{50\text{–}95}$ performance, it improves by 14.40%, and its advantage over other methods is up to 30%. The results demonstrate that WormNet not only maintains near perfect detection accuracy but also maintains high accuracy within a strict IoU threshold range of 0.50–0.95, exhibiting significant robustness.

On the WormDish-Real dataset with more complex experimental conditions (various lighting, bacterial patches, and petri dish textures), WormNet still achieved the best detection performance, AP_{50} of 98.80% and $\text{mAP}_{50\text{–}95}$ of 96.80%. Compared to FCOS (94.50% / 88.30%) and Cascade R-CNN (96.00% / 88.10%), AP_{50} improved by 4.30% and 2.80%, and $\text{mAP}_{50\text{–}95}$ improved up by approximately 8.70%. In contrast, while YOLOX had higher recall on AP_{50}, its $\text{mAP}_{50\text{–}95}$ on WormDrop-Micro was only 64.70%, revealing insufficient localization at high IoU thresholds, while WormNet maintained high levels on both AP_{50} and $\text{mAP}_{50\text{–}95}$, making it more suitable as a detection front-end for subsequent fine-grained behavioral and morphological analysis.

In terms of computational performance, WormNet achieves an inference speed of 20.45 FPS, higher than TridentNet (10.43 FPS), YOLOX (9.68 FPS), Faster R-CNN (17.99 FPS), and FCOS (18.12 FPS). This processing speed enables batch analysis of multi-worm videos containing up to approximately 30 worms per field of view, supporting high-throughput behavioral quantification. While its computational cost is slightly higher (173.50 GFLOPs, compared to FCOS 123 GFLOPs and Faster R-CNN 134 GFLOPs), WormNet achieves significantly higher AP_{50} and $\text{mAP}_{50\text{–}95}$ with a parameter size of 31.90M, and is far lower than Cascade R-CNN 69.20M, achieving a good trade-off between speed, accuracy, and model complexity. This allows WormNet to meet the real-time requirements of high-throughput worm behavior experiments while providing a reliable foundation for subsequent instance segmentation, skeleton extraction, and behavior parameter quantification.

C. Segmentation Results and Analysis

1) Segmentation Performance Evaluation

TABLE I
COMPARISON OF DETECTION METRICS.

Method	FPS / GFLOPs	Parameters/M	WormDrop-Micro		WormDish-Real	
			AP_{50} / %	$mAP_{50–95}$ / %	AP_{50} / %	$mAP_{50–95}$ / %
TridentNet	10.43 / 797	33.04	91.60	71.00	94.90	79.10
FCOS	18.12 / 123	32.11	92.60	78.90	94.50	88.30
Cascade R-CNN	15.27 / 162	69.20	93.70	81.00	96.00	88.10
YOLOX	9.68 / 26.76	8.94	97.40	64.70	95.70	73.60
Faster R-CNN	17.99 / 134	41.35	97.60	79.40	97.40	68.30
Ours	**20.45 / 173.50**	**31.90**	**99.50**	**95.40**	**98.80**	**96.80**

TABLE II
COMPARISON OF SEGMENTATION METRICS.

Method	WormDrop-Micro		WormDish-Real	
	Dice / %	mIoU / %	Dice / %	mIoU / %
OCRNet	66.43	49.73	66.26	49.54
DeepLabV3	88.07	78.69	90.33	82.37
DDRNet	88.09	78.71	78.43	64.51
SegFormer	91.59	84.49	92.79	86.55
Unet	92.06	85.28	76.41	61.82
Ours	**93.77**	**88.28**	**93.46**	**87.90**

To verify the segmentation performance of WormNet, we compared it with five mainstream instance segmentation networks OCRNet [47], DeepLabV3 [48], UNet [49], DDRNet [50], and SegFormer [51] on WormDrop-Micro and WormDish-Real. The quantitative results of its Dice coefficient and mIoU are shown in Tab. II.

