

1 MICRA-Net: MICRoscropy Analysis Neural Network to solve 2 detection, classification, and segmentation from a single simple 3 auxiliary task

4 Anthony Bilodeau¹, Constantin V.L. Delmas^{1,5}, Martin Parent^{1,5}, Paul De Koninck^{1,2},
5 Audrey Durand^{3,4,6}, and Flavie Lavoie-Cardinal^{1,5*}

6 ¹CERVO Brain research center, Québec (QC), Canada

7 ²Département de biochimie, microbiologie et bio-informatique, Université Laval, Québec
8 (QC), Canada

9 ³Département d'informatique et de génie logiciel, Université Laval, Québec (QC), Canada

10 ⁴Département de génie électrique et de génie informatique, Université Laval, Québec (QC),
11 Canada

12 ⁵Département de psychiatrie et de neurosciences, Université Laval, Québec (QC), Canada

13 ⁶Canada CIFAR AI chair, Mila, Canada

14 *corresponding author: flavie.lavoie-cardinal@cervo.ulaval.ca

15 **Abstract**

16 High throughput quantitative analysis of microscopy images presents a challenge due to the complexity
17 of the image content and the difficulty to retrieve precisely annotated datasets. In this paper we intro-
18 duce a weakly-supervised MICRoscropy Analysis neural network (MICRA-Net) that can be trained on a
19 simple main classification task using image-level annotations to solve multiple the more complex auxiliary
20 semantic segmentation task and other associated tasks such as detection or enumeration. MICRA-Net
21 relies on the latent information embedded within a trained model to achieve performances similar to
22 state-of-the-art architectures when no precisely annotated dataset is available. This learnt information is
23 extracted from the network using gradient class activation maps, which are combined to generate detailed
24 feature maps of the biological structures of interest. We demonstrate how MICRA-Net significantly alle-
25 viates the Expert annotation process on various microscopy datasets and can be used for high-throughput
26 quantitative analysis of microscopy images.

27 **1 Introduction**

28 The development of powerful microscopy techniques that allow to characterize biological structures with
29 subcellular resolution and on large field of views tremendously increased the complexity of quantitative
30 image analysis tasks [1]. The resulting images exhibit a wide range of structures that need to be identified,
31 counted, precisely located, and segmented. Expert knowledge is commonly required to achieve successful
32 identification and segmentation of the multiple structures of interest in microscopy images [2, 3]. These
33 tasks can be tedious and time consuming especially for large databanks or for the comparison of multiple
34 biological conditions. It was recently demonstrated that deep convolutional neural networks (CNN) are
35 excellent feature extractors [4]. They were successfully applied to segmentation (e.g. whole cells, nuclei,
36 dendritic spines), enumeration (e.g. cell counting), and classification (e.g. state of cell) of structures in
37 microscopy images [5–12]. The most common deep learning (DL) approaches applied to microscopy and
38 biomedical images are fully-supervised and require precisely annotated datasets [9, 11, 12]. Hence, it is
39 often a limiting step in the application of DL for quantitative analysis of biomedical imaging [3, 13, 14].
40 To alleviate the annotation process, weakly-supervised DL methods were introduced [14–17]. Bounding box

41 annotations are commonly used for weakly-supervised segmentation tasks as they are simple, allow the task
42 to be spatially constrained [2, 16, 18–20], and were shown to decrease the annotation phase by 15-fold
43 compared to precise identification of structure boundaries [21]. Methods for training with binary, image-
44 level targets, reducing even further the complexity and duration of the annotation task, have been proposed
45 when multiple instances are displayed on a single image [22]. Unfortunately, when applied to microscopy
46 and biomedical image analysis, such weakly-supervised approaches using whole image annotations, resulted
47 in lower segmentation precision compared to approaches using precisely identified structures [23–25].

48 In this paper we propose MICRA-Net (MICRoscropy Analysis Neural Network), a new approach relying
49 only on image-level classification annotations for training a deep neural network to perform different type
50 of microscopy image analysis tasks such as semantic segmentation, cell counting, and detection of sparse
51 features. MICRA-Net builds on *latent learning* [26], which refers to a model retaining information (*i.e.* latent
52 space) that is not required for the task at hand in order to learn new auxiliary complementary tasks [26].
53 In this work, we leverage the information embedded within a trained classification network to solve multiple
54 complementary, yet very different, tasks relevant to microscopy image analysis. The network uses binary
55 classification targets as input to build a general representation of the specific dataset and generates detailed
56 feature maps from which specific tasks, such as instance segmentation, semantic segmentation, detection,
57 and classification, can be addressed. Even further this showcases the potential of MICRA-Net for addressing
58 various high-throughput microscopy analysis challenges, relying solely on weak image-level annotations for
59 training.

