

# A CNN FRAMEWORK BASED ON LINE ANNOTATIONS FOR DETECTING NEMATODES IN MICROSCOPIC IMAGES

Long Chen,<sup>1</sup> Martin Strauch,<sup>1</sup> Matthias Daub,<sup>2</sup> Xiaochen Jiang,<sup>1</sup> Marcus Jansen,<sup>3</sup>  
Hans-Georg Luigs,<sup>3</sup> Susanne Schultz-Kuhlmann,<sup>4</sup> Stefan Krüssel,<sup>4</sup> Dorit Merhof<sup>1</sup>

<sup>1</sup>Imaging and Computer Vision, RWTH Aachen University, Aachen, Germany

<sup>2</sup>Julius Kühn Institute: Federal Research Centre for Cultivated Plants, Elsdorf, Germany

<sup>3</sup>LemnaTec GmbH, Aachen, Germany

<sup>4</sup>Landwirtschaftskammer Niedersachsen - Pflanzenschutzamt (PSA), Hannover, Germany

## ABSTRACT

Plant parasitic nematodes cause damage to crop plants on a global scale. Robust detection on image data is a prerequisite for monitoring such nematodes, as well as for many biological studies involving the nematode *C. elegans*, a common model organism. Here, we propose a framework for detecting worm-shaped objects in microscopic images that is based on convolutional neural networks (CNNs). We annotate nematodes with curved lines along the body, which is more suitable for worm-shaped objects than bounding boxes. The trained model predicts worm skeletons and body endpoints. The endpoints serve to untangle the skeletons from which segmentation masks are reconstructed by estimating the body width at each location along the skeleton. With light-weight backbone networks, we achieve 75.85% precision, 73.02% recall on a potato cyst nematode data set and 84.20% precision, 85.63% recall on a public *C. elegans* data set.

**Index Terms**— worm detection, CNN, line annotations

## 1. INTRODUCTION

### 1.1. Motivation

Many nematode (roundworm) species are parasitic on crops, such as potatoes, sugar beets and soybeans, causing billions of losses in agriculture worldwide [1]. Robust nematode detection in microscopic images is a prerequisite for quantifying nematode infestation based on (soil) samples and for phenotyping, i.e. measuring quantitative features that characterize the nematodes. Moreover, the nematode *C. elegans* is an important model organism in biology, having, for example, been used for high-throughput screening of antimicrobial drugs [2]. A lesioning study of *C. elegans* motorneurons served to infer the function of individual neurons [3]. Large-scale chemical and RNAi screens using nematodes are also widespread [4].

### 1.2. Contribution

We introduce a CNN-based approach for detecting worm-shaped objects in microscopic images (Section 2.1). The worms are long and thin, i.e. they extend over a large range, but cover only a small number of pixels. Thus, we propose to represent worms by curved lines along the body instead of bounding boxes as they are used by most object detection approaches [5]. Given well-estimated worm skeletons and endpoints, overlapping worms can be untangled (Section 2.2) and segmentation masks can be reconstructed (Section 2.3).

We evaluate our method on a potato cyst nematode (*Globodera* spp.) data set generated by Pflanzenschutzamt (PSA) Niedersachsen and a public *C. elegans* data set (BBBC010 [2]), which are representative of sparse- and dense-object data with dirty and clean background, respectively (Sections 3-4).

### 1.3. Relationship to prior work

Previous work focusses largely on untangling worms on clean backgrounds and on phenotyping. In [6, 7, 8], a graph is constructed from skeleton segments for each worm cluster. Individuals are untangled by searching for best fits of the learned worm model while minimizing overlap. Worm Tracker 2.0 [9] was developed to record the behavior of worms and can extract 702 features relevant to behavioral phenotypes.

However, the approaches mentioned above can only be applied to pure samples in which worms are relatively easy to segment (as in [6]). In contrast, microscopic images acquired from soil samples (Section 3) contain eggs, nematode cyst fragments and other organic debris. Recent advances in deep learning have greatly improved detection performance for objects in complex contexts. Standard detection approaches [5] make predictions in the form of bounding boxes, which are a good representation of approximately convex objects, but not very informative for elongated worms. Instead, we use curved lines along the worm body as a more suitable representation.