Based on the WormDrop-Micro dataset, WormNet achieved the best performance in both Dice coefficient and mIoU (93.77% and 88.28%, respectively), improving upon the second-best method, UNet (92.06% and 85.28%), by approximately 1.71% and 3.00% respectively. Compared to other methods, it achieved improvements up to 27.34% in Dice and 38.55% in mIoU. On the more challenging WormDish-Real dataset, WormNet also achieved the best Dice and mIoU (93.46% and 87.90%, respectively), representing improvements of 0.67% and 1.35% compared to the closest performer, SegFormer (92.79% and 86.55%). Improvements over UNet reached 17.05% and 26.08% respectively. Notably, some networks (such as UNet and DDRNet) experienced significant performance degradation when migrating from WormDrop-Micro to WormDish-Real (Dice coefficients decreased by approximately 15.65% and 9.66% respectively), while WormNet maintained stable, near upper level segmentation accuracy under both imaging conditions, demonstrating excellent cross-scene generalization ability.

Overall, WormNet significantly outperforms baseline methods on both WormDrop-Micro and WormDish-Real, improving pixel-level overlap while maintaining single unit separation and contour continuity in complex backgrounds, low-contrast scenarios, and multi-scale coexistence scenarios. This verifies the effectiveness of multi-scale feature fusion and the integrated detection segmentation design in improving the accuracy and robustness of worm instance-level segmentation, providing a reliable prerequisite for subsequent motility and morphological parameter quantification.

2) Visualization and Error Analysis

In addition to the quantitative results, we also compared the segmentation results of top ranked comparison methods (DDRNet, DeepLabV3, SegFormer). These three comparison networks generally suffer from problems such as contour breaks, individual worm adhesion, and missing slender tails in both types of scenes. These typical errors reflect their insufficient ability to characterize the boundaries of slender targets under complex backgrounds, low contrast, and multiple worm adhesion conditions.

In contrast, the overall shape of the instance masks generated by WormNet is closer to that of human annotations, maintains a coherent overall contour and slender tail, and clearly distinguishes adjacent individuals even when in contact or partially overlapping, significantly reducing breakage and adhesion. These results are consistent with improvements in Dice and mIoU, indicating that WormNet achieves higher segmentation quality in terms of instance-level contour integrity and individual separability, providing a more reliable mask foundation for subsequent skeleton extraction, bending angle statistics, and behavioral feature calculation.

D. Ablation Study

We performed ablation experiments on the detection and segmentation parts of WormNet to quantify the contribution of each module. As shown in Tab. III, adding SimSnake Conv increases AP_{50} / $mAP_{50–95}$ up to 0.50% / 27.10%, and further introducing BiFPN raises them to 0.90% / 29.80%. For segmentation, the MCGF module increases the Dice coefficient to 2.68% and mIoU to 4.65% by fusing detection priors. These gains mainly come from the integrated framework, SimSnake Conv, and detection segmentation interaction, which together improve performance and robustness in complex high-throughput worm analysis.

E. Worm Motion and Morphological Parameter Analysis

Kinematically, WormNet calculates the centroid coordinates of each frame using instance masks, and obtains instantaneous speed by combining the Euclidean distance between the centroids of adjacent frames with the FPS. Simultaneously, it extracts the central skeleton from the refined mask, several representative points in the middle of the skeleton and combined with the head and tail endpoints to form bending angles to quantify posture changes and sway amplitude.

TABLE III
ABLATION STUDY ON WORMDROP-MICRO.

Task	Module			Metrics	
	Basic	SimSnake Conv	BiFPN	AP_{50}/%	mAP_{50-95}/%
Det part	✓	–	–	98.60	65.60
Det part	✓	✓	–	99.10	92.70
Det part	✓	✓	✓	**99.50**	**95.40**
	Basic	MCGF		Dice	mIoU
Seg part	✓	–		91.09	83.63
Seg part	✓	✓		**93.77**	**88.28**

Morphologically, WormNet estimates body length and width using skeleton and contour information. Body length is obtained by accumulating the Euclidean distances between adjacent points on the skeleton, body width is measured by searching for intersections with the contour along the local normal direction of the skeleton, and the average of the entire skeleton yields a representative body width. Based on this process, the system can stably output multi-dimensional indicators such as speed, bending angle, body length, and body width under parallel multi-worm views. Tab. IV shows an example of automated analysis of five worms under the same experimental conditions. Speed and bending angle describe complementary aspects of worm behavior, reflecting locomotor activity, body-wave propagation, and posture coordination. Body length and width further characterize morphology related to growth and treatment-induced phenotypic changes. Together, these parameters provide a quantitative basis for worm-based high-throughput drug screening and toxicity assessment.