60 2 Results

61 The generation of precisely annotated large datasets to train deep neural networks in a fully-supervised
62 manner remains a challenge in the field of microscopy and biomedical imaging. MICRA-Net, a CNN-based
63 method, addresses this challenge by using solely whole-image binary targets for training. This approach
64 outperforms state-of-the-art DL baselines trained in a weakly-supervised manner for the semantic segmenta-
65 tion of diverse biological structures. It is therefore of great interest for the automated quantitative analysis
66 of microscopy datasets for which no fully-supervised training dataset is available. In the following we first
67 investigate the impacts of the annotation burden, before characterizing the performance of MICRA-Net on
68 synthetic and real data for various tasks. We then evaluate how MICRA-Net can be fine-tuned in order to
69 leverage information from a previously acquired, but different, dataset. Finally, we show how the proposed
70 approach could be used to support Experts in the annotation of sparse and small structures in large images.

71 2.1 Annotation task reduction analysis

72 MICRA-Net is trained on a simple multi-class classification task and therefore only requires the Expert to
73 identify class-specific positive and negative images with respect to the structures of interest. In contrast to
74 the identification of the structure boundaries using precise or bounding box contours, image-level annotations
75 do not require to specify the positions of the object in the field of view of the microscopy images (Figure 1a).

76 We quantified the required time to generate annotations with different levels of precision (precise, bound-
77 ing boxes, and points) by conducting a User-Study in which we asked participants to annotate the testing
78 images from the Cell Tracking Challenge on 6 different cell lines [8] (see Methods). We analysed the inter-
79 participant variability by comparing the annotations of the participants in a one-versus-all manner. The
80 metric used to assess this variability combines both the level of association between objects (F1-score) and
81 the precision on the contour of annotated objects [27] (IOU, Figure 1b and Supplementary Fig. 1-3). Since
82 it is not possible to report the IOU between points annotations, we show the average F1-score as a constant
83 line on Figure 1b. As a general tendency, simpler annotation tasks reduced the inter-participant variabil-
84 ity (higher F1-score at given IOU). For each selected cell lines, we report the median distances between
85 associated point markers (centroid of objects, Figure 1c) and the average distance between the contours of
86 associated objects (Figure 1d) as a mean to probe the variability of annotations. We measured a median
87 error on the cell boundaries ranging from 2 to 7 pixels depending on the cell line (Figure 1d). Several factors
88 can reduce the precision of the annotations, such as the contrast (Fluo-N2DL-HeLa - high contrast vs PhC-
89 C2DL-PSC - low contrast) and the shape (Fluo-N2DH-GOWT1 - round vs PhC-C2DH-U373 - irregular)

90 (Figure 1c,d and Supplementary Fig. 3). The required time to annotate a single cell is increased by approx-
 91 imately 2 folds when going from points annotations to bounding boxes, and from bounding boxes to precise
 92 annotations (Figure 1e). Finally, we evaluated the difference between weak-supervision using MICRA-Net's
 93 training scheme and fully-supervised training both in terms of interactions and annotation time (Figure 1f
 94 and Methods). Compared to the precise annotations required to train fully-supervised DL approaches, the
 95 generation of whole image binary annotations reduces on average by 6 folds the required annotation duration.