Worms in the image are likely to overlap, especially in case of a large number of worms per area. Different from the model search approaches used by [6, 7, 8], we find that indi-

\*This work was funded by the Germany Ministry of Education and Research, project KMU-innovativ-19: PheNeSens, grant number 031B0474C

viduals can be well untangled with simple geometric criteria, when the endpoints (head and tail) are known. Therefore, the network is trained to output both the skeleton and endpoints. To handle the instability of the training due to highly unbalanced positive and negative pixels, we adopt a focal loss [10] with reduced penalty around the positive pixels, which is inspired by CornerNet [11].

After skeletons of individuals are obtained, we reconstruct segmentation masks by estimating the body width at each skeleton pixel. The entire pipeline outputs object segmentations, requiring only skeleton annotations for training.

## 2. OUR APPROACH

### 2.1. Skeleton and endpoint prediction

We employ the standard U-Net architecture [12] as the CNN component of our framework. Two branches are added on the last feature map for predicting the worm *skeleton* and body *endpoints*, respectively. Each branch consists of one feature convolutional layer (ReLU activation) with 64 channels and a following output layer with 2 classes (softmax activation).

A common problem of dense prediction at each pixel with highly unbalanced positive and negative labels is that the model will degenerate to predict all pixels as the majority. To make the training gradually focus on the mispredicted minority, we apply a variant [11] of the focal loss [10]:

$$L = \frac{-1}{NHW} \sum_{i=1}^H \sum_{j=1}^W \begin{cases} (1-p)^\gamma \log p, & y_{ij} = 1 \\ (1-w_{ij})^\beta (p)^\gamma \log(1-p), & y_{ij} = 0 \end{cases}$$

where  $H$  and  $W$  are the image height and width,  $N$  is the number of objects in the image,  $p$  is the probability of being a positive label. The focusing parameter  $\gamma$  [10] is set to 2.

We compute the weight map  $w_{ij}$  by applying a 1D unnormalized Gaussian function to the distance transforms [13] of the ground truth. The distance-based weighting reduces the penalty of negative pixels around positive pixels, giving a certain degree of tolerance to the offset in both annotation and training. The standard deviation  $\sigma$  of the Gaussian determines the distance within which the penalty is reduced. During the training process, this "ground truth slack" is applied separately to the skeletons and to the endpoints.

The hyperparameter  $\beta$  is set to 4 for all experiments. The total training loss is the sum of the losses of both branches.

### 2.2. Worm untangling

Skeletons of overlapping worms can be untangled with the help of the *endpoints* (head/tail;  $ep$  for brevity) predicted by the CNN. Based only on the geometry of the skeletons, points on the skeletons ( $p\_skel$ ) can be classified as *geometric endpoints*, intersections and line points ( $ep\_geo$ ,  $ip\_geo$ ,  $lp\_geo$ ).

---

### Algorithm 1 Untangle individual skeletons

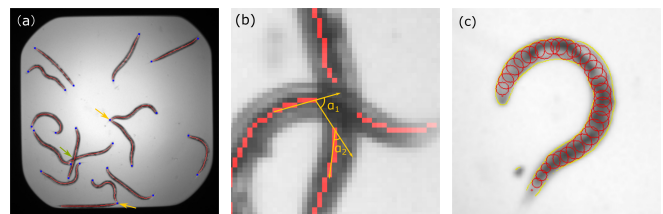
---

**Require:** skeleton  $skel$ , predicted endpoints  $ep$

- 1: find endpoints  $ep\_geo$ , intersections  $ip\_geo$  of  $skel$
  - 2: **for each**  $p \in \{ep\}$  **do**
  - 3:   **if**  $p$  matches\_one\_of  $\{ep\_geo\}$  **then**
  - 4:     continue
  - 5:   **else if**  $p$  matches\_one\_of  $\{p\_skel\}$  **then**
  - 6:     cut the skeleton
  - 7:   **end if**
  - 8: **end for**
  - 9: recompute  $ip\_geo$  of the cut  $skel$
  - 10: **for each**  $p \in \{ip\_geo\}$  **do**
  - 11:   cut the skeleton at intersection  $p$
  - 12:   get all cut segments
  - 13:   compute steering angles  $sA$  between segment pairs
  - 14:   **while** #remaining\_segments  $\geq 2$  **do**
  - 15:     connect segments with the smallest  $sA$
  - 16:   **end while**
  - 17: **end for**
- 

The skeletons of individuals are computed with Algorithm 1 that consists of two steps: 1) cutting the skeleton (line 2-7) to separate worms fused at the endpoints, and 2) resolving intersections by cutting worms and connecting the cut segments from the same worm (line 10-17): see Figure 1a. The "matches\_one\_of" operations localize the closest target inside a search area (a search circle with radius 5).