Furthermore, WormNet processes frames containing about five worms in approximately 1 second per frame and remains suitable for real-time analysis under routine conditions, although runtime increases in denser scenes. In a challenging five-worm video with frequent overlap and entanglement, WormNet showed a frame-level failure rate of 24.3% (243 / 1000), defined as the proportion of frames with missed detection, mask loss or breakage. These failures were partially mitigated by temporal continuity and adjacent-frame information.

F. Ferulic Acid Behavioral Case Study

This experiment used ferulic acid (FA) to verify whether a WormNet based automated behavior quantification system could sensitively capture subtle differences in drug effects over time and dosage in real-world multi-worm experimental scenarios, rather than simply serving as a pharmacological experiment. Therefore, we set five FA concentrations (0, 100, 300, 500, and 700 μM) and collected videos on days 3 (D3), days 5 (D5), and days 7 (D7) after treatment. Approximately 12 worms were collected at each time point in each group. All videos were automatically processed by WormNet, and batch calculation of parameters such as speed and bending angle to evaluate the system's stability and validity under high-throughput, multi-worm parallel conditions.

As shown in Fig. 6a, the results of two-way ANOVA was performed on the mean crawling speed showed that the main effect of time explained 37.85% of the total variation

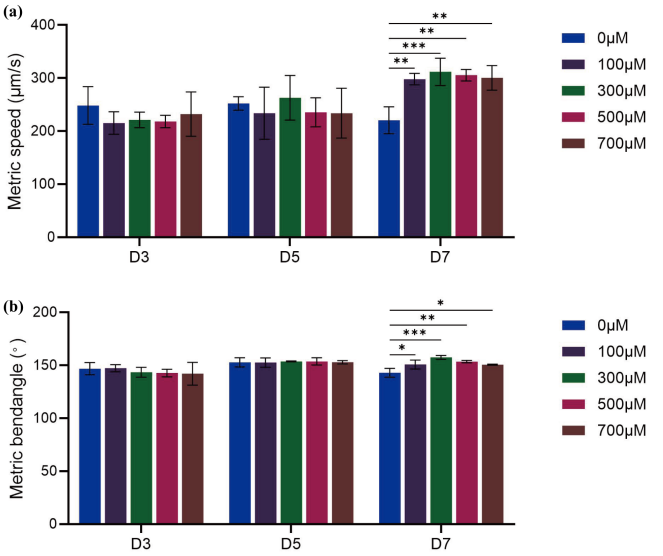

Fig. 6. Effects of different ferulic acid concentrations and exposure times on worm locomotion. (a) Two-way ANOVA results of the average speed of worms after exposure to different concentrations of ferulic acid for different days. (b) Two-way ANOVA results of the bending angle of worms under the same experimental conditions.

($P < 0.0001$), while the main effect of concentration alone accounted for only 3.89% ($P = 0.506$). The time $\times$ concentration interaction term was significant and contributed 23.91% of the variation ($P = 0.0269$). Therefore, the speed index automatically extracted by WormNet clearly reveals that the effect of FA is not a simple higher dose, faster or slower speed, but rather a time-dependent dose-response that evolves with exposure time. The comparative results showed that at D3 and D5, there were no statistically significant differences in speed between all FA groups and the control group (0 μM), suggesting that motility had not undergone systematic changes under short term exposure. However, at D7, the mean speed of all dose groups from 100 to 700 μM was significantly higher than that of the control group, with an increase of approximately 80–90 μm/s. The control group showed a slight age-dependent deceleration from D3, D5 to D7, while the continuous FA treatment showed deceleration suppressed or even transformed into increased activity.

The two-way ANOVA of the bending angle showed a trend highly consistent with that of speed (Fig. 6b). At days 3 and 5, the mean bending angle of each dose group remained relatively stable within the range of 145°–155°, with minimal differences between groups, indicating that short-term FA treatment had a limited impact on the amplitude of body swaying. By days 7, the bending angle of the control group decreased to approximately 140°, showing a similar age related decline as speed, while the bending angle of the 100–700 μM FA group increased to approximately 150°–160°, significantly higher than the control group, suggesting that long term FA exposure can maintain or even enhance the activity of worm body swaying. The speed and bending behaviors exhibited a consistent late activation pattern on the time-concentration plane, indicating that the multi-index behavioral spectrum output by

TABLE IV
AUTOMATED PARAMETER ANALYSIS RESULTS OF WORM SAMPLES.