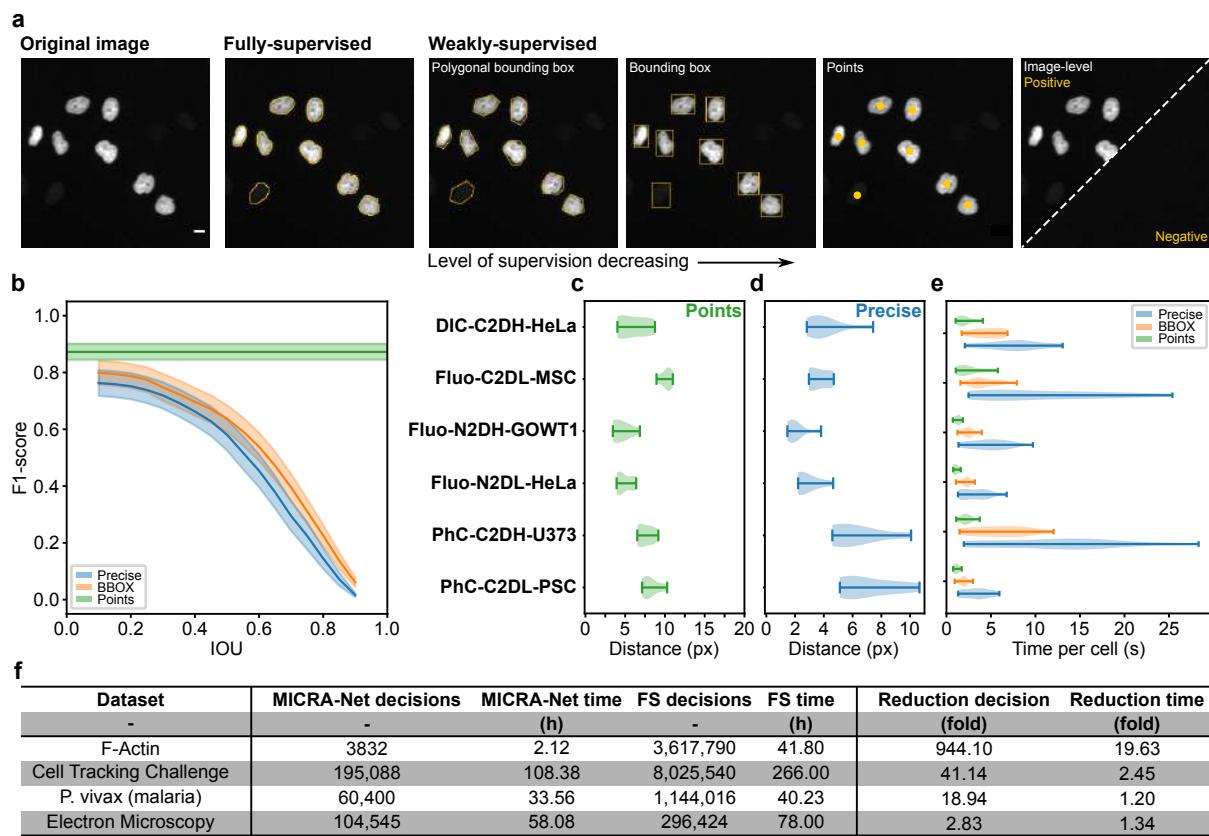


Figure 1 : Caption is on the next page

96 2.2 MICRA-Net architecture and baselines

97 Figure 2a shows the architecture of MICRA-Net, which was designed around a CNN architecture, more
 98 specifically using a U-Net-like encoder, composed of 8 convolutional layers (L^1 to L^8) followed by a fully
 99 connected layer. The rationale is that U-Net is an established method able to solve multiple biomedical
 100 tasks. The gradient class activated maps (Grad-CAM, see Methods) were extracted for each predicted class
 101 and at every layer of the network (Figure 2a-c & Supplementary Fig. 5a,b). Thereafter, Rectified Linear
 102 Unit (ReLU) activation and thresholding on the Grad-CAM of the last convolutional layer (L^8) were applied
 103 to generate a coarse class-specific feature map [28]. To increase the information contained in the extracted
 104 feature map, local maps from layers L^{1-7} were concatenated, resulting in a class-specific 7-dimensions feature
 105 space (Figure 2b,c). We retrieved the first principal component of every pixel using principal component
 106 analysis (PCA) decomposition on the feature space to generate a single feature map that was used to solve
 107 different sets of specific auxiliary tasks (Figure 2b,c & Methods).

108 To characterize the performance of MICRA-Net we compared the results obtained on different datasets
 109 with three established baselines: i) pretrained U-Net (in the following sections referred to as U-Net) [9], ii)
 110 Mask R-CNN [10], and iii) Ilastik [29]. These baselines were chosen as they are widely used in the literature
 111 and they allow semantic segmentation with none or simple modifications (see Supplementary Note 2 & 3 for

Figure 1: Various supervision levels can be employed for training a DL model to segment structures of interest in microscopy images. a) Representative image from the Cell Tracking Challenge dataset [8] overlayed with the corresponding fully- and weakly-supervised annotations. Annotated images are presented in decreasing spatial level of supervision and required annotation time (*from left to right*). b) We report the averaged inter-participant variability from the User-Study from 6 selected cell lines of the Cell Tracking Challenge using three levels of supervision (precise, bounding boxes (BBOX), and points). Representative examples from the participants may be found in Supplementary Fig. 1-3 as well as the specific curves per cell line in Supplementary Fig. 4. The inter-participant agreement was calculated using the F1-score as a function of IOU for precise (blue) and BBOX (orange) annotations in a all versus one manner [27]. The F1-score for points annotation (green) was calculated with a maximal distance of association of 30 pixels. Plotted are the bootstrapped mean (line) and 95% confidence interval (shade, 10 000 repetitions). c-e) Shown are the distribution of median scores from the inter-participant comparison calculated in a all versus one manner. c) Distance between associated point markers. d) Average distance between the precise contours of participants annotations was calculated for precise annotations. e) Average required time per objects on different cell lines for each supervision level. f) Evaluation of the annotation task required to generate the training set for all microscopy datasets used throughout the paper for fully-supervised (FS) and MICRA-Net approaches. Reported above is the effective number of decisions (number of extracted crops for MICRA-Net and number of edge pixels for fully-supervised learning) and the required time in hours. For MICRA-Net the number of decisions corresponds to the number of extracted crops and the annotation time per crop (assignation of a positive or negative annotation) was on average 2 seconds for all datasets. For fully-supervised learning, the decision and annotation time was evaluated for each dataset separately on a precisely annotated subset of images (see Methods).