At an intersection, our heuristic decides to connect segments that form a more or less straight line: It connects the segments with the smallest steering angle  $sA$ , where  $sA$  is defined as the sum of the steering angles between the cut segments and the line that connects the cut segments (Figure 1b).



**Fig. 1. a)** Cutting the skeleton (yellow arrow) and connecting segments (green arrow). **b)** Steering angle between two segments at an intersection  $sA = \alpha_1 + \alpha_2$ . **c)** Mask reconstruction by estimating the body width.

### 2.3. Reconstructing masks from the skeletons

To reconstruct segmentation masks from the skeletons (Figure 1c), we need to estimate the width of the worm body at each skeleton point. First, edges are detected in the original image with the Canny edge detector [14]. For each skeleton

pixel, we use the shortest distance to the edge as the radius and fill a circle centered at the skeleton pixel.

In order to make the reconstruction more stable, the estimated radius is smoothed (two pixels before and after) along the skeleton. In addition, we limit the radius to be less than the path length from the skeleton pixel to the skeleton end, so that the segmentation will form a tip at the endpoint.

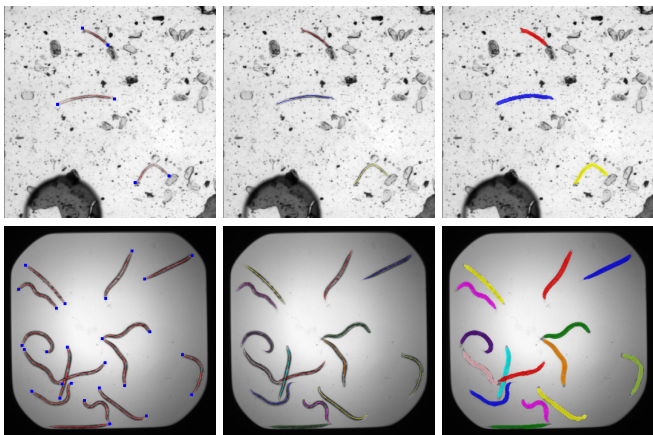
### 3. IMAGE DATA

We evaluated our method on two data sets (Figure 2): a motivating potato cyst nematode (*Globodera* spp.) data set recorded by PSA, and a public *C. elegans* reference data set (Broad Bioimage Benchmark Collection: BBBC010 [2]).

The PSA data set contains 3376 nematodes in 1973 microscopic images. Images without nematodes account for 43.6% in the data set, 26.6% contain only one, 16.8% two objects. Since cyst nematodes like *Globodera* spp. produce their offspring in a cyst, which needs to be physically crushed to release worm shaped nematode stages, the samples contain large numbers of distracting objects, such as cysts wall fragments and cyst attached organic material. In contrast, the images of the BBBC010 data set contain only nematodes, such that segmentation is easy. However, overlap occurs more frequently due to the worm density that is higher than for PSA.

## 4. RESULTS AND DISCUSSION

Figure 2 shows exemplary qualitative results demonstrating that our approach works robustly on samples with and without distractors, as well as for dense and sparse object collections.



**Fig. 2.** Qualitative results for PSA (top row) and BBBC010 (bottom row). From left to right: predicted skeleton and endpoints, untangled individual worm skeletons and reconstructed segmentation masks.

### 4.1. Evaluation metric

We quantified object detection performance for individual worms by computing F-scores. This measure is similar to the IoU commonly used in object detection, and it allows us to compare results on BBBC010 with those reported in [6]. We measured precision and recall for different F-score thresholds above which a worm detection was considered to be correct.

We evaluate both skeletons and masks in this work. When evaluating skeletons (single-pixel), small deviations should be allowed. We formulate the calculation of overlap between prediction and the ground truth as a maximum bipartite matching problem, with each predicted skeleton pixel connected to ground truth pixels within a range of 3.

### 4.2. Experiments on the PSA data set

The PSA data set contains line and mask annotations as ground truth. It was collected on 10 different dates, and we chose the 367 images from the last three dates as test and the others as training data. To avoid training instability, we ignored images without worms during training.