Worm No.	Morphology		Motility	
	Length (mm)	Width (mm)	Speed (μm/s)	Average Bending Angle (°)
1	0.9127	0.0673	176.89	149.5753
2	0.9347	0.0705	91.22	129.6920
3	0.8845	0.0579	276.65	117.2135
4	0.9101	0.0698	95.53	140.9074
5	0.9589	0.0710	109.51	119.9633

TABLE V
CROSS-DATASET GENERALIZATION EVALUATION OF WORMNET.

Dataset	Detection Metrics		Segmentation Metrics	
	AP_{50}/%	$mAP_{50–95}$/%	Dice	mIoU
WormDrop-Micro ↓ WormDish-Real	74.80	57.50	78.32	66.24
WormDish-Real ↓ WormDrop-Micro	93.70	76.50	86.21	75.79

WormNet has good internal consistency and explanatory power in capturing subtle drug-induced kinetic changes, providing a reliable automated evaluation method for subsequent worm-based efficacy assessment and toxicity screening.

G. Cross-dataset Generalization Evaluation

Cross-dataset validation on two imaging setups demonstrated good generalization of WormNet. Training on WormDrop-Micro and testing on WormDish-Real achieved 74.80% AP_{50}, 57.50% $mAP_{50–95}$, 78.32% Dice, and 66.24% mIoU, while the reverse setting achieved 93.70% AP_{50}, 76.50% $mAP_{50–95}$, 86.21% Dice, and 75.79% mIoU shown in Tab. V, indicating that training on more complex scenes improves robustness.

V. DISCUSSION

The WormNet platform developed in this study improves existing worm behavior analysis methods at three levels: structure, algorithm, and application. Through integrated hardware and software design, it integrates microscopic imaging, detection, segmentation, tracking, and behavior parameter quantification into a unified deep learning closed-loop process, avoiding information loss caused by the fragmentation of traditional detection, segmentation, and tracking processes. At the algorithm level, modules such as SimSnake Conv and MCGF are introduced, which significantly improves robustness and accuracy in low-contrast, complex background, and high-density worm scenarios. Through experiments on two representative datasets, WormDrop-Micro and WormDish-Real, the system achieves maximum to 1.90% improvement in AP_{50} and 28.50% improvement in $mAP_{50–95}$ over SOTA detectors, and in terms of segmentation accuracy, the Dice coefficient was improved by up to 1.71%, and mIoU by 3.00%, indicating that WormNet shows excellent robustness under common variations in illumination and background interference. In addition, the WormNet pipeline achieved a processing speed of 20.45 FPS and supports batch analysis of multi-worm videos containing up to approximately 30 worms per field of view, further demonstrating its suitability for high-throughput behavioral quantification. Multi-dose, multi-time point experiments using ferulic acid (FA) as a model drug further demonstrated that the speed and bending angle automatically extracted by WormNet can clearly characterize the time-dose interaction effect. Nevertheless, under high-density overlap and entanglement, WormNet shows reduced performance, with a frame-level failure rate of 24.3% in a representative test video. These cases remain important directions for future improvement. Overall, WormNet effectively addresses the shortcomings of existing methods in instance-level profiling, skeleton acquisition, robustness to complex scenarios, and integrated evaluation processes, providing an engineerable quantitative tool for high-throughput drug screening, toxicology studies, and behavioral phenotypic analysis of other model organisms.