112 dataset specific implementation details). This rendered a similar task between the baselines and MICRA-Net.

113 2.3 Multi-class segmentation of synthetic images

114 To validate the classification and segmentation performance of MICRA-Net, we created a synthetic dataset
115 containing N randomly sampled cluttered handwritten digits from the MNIST dataset [30] (Modified MNIST
116 dataset, Figure 2c & Methods). Each image may contain several instances of digits (from 0 to 9), as well
117 as variable levels of noise and signal to mimic slight variations akin to those that may be observed in
118 microscopy images (see Methods). The first step was to classify the digits appearing on each image to
119 validate the representation capability of the network, which is confirmed by the obtained class-wise mean
120 classification testing accuracy of $(98.9 \pm 0.5)\%$ (mean \pm std).

121 In addition to the classification task, MICRA-Net generates class-specific segmentation maps of the digits
122 in the modified MNIST dataset. Using the information embedded in the Grad-CAMs of the hidden layers
123 (L^{1-7}) to precisely locate each digit in the image significantly increased the segmentation performance of the
124 network when compared to the maps obtained from the Grad-CAMs of the last layer only (L^8) (Figure 2d,
125 Supplementary Fig. 5c,d & Supplementary Fig. 6). A U-Net [31] trained on the same dataset using a
126 fully- and weakly-supervised training scheme was used as a baseline to better evaluate the performance of
127 MICRA-Net. Fully-supervised learning consisted in training with the binary digits contours from MNIST,
128 while weak contours were generated by a dilation of the digits with a square of size $\{5, 10, 25\}$ pixel as
129 a structuring element (see Supplementary Note 1). Figure 2e shows that MICRA-Net achieves similar
130 or superior segmentation performance compared to all weakly-supervised training instances of the U-Net
131 and is only outperformed on all measured metrics (F1-score, intersection over union (IOU), and symmetric
132 boundary dice (SBD)) by fully-supervised training (Supplementary Fig. 7 & Supplementary Tab. 1).

133 2.4 Class-specific segmentation of super-resolution microscopy images

134 The next question that needed to be addressed was the applicability of our approach for super-resolution
135 microscopy image segmentation, for which precisely annotated datasets are rarely available. The auxiliary
136 task was the semantic segmentation of STimulated Emission Depletion (STED) microscopy images of two

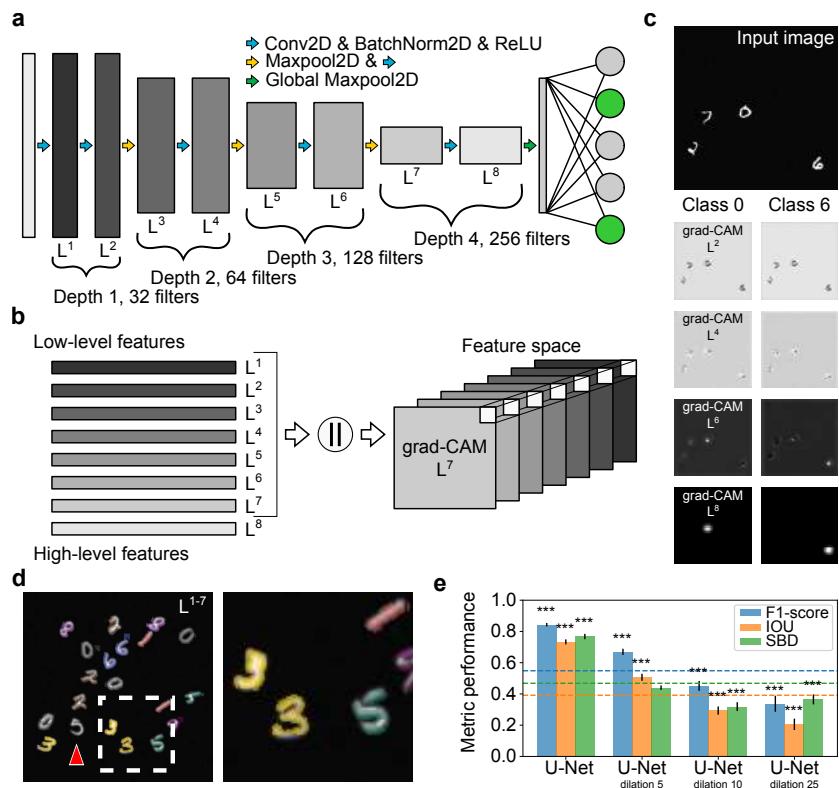


Figure 2: MICRA-Net architecture and experimental results on the modified MNIST dataset. a) MICRA-Net architecture (detailed in the Methods section). Each depth is composed of two sequential convolutional layers (Conv2D), batch normalization (BatchNorm2D), and Rectified Linear Unit activation (ReLU). A 2×2 max pooling (MaxPool2D) was employed to increase the richness of the representation from the model. A linear layer is used to project the globally pooled L^8 layer (256 filters, Global Maxpool2D) to the specified number of classes. b) Concatenation of low- and high-level feature maps obtained from the Grad-CAMs of every layer is performed to generate the multi-dimensional feature space for every predicted class. c) Feature maps generated from the calculated Grad-CAMs for class 0 and 6 on the modified MNIST dataset. Each activated class is backpropagated through the network and a local map for each layer of the network (L^{1-8}) is computed. See Supplementary Fig. 5 for layer specific grad-CAMs. d) Detailed segmentation maps of the digits of a representative image (256 \times 256 pixel) and insets (right, dashed white box) from the modified MNIST dataset using MICRA-Net. The color code corresponds to the digit class and the red arrow indicates a missed digit in the field of view. e) Mean performance over the 10 classes obtained with the U-Net trained with and without dilation of the ground truth contours. The segmentation maps are presented in Supplementary Fig. 7a. MICRA-Net segmentation performance (color-coded dashed lines, see Supplementary Fig. 5 for distributions) surpasses the U-Net trained with 10 pixels dilation and is not statistically different from the U-Net trained with 5 pixels dilation on all measured metrics. Only fully-supervised training outperforms MICRA-Net segmentation on all measured metrics. p -values are calculated using resampling (see Methods) and are reported in Supplementary Tab. 1. Bar graphs show the mean values and standard deviation.

137 nanostructures of the F-actin cytoskeleton in neurons: 1) a periodical lattice structure (rings) and 2) longitudinal fibers (Figure 3a,b) [2]. The F-actin nanostructure segmentation task is challenging since the
 138 morphology of neurons is highly variable throughout the dataset, and there are many distractors around
 139 the structures of interest [2]. Figure 1f shows that image-level annotation reduced by more than 19 folds
 140 the time required by an Expert to generate the training dataset compared to precise identification of the
 141 structure boundaries that would be required for fully-supervised DL approaches. This also corresponds to a
 142 reduction of the annotation time of more than 3 folds compared to the tracing of polygonal bounding boxes,
 143

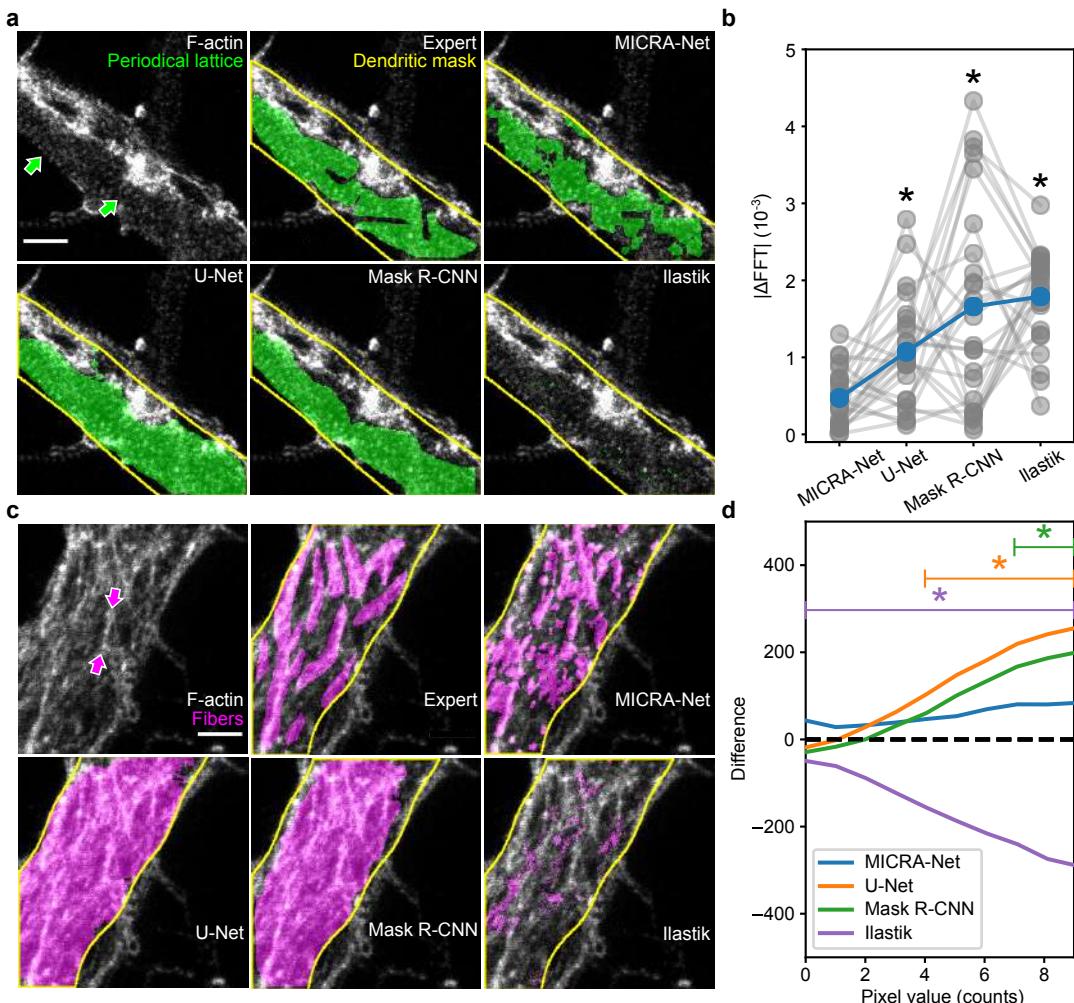


Figure 3: Semantic segmentation of F-actin nanostructures observed on super-resolution microscopy images. a,c) Representative raw images from a dataset of STimulated Emission Depletion (STED) microscopy images of two F-actin nanostructures in fixed cultured hippocampal neurons: periodical lattice (a) and longitudinal fibers (c). Arrows point towards the periodical lattice (green) and longitudinal fibers (magenta). Segmentation masks obtained from an Expert, MICRA-Net, weakly-supervised U-Net, weakly-supervised Mask R-CNN, and weakly-supervised Ilastik are also reported for both structures as comparison. b) Performance evaluation of MICRA-Net and weakly-supervised baselines segmentation on the precisely annotated testing dataset using custom metrics for periodical lattice (rings). The FFT metrics compares the frequency content of the provided masks. The segmentation resulting from MICRA-Net is not significantly different from the Expert annotations, while the other baselines are (U-Net, Mask R-CNN, and Ilastik). d) The intensity distribution metric evaluates the difference between the pixels found within the precise Expert annotations and the DL-based segmentation approaches for the F-actin fibers nanostructures (see Methods). The raw number of low intensity pixel segmented by MICRA-Net is not significantly different for any low value of intensity pixel from the Expert. This is not the case for all baselines (U-Net, Mask R-CNN, and Ilastik) which annotated a significantly different number of low intensity pixels. The complete range of pixel values is shown in Supplementary Fig 12. p -values are calculated using resampling (see Methods) and are reported in Supplementary Tab. 4, 5. Performance evaluation was performed within the dendritic mask (a,c: yellow line). a,c) Scale bars: 1 μ m.

144 which were recently used for weakly-supervised training of the U-Net architecture on this dataset [2].

145 On the main classification task, MICRA-Net achieves an accuracy of 75.2% and 83.7% on the testing

146 dataset for the F-actin periodical lattice and longitudinal fibers, respectively. This is inline with a mean
147 inter-participant classification accuracy of $(80 \pm 5)\%$ and $(75 \pm 7)\%$ for periodical lattice and longitudinal
148 fibers respectively (calculated from 6 participants using a leave-one-out scheme from 50 images), confirming
149 the model capability to handle data of this nature (Supplementary Fig. 8). As described in the previous
150 section, an informative feature map was generated from the PCA decomposition of the combined L^{1-7}
151 extracted features. Thresholding of this feature map resulted in detailed binary masks that were used to
152 solve the segmentation task. We relied on a *precisely annotated dataset* consisting of 25 images of each
153 structure (Supplementary Fig. 9) to evaluate the performance of all trained models: i) MICRA-Net, ii)
154 multi-participants polygonal bounding box annotations (6 participants on 25 images of each structure: *User-
155 Study*), iii) U-Net trained with polygonal bounding boxes [2], iv) Mask R-CNN trained with polygonal
156 bounding boxes, and v) Ilastik trained using scribbles (see Methods & Supplementary Note 2 for specific
157 details). MICRA-Net achieved equivalent or superior segmentation performance on the *precisely annotated
158 dataset* in comparison to both the *User-Study* and all baselines when comparing the common segmentation
159 metrics (Supplementary Fig. 9-11 & Supplementary Tab. 2, 3). Thus, even if trained with weak image-level
160 annotations, MICRA-Net can extract the necessary structural information to generate detailed segmentation
161 maps for both nanostructures.

162 A qualitative visual inspection of the segmentation masks suggested that MICRA-Net segmentation
163 produced a finer detailed mask compared to the weakly-supervised baseline segmentation [2], especially for
164 fibers, for which it provides detailed segmented contours of single fiber strains (Figure 3c, Fibers). Custom
165 performance metrics that were adapted to the F-actin nanostructures were required to better characterize this
166 observation. For the F-actin periodical lattice, we measured the Fourier Transform (FFT) of the segmented
167 areas for frequencies corresponding to the periodicity of the lattice (180-190 nm [32]) (Figure 3b & Methods).
168 The FFT-metric calculated on the areas segmented with MICRA-Net is not significantly different from the
169 one obtained from the *precisely annotated dataset* (Figure 3b). For all other baselines, evaluation of the
170 FFT-metric on the segmented areas shows a significant difference with the *precisely annotated dataset*. This
171 suggests a better segmentation of the periodic structure for our approach over weakly-supervised baselines
172 (Supplementary Tab. 3, 4). Similarly, a custom metric based on the pixel intensity distribution of the
173 segmented areas was developed to evaluate the approaches on the fiber segmentation task (see Methods).
174 While no difference was observed for the regions identified with MICRA-Net compared to the regions from the
175 *precisely annotated dataset*, a significant increase in the proportion of low-intensity pixels (regions between
176 single fibers) was observed for all weakly-supervised baselines (Figure 3d & Supplementary Tab. 5). This
177 supports a higher accuracy to precisely identify the contours of individual fibers or periodical lattice regions
178 of MICRA-Net over weakly-supervised U-Net segmentation.

179 2.5 Single cell semantic segmentation

180 Cell counting and segmentation is a common challenge in high-throughput analysis of optical microscopy
181 images [8, 9, 12, 33, 34]. Both fully- and weakly-supervised DL approaches were shown to be very powerful
182 to assess these tasks on multiple cell lines [7, 25]. We first highlight some prerequisite of the dataset to train
183 MICRA-Net (and baselines) at solving an instance segmentation task using 6 selected cell lines from the Cell
184 Tracking Challenge (CTC) [8]. For weakly-supervised learning from image-level targets, a sufficient amount
185 of negative samples (images not containing the object of interest) is required to extract informative context
186 from an image, *i.e.* to distinguish the cells in the field of view. We trained MICRA-Net on 256×256 pixel
187 crops from the resampled images of the CTC (with an effective pixel size of $0.5 \mu\text{m}$, Supplementary Tab. 6)
188 and obtained a classification accuracy of $(95.8 \pm 0.4)\%$ (calculated from 5 network instances). Despite
189 having a high classification accuracy, MICRA-Net detection and segmentation performances were strongly
190 reduced when no negative samples were provided (Supplementary Fig. 13, DIC-C2DH-HeLa and Fluo-N2DH-
191 GOWT1). It is therefore necessary to adapt the size of the training images that are provided to the network
192 to the size of the structures of interest, ensuring that enough images contain only background (Supplementary
193 Tab. 6 for selected factors). Another requirement when training a deep learning architecture is that the object
194 of interest can fit entirely within the field of view. Otherwise the model has no information on how different
195 parts of an object should be tied together. To reflect this statement, we trained both U-Net and Mask
196 R-CNN on a resized version of the CTC dataset containing positive and negative samples on all cell lines
197 (Supplementary Fig. 14 and Supplementary Tab. 6 for scale factors). We observe that the performance of all

198 models is significantly lower on the DIC-C2DH-HeLa cell line at this scale. Since both training conditions
 199 cannot be met on this cell line, we removed it from training. Hence, we report the performance of all trained
 200 models on 5 selected cell lines from the CTC.

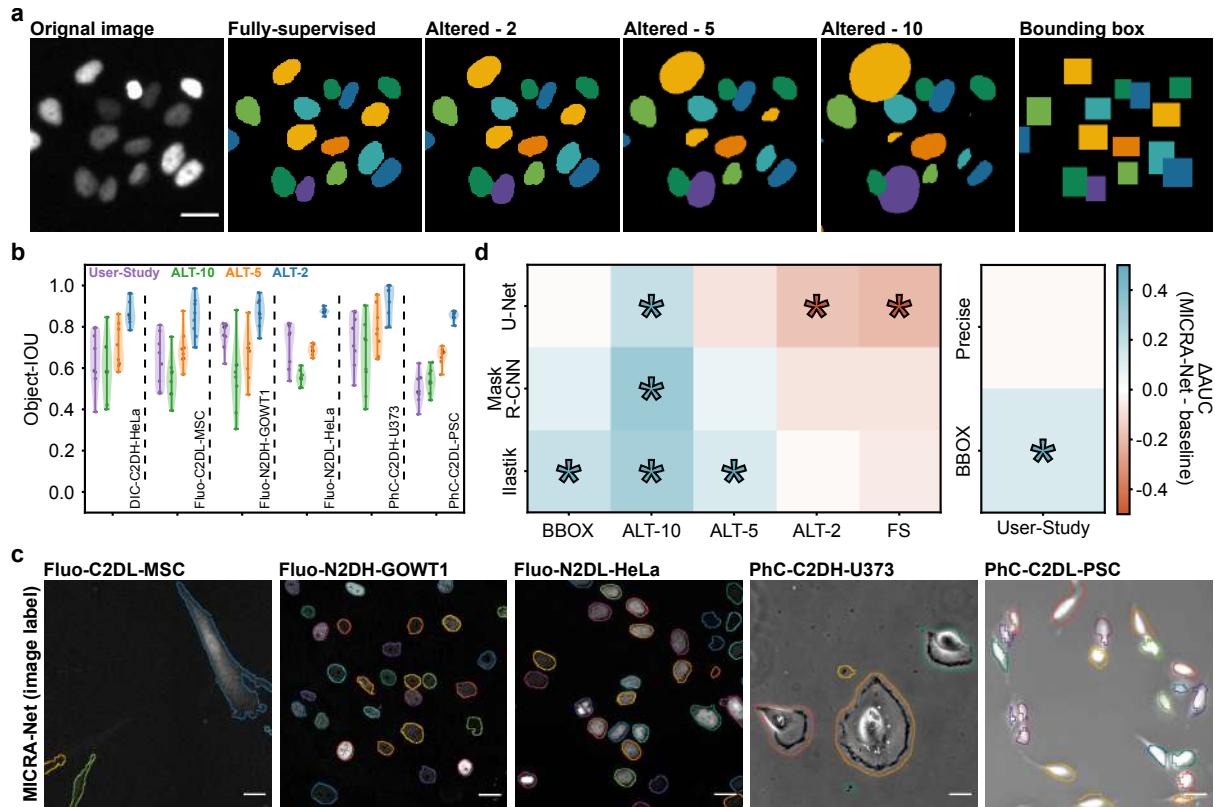


Figure 4: Cell counting and segmentation on 5 selected cell lines of the *Cell Tracking Challenge dataset* (CTC). a) Representative examples of various levels of supervision used to train the selected baselines. The altered- \mathcal{X} are obtained from a binary object dilation/erosion where the transformation is sampled from a normal distribution with 0-mean and \mathcal{X} -standard-deviation. b) Quantification of the IOU between associated objects for the User-Study and altered versions (ALT- \mathcal{X}) of the testing set with the ground truth objects for each cell line of the CTC. The precision of the participants is equivalent to an ALT-5 version of the testing set. c) Representative examples of MICRA-Net semantic instance segmentation. Each outline color depicts a different segmented object. See Supplementary Fig. 16-18 for baseline examples. d, left) We compared the difference of the pooled area under the curve (AUC, F1-score vs. IOU) of all cell lines for MICRA-Net over the baselines on the precisely annotated dataset. The raw curves are available in Supplementary Fig. 19-23, and the non-pooled data in Supplementary Fig. 24. Higher and lower performance of MICRA-Net are reported in blue and red respectively. MICRA-Net is only outperformed by U-Net trained using ALT-2 or fully-supervised training. d, right) We compared the pooled AUC for all cell lines for the conducted User-Study using precise annotations and bounding boxes. The precision of the segmentation masks generated with MICRA-Net is similar to the precise annotations and better than the bounding boxes obtained in the User-Study. Stars are used to highlight a significant change (Supplementary Tab. 9, 10). All scale bars are 25 μ m.

201 As a proof of concept, using the CTC, for which precise annotations are available, we compared the
 202 semantic instance segmentation of MICRA-Net with fully- and weakly-supervised baselines: U-Net [9], Mask
 203 R-CNN [10], and Ilastik [29] (see Supplementary Note 3 for specific implementation details). The weak
 204 supervision consisted in dilating/eroding each object of the fully-supervised dataset by a value sampled from
 205 a normal distribution with 0 mean and standard deviation in {2, 5, 10} (Altered- \mathcal{X} or ALT- \mathcal{X}), or by taking
 206 the bounding boxes of each objects (see Methods and Figure 4a). Since no precisely