Our approach achieved 90.34% precision and 86.28% recall for mask detection at a F-score threshold of 0.5. After increasing the F-score threshold to 0.8, the precision was still 75.85%, with 73.02% of the worms detected (Table 4.3).

It is worth noting that the performance drop from skeleton to mask was more pronounced on the PSA data set, for example  $\approx 7\%$  at a F-score threshold of 0.8, while no significant decline could be observed on BBBC010. This resulted from the smaller worm widths in the PSA data set for which the error of width estimation (Section 2.3) was relatively larger.

Analyzing the effect of balanced labels (loss  $L$ , Section 2.1), we trained a standard segmentation U-Net with unbalanced labels using a binary cross-entropy loss: The model turned out to predict all pixels as background, the larger class (data not shown).

Finally, studying the effect of the ground truth slack (Section 2.1), we trained the model using Gaussians with different standard deviations. Table 1 shows that the variants with ground truth slack performed clearly better than those without.

### 4.3. Experiments on the BBBC010 data set

For BBBC010, object masks are provided as ground truth. We hence used the morphological skeleton of the ground truth masks to train our model with line annotations.

In addition, we employed data augmentation to increase the size of the training set: We performed gamma correction with  $\gamma = 0.5, 2$  to generate images with different contrast. Afterwards, original and generated images were rotated with a step size of 30 degrees. Overall, the training set was expanded to 36 times its original size.

For training, we split the BBBC010 data set into two parts (A01-B24, C01-E04) and performed two-fold cross valida-

Pre./Rec. (%)		F-score				
		0.5	0.6	0.7	0.8	0.9
PSA_skeleton	no slack	89.67 / 84.35	76.93 / 82.24	56.56 / 70.71	42.51 / 53.76	28.65 / 36.24
	slack_2_3	90.38 / 88.00	<b>89.55 / 87.41</b>	85.75 / 84.82	81.00 / 80.24	71.14 / 70.47
	slack_3_5	<b>90.94 / 87.29</b>	<b>89.86 / 86.82</b>	<b>87.44 / 85.18</b>	<b>82.73 / 80.59</b>	<b>75.24 / 73.29</b>
PSA_mask	no slack	85.49 / 82.91	64.19 / 76.16	47.91 / 59.88	31.91 / 39.88	4.00 / 5.00
	slack_2_3	<b>89.90 / 86.98</b>	<b>86.82 / 85.00</b>	82.19 / 80.47	72.45 / 70.93	<b>9.74 / 9.53</b>
	slack_3_5	<b>90.34 / 86.28</b>	<b>87.92 / 84.65</b>	<b>83.70 / 80.58</b>	<b>75.85 / 73.02</b>	9.66 / 9.30
BBBC010_skeleton	no slack	97.68 / 94.51	85.89 / 93.98	62.68 / 84.30	47.35 / 65.47	37.32 / 51.60
	slack_2_3	96.89 / 94.66	94.43 / 94.51	89.08 / 92.99	82.72 / 87.20	78.02 / 82.24
	slack_3_5	<b>97.82 / 95.12</b>	<b>95.07 / 94.59</b>	<b>89.91 / 93.22</b>	<b>83.02 / 87.20</b>	<b>78.45 / 82.39</b>
BBBC010_mask	no slack	87.86 / 96.44	72.32 / 89.83	59.20 / 76.67	46.51 / 60.24	19.44 / 25.18
	slack_2_3	<b>95.94 / 97.08</b>	<b>93.92 / 95.52</b>	<b>90.91 / 92.46</b>	<b>84.20 / 85.63</b>	<b>35.03 / 35.63</b>
	slack_3_5	<b>96.01 / 96.80</b>	<b>94.12 / 95.45</b>	90.34 / 91.75	82.91 / 84.21	27.73 / 28.17

**Table 1.** Detection precision / recall at different F-score thresholds for skeleton/mask detection on the PSA/BBBC010 data sets. Ground truth slack (Section 2.1) variants: slack\_2\_3 refers to the skeleton slack with the standard deviation of the Gaussian  $\sigma = 2$  and the endpoint slack with  $\sigma = 3$ .

tion. Our approach achieved 84.20% precision and 85.63% recall for masks at a F-score threshold of 0.8 (Table 4.3). As for PSA, we found that the variants with ground truth slack performed better than those without (Table 4.3).