VI. CONCLUSION

This work constructs WormNet, an integrated platform for high-throughput behavioral analysis of *C. elegans*, achieving microscopic imaging, detection, segmentation, tracking, and quantification of behavioral parameters within a unified framework. Thanks to multi-scale feature extraction and cross-modal fusion design, WormNet achieves improvements in AP_{50} of up to 1.90%, $mAP_{50–95}$ of up to 28.50%, and Dice, mIoU of up to approximately 1.71%, and 3.00% on WormDrop-Micro and WormDish-Real compared to existing best methods, demonstrating excellent robustness under complex imaging conditions. Beyond accuracy, the WormNet workflow reaches a processing speed of 20.45 FPS and allows batched analysis of multi-worm recordings with approximately 30 worms in one field of view, further supporting high-throughput behavioral assessment. Combined with ferulic acid treatment experiments, WormNet automatically extracts indicators such as speed and bending angle, which can distinguish time-dose-dependent behavioral differences, demonstrating its ability to quantitatively characterize the effects of real drugs or compounds. The main limitation of the current platform is reduced reliability in severely overlapped, occluded, self-coiled, or defocused cases, which remain important directions for future improvement. Overall, WormNet provides an engineering path for high-throughput, automated, and data-driven analysis of worm behavioral phenotypes and has the potential for application to other small model organisms.

REFERENCES

[1] J. M. Gray, J. J. Hill, and C. I. Bargmann, "A circuit for navigation in caenorhabditis elegans," *Proceedings of the National Academy of Sciences*, vol. 102, no. 9, pp. 3184–3191, 2005.

[2] E. M. Hedgecock and R. L. Russell, "Normal and mutant thermotaxis in the nematode caenorhabditis elegans." *Proceedings of the National Academy of Sciences*, vol. 72, no. 10, pp. 4061–4065, 1975.

[3] X. Dong, P. Song, and X. Liu, "Automated robotic microinjection of the nematode worm caenorhabditis elegans," *IEEE Transactions on Automation Science and Engineering*, vol. 18, no. 2, pp. 850–859, 2020.

[4] A. K. Ray, A. Priya, M. Z. Malik, T. A. Thanaraj, A. K. Singh, P. Mago, C. Ghosh, Shalimar, R. Tandon, and R. Chaturvedi, "A bioinformatics approach to elucidate conserved genes and pathways in c. elegans as an animal model for cardiovascular research," *Scientific Reports*, vol. 14, no. 1, p. 7471, 2024.

[5] M. C. Roozen and M. J. Kas, "Assessing genetic conservation of human sociability-linked genes in c. elegans," *Behavior Genetics*, vol. 55, no. 2, pp. 141–152, 2025.

[6] Y. Kwon, J. Kim, Y. B. Son, S. A. Lee, S. S. Choi, and Y. Cho, "Advanced neural functional imaging in c. elegans using lab-on-a-chip technology," *Micromachines*, vol. 15, no. 8, p. 1027, 2024.

[7] H. Yuan, W. Yuan, S. Duan, K. Jiao, Q. Zhang, E. G. Lim, M. Chen, C. Zhao, P. Pan, X. Liu *et al.*, "Microfluidic-assisted caenorhabditis elegans sorting: current status and future prospects," *Cyborg and Bionic Systems*, vol. 4, p. 0011, 2023.

[8] S. Sohrabi, D. E. Mor, R. Kaletsky, W. Keyes, and C. T. Murphy, "High-throughput behavioral screen in c. elegans reveals parkinson's disease drug candidates," *Communications Biology*, vol. 4, no. 1, p. 203, 2021.

[9] P. Song, W. Zhang, A. Sobolevski, K. Bernard, S. Hekimi, and X. Liu, "A microfluidic device for efficient chemical testing using caenorhabditis elegans," *Biomedical Microdevices*, vol. 17, no. 2, p. 38, 2015.

[10] X. Dong, P. Song, and X. Liu, "An automated microfluidic system for morphological measurement and size-based sorting of c. elegans," *IEEE Transactions on Nanobioscience*, vol. 18, no. 3, pp. 373–380, 2019.

[11] K. Kuze, U. T. Tazawa, K. Suwazono, C. Chen, Y. Toyoshima, and Y. Iino, "Wormtracer: A precise method for worm posture analysis using temporal continuity," *Journal of Neuroscience Methods*, vol. 427, p. 110644, 2025.

[12] T. B. Mahbub, P. Safaeian, and S. Sohrabi, "Automated platforms in c. elegans research: Integration of microfluidics, robotics, and artificial intelligence," *Micromachines*, vol. 16, no. 10, p. 1138, 2025.

[13] E. Fryer, S. Guha, L. E. Rogel-Hernandez, T. Logan-Garbisch, H. Farah, E. Rezaei, I. N. Mollhoff, A. L. Nekimken, A. Xu, L. S. Seyahi *et al.*, "A high-throughput behavioral screening platform for measuring chemotaxis by c. elegans," *PLoS Biology*, vol. 22, no. 6, p. e3002672, 2024.

[14] R. A. Kerr, A. E. Roux, J. Goudeau, and C. Kenyon, "The c. elegans observatory: High-throughput exploration of behavioral aging," *Frontiers in Aging*, vol. 3, p. 932656, 2022.

[15] A. Javer, M. Currie, C. W. Lee, J. Hokanson, K. Li, C. N. Martineau, E. Yemini, L. J. Grundy, C. Li, Q. Ch'ng *et al.*, "An open-source platform for analyzing and sharing worm-behavior data," *Nature Methods*, vol. 15, no. 9, pp. 645–646, 2018.

[16] N. A. Swierczek, A. C. Giles, C. H. Rankin, and R. A. Kerr, "High-throughput behavioral analysis in c. elegans," *Nature Methods*, vol. 8, no. 7, pp. 592–598, 2011.

[17] L. Hebert, T. Ahamed, A. C. Costa, L. O'shaughnessy, and G. J. Stephens, "Wormpose: Image synthesis and convolutional networks for pose estimation in c. elegans," *PLoS Computational Biology*, vol. 17, no. 4, p. e1008914, 2021.

[18] S. J. Husson, W. S. Costa, C. Schmitt, and A. Gottschalk, "Keeping track of worm trackers," *WormBook: The Online Review of C. elegans Biology*, 2018.

[19] S. Wang and Z. Wang, "Track-a-worm, an open-source system for quantitative assessment of c. elegans locomotory and bending behavior," *PloS One*, vol. 8, no. 7, p. e69653, 2013.

[20] A. Javer, L. Ripoll-Sánchez, and A. E. X. Brown, "Powerful and interpretable behavioural features for quantitative phenotyping of caenorhabditis elegans," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 373, no. 1758, p. 20170375, 2018.

[21] E. Itskovits, A. Levine, E. Cohen, and A. Zaslaver, "A multi-animal tracker for studying complex behaviors," *BMC Biology*, vol. 15, no. 1, p. 29, 2017.

[22] P. D. McClanahan, L. Golinelli, T. A. Le, and L. Temmerman, "Automated scoring of nematode nictation on a textured background," *PLoS One*, vol. 18, no. 8, p. e0289326, 2023.

[23] P. E. L. Castro, K. Kounakis, A. G. Garví, I. Gkikas, I. Tsiamantas, N. Tavernarakis, and A. J. Sánchez-Salmerón, "Segelegans: Instance segmentation using dual convolutional recurrent neural network decoder in caenorhabditis elegans microscopic images," *Computers in Biology and Medicine*, vol. 190, p. 110012, 2025.

[24] W. H. Weheliye, J. Rodriguez, L. Feriani, A. Javer, V. Uhlmann, and A. E. Brown, "A neural network model enables worm tracking in challenging conditions and increases signal-to-noise ratio in phenotypic screens," *PLOS Computational Biology*, vol. 21, no. 8, p. e1013345, 2025.

[25] A. García-Garví and A. J. Sánchez-Salmerón, "High-throughput behavioral screening in caenorhabditis elegans using machine learning for drug repurposing," *Scientific Reports*, vol. 15, no. 1, p. 26140, 2025.

[26] W. Dai, Z. Wu, R. Liu, J. Zhou, M. Wang, T. Wu, and J. Liu, "Sosegformer: A cross-scale feature correlated network for small medical object segmentation," *IEEE International Symposium on Biomedical Imaging (ISBI)*, pp. 1–4, 2024.

[27] W. Dai, R. Liu, Z. Wu, T. Wu, M. Wang, J. Zhou, Y. Yuan, and J. Liu, "Exploiting scale-variant attention for segmenting small medical objects," *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–18, 2026.

[28] K. Bates, K. N. Le, and H. Lu, "Deep learning for robust and flexible tracking in behavioral studies for c. elegans," *PLOS Computational Biology*, vol. 18, no. 4, p. e1009942, 2022.

[29] J. Zhang, S. Liu, H. Yuan, R. Yong, S. Duan, Y. Li, J. Spencer, E. G. Lim, L. Yu, and P. Song, "Deep learning for microfluidic-assisted caenorhabditis elegans multi-parameter identification using yolov7," *Micromachines*, vol. 14, no. 7, p. 1339, 2023.

[30] B. Dong and W. Chen, "A high precision method of segmenting complex postures in caenorhabditis elegans and deep phenotyping to analyze lifespan," *Scientific Reports*, vol. 15, no. 1, p. 8870, 2025.

[31] M. Deserno and K. Bozek, "Wormswin: Instance segmentation of c. elegans using vision transformer," *Scientific Reports*, vol. 13, no. 1, p. 11021, 2023.

[32] X. Liu, J. Liu, W. Teng, Y. Peng, B. Li, X. Han, and J. Huo, "Automated c. elegans behavior analysis via deep learning-based detection and tracking," *PLOS Computational Biology*, vol. 21, no. 11, p. e1013707, 2025.

[33] S. Escobar-Benavides, A. García-Garví, P. E. Layana-Castro, and A. J. Sánchez-Salmerón, "Towards generalization for caenorhabditis elegans detection," *Computational and Structural Biotechnology Journal*, vol. 21, pp. 4914–4922, 2023.

[34] X. Cai, Q. Lai, Y. Wang, W. Wang, Z. Sun, and Y. Yao, "Poly kernel inception network for remote sensing detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 27 706–27 716.

[35] M. Tan, R. Pang, and Q. V. Le, "Efficientdet: Scalable and efficient object detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 10 781–10 790.

[36] Y. Du, Z. Zhao, Y. Song, Y. Zhao, F. Su, T. Gong, and H. Meng, "Strongsort: Make deepsort great again," *IEEE Transactions on Multimedia*, vol. 25, pp. 8725–8737, 2023.

[37] S. Schneider, G. W. Taylor, and S. C. Kremer, "Similarity learning networks for animal individual re-identification-beyond the capabilities of a human observer," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision Workshops*, 2020, pp. 44–52.

[38] N. Ravi, V. Gabeur, Y. Hu, R. Hu, C. Ryali, T. Ma, H. Khedr, R. Rädle, C. Rolland, L. Gustafson, E. Mintun, J. Pan, K. V. Alwala, N. Carion, C. Wu, R. Girshick, P. Dollár, and C. Feichtenhofer, "Sam 2: Segment anything in images and videos," *arXiv:2408.00714*, 2024.

[39] K. Chen, J. Wang, J. Pang, Y. Cao, Y. Xiong, X. Li, S. Sun, W. Feng, Z. Liu, J. Xu, Z. Zhang, D. Cheng, C. Zhu, T. Cheng, Q. Zhao, B. Li, X. Lu, R. Zhu, Y. Wu, J. Dai, J. Wang, J. Shi, W. Ouyang, C. C. Loy, and D. Lin, "MMDetection: Open mmlab detection toolbox and benchmark," *arXiv:1906.07155*, 2019.

[40] M. Contributors, "MMSegmentation: Openmmlab semantic segmentation toolbox and benchmark," https://github.com/open-mmlab/mmsegmentation, 2020.

[41] Y. Li, Y. Chen, N. Wang, and Z. Zhang, "Scale-aware trident networks for object detection," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 6054–6063.

[42] Z. Tian, C. Shen, H. Chen, and T. He, "Fcos: Fully convolutional one-stage object detection," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 9627–9636.

[43] T. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2017, pp. 2980–2988.

[44] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137–1149, 2016.

[45] Z. Cai and N. Vasconcelos, "Cascade r-cnn: High quality object detection and instance segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 5, pp. 1483–1498, 2019.

[46] Z. Ge, S. Liu, F. Wang, Z. Li, and J. Sun, "Yolox: Exceeding yolo series in 2021," *arXiv:2107.08430*, 2021.

[47] Y. Yuan, X. Chen, and J. Wang, "Object-contextual representations for semantic segmentation," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2020, pp. 173–190.

[48] L. C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 801–818.

[49] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical Image Computing and Computer-assisted Intervention*, 2015, pp. 234–241.

[50] H. Pan, Y. Hong, W. Sun, and Y. Jia, "Deep dual-resolution networks for real-time and accurate semantic segmentation of traffic scenes," *IEEE Transactions on Intelligent Transportation Systems*, vol. 24, no. 3, pp. 3448–3460, 2022.

[51] E. Xie, W. Wang, Z. Yu, A. Anandkumar, J. M. Alvarez, and P. Luo, "Segformer: Simple and efficient design for semantic segmentation with transformers," *Advances in Neural Information Processing Systems*, vol. 34, pp. 12 077–12 090, 2021.

Qingyu Shi received the M.E. degree in Multimedia Telecommunications from Xi'an Jiaotong-Liverpool University, Suzhou, China, in 2025. She is currently pursuing the Ph.D. degree in Electrical Engineering and Electronics at Xi'an Jiaotong-Liverpool University, Suzhou, China. Her primary research interests include computer vision, machine learning, and their applications in intelligent image analysis.

Yining Shi received the M.S. degree in Crop Genetics and Breeding from Nanjing Agriculture University, Nanjing, China, in 2023, and is a Ph.D. candidate in Electrical Engineering and Electronics from Xi'an Jiaotong-Liverpool University, Suzhou, China. Her research interests include the biosensing, lateral flow assay (LFA) technologies and point-of-care testing.

Shiyan Li is currently pursuing the B.E. degree in Computer Science and Technology with Xi'an Jiaotong-Liverpool University, Suzhou, China. His research interests include computer vision and deep learning.

Lei Fu received the Ph.D. degree in Chemistry from Stanford University, Stanford, CA, USA, in 1997. He is currently a Professor and Executive Dean of the Academy of Pharmacy, Xi'an Jiaotong-Liverpool University, Suzhou, China, and a Professor of Medicinal Chemistry at Shanghai Jiao Tong University, Shanghai, China. His research interests include drug discovery and development, process research on drug APIs and related substances, traditional Chinese medicines as botanical cosmeceuticals, medical devices as targeted drug delivery systems, and human hibernation studies using biomimetic models of cytochrome c oxidase.

Jinxin Gu received the Ph.D. degree in Pharmaceutical Sciences from Shanghai Jiao Tong University, Shanghai, China, in 2022. He is currently an Assistant Professor at the Academy of Pharmacy, Xi'an Jiaotong-Liverpool University, Suzhou, China. His research interests include neurodegenerative diseases, mitochondrial metabolism, aging-associated diseases, and drug discovery using the *C. elegans* model.

Yixue Qiao received the Ph.D. degree in Reproductive Medicine from the University of Newcastle, Australia, in 2021. She is currently an Assistant Professor at the Academy of Pharmacy, Xi'an Jiaotong-Liverpool University, Suzhou, China. Her research interests include mitochondrial diseases, autoimmune diseases, cancer, drug discovery, and in vitro and in vivo efficacy validation.

Ian Sandall received the Ph.D. degree in Quantum Dot Lasers from Cardiff University, Cardiff, U.K., in 2008. Since 2016, he has been a Lecturer at the University of Liverpool, Liverpool, U.K. His research interests include photonic biosensors, integrated semiconductor lab-on-a-chip systems, medical diagnostics, and point-of-care applications.

Eng Gee Lim received the Ph.D. degree in Electrical and Electronic Engineering from Northumbria University, Newcastle upon Tyne, U.K., in 2002. He is currently the Inaugural Academy Dean of Artificial Intelligence, Inaugural Director of the AI University Research Centre at Xi'an Jiaotong-Liverpool University, Suzhou, China. His research interests include artificial intelligence, robotics, AI healthcare, microwave engineering, antennas, wireless communication networks, smart-grid communication, energy harvesting, and wireless networks for smart and green cities.

Pengfei Song received the Ph.D. degree in Mechanical Engineering from McGill University, Montreal, Canada, in 2018. He is currently a Senior Associate Professor of Mechatronics and Robotics at Xi'an Jiaotong-Liverpool University, Suzhou, China. His research interests include microfluidic biosensors, autonomous microscale robotics, smart healthcare instrumentation, AI-integrated paper-based microfluidic diagnostics, and automated *C. elegans* drug screening systems.