On the clean background data BBBC010, also the segmentation/untangling method from [6] can be applied. Our results are superior to the "81% correctly segmented worms" (mask recall) reported at the 0.8 threshold in [6].

## 5. CONCLUSIONS

We have proposed a CNN-based framework for detecting worm-shaped objects. With the focal loss and the ground truth slack strategy, the CNN model can predict worm skeletons and endpoints robustly. Individuals are untangled from worm clusters, and finally we reconstruct a segmentation mask based on the skeleton. The overall pipeline requires only line (skeleton) annotations for training and outputs segmentation masks. Employing a CNN enables our framework to cope also with images with complex background as they occur in the PSA data set. Future work will focus on improving segmentation accuracy, as well as on extracting descriptive features for nematode phenotyping.

## 6. REFERENCES

- [1] T. H. Askary, "Limitations, research needs and future prospects in the biological control of phytonematodes," in *Biocontrol Agents of Phytonematodes*, chapter 19, pp. 446–454. CABI, Wallingford, UK, Oxford, 2015.
- [2] T. I. Moy, A. L. Conery, J. Larkins-Ford, W. Gang, R. Mazitschek, G. Casadei, K. S. Lewis, A. Carpenter, and F. M. Ausubel, "High-throughput screen for novel antimicrobials using a whole animal infection model.," *ACS chemical biology*, vol. 4, no. 7, pp. 527–533, 2009.
- [3] Y. L. Chew, D. S. Walker, E. K. Towilson, P. E. Vértés, G. Yan, A.-L. Barabási, and W. R. Schafer, "Recordings of caenorhabditis elegans locomotor behaviour following targeted ablation of single motoneurons.," *Scientific Data*, vol. 4, 2017.
- [4] R. S. Kamath and J. Ahringer, "Genome-wide RNAi screening in Caenorhabditis elegans.," *Methods*, vol. 30, 2003.
- [5] J. Huang, V. Rathod, C. Sun, M. Zhu, A. Korattikara, Fathi A, I. Fischer, Z. Wojna, Y. Song, S. Guadarrama, and K. Murphy, "Speed/accuracy trade-offs for modern convolutional object detectors.," in *Proc. IEEE CVPR*, 2016, pp. 3296–3297.
- [6] C. Wählby, L. Kamentsky, Z. H. Liu, T. Riklin-Raviv, A. L. Conery, E. J. O'Rourke, K. L. Sokolnicki, O. Visvikis, V. Ljosa, J. E. Irazoqui, P. Golland, G. Ruvkun, F. M. Ausubel, and A. E. Carpenter, "An image analysis toolbox for high-throughput C. elegans assays.," *Nature Methods*, vol. 9, 2012.
- [7] C. Wählby, T. Riklin-Raviv, V. Ljosa, A.L. Conery, P. Golland, F.M. Ausubel, and A.E. Carpenter, "Resolving clustered worms via probabilistic shape models.," in *Proc. IEEE ISBI*, 2010, pp. 552–555.
- [8] T. R. Raviv, V. Ljosa, A.L. Conery, F.M. Ausubel, A.E. Carpenter, P. Golland, and C. Wählby, "Morphology-guided graph search for untangling objects: C. elegans analysis," in *Proc. MICCAI*, 2010, pp. 634–641.

- [9] E. Yemini, T. Jucikas, L. J. Grundy, A. E. Brown, and W. R. Schafer, "A database of caenorhabditis elegans behavioral phenotypes.," *Nature Methods*, vol. 10, 2013.
- [10] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, "Focal loss for dense object detection.," in *Proc. IEEE CVPR*, 2017, pp. 2999–3007.
- [11] H. Law and J. Deng, "CornerNet: Detecting objects as paired keypoints," in *ECCV 2018*, 2018, pp. 734–750.
- [12] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation.," in *Proc. MICCAI*, 2015, pp. 234–241.
- [13] C. R. Maurer, R. Qi, and V. Raghavan, "A linear time algorithm for computing exact euclidean distance transforms of binary images in arbitrary dimensions.," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 2, pp. 265–270, 2003.
- [14] J. Canny, "A computational approach to edge detection.," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 8, no. 6, pp. 679–698, 1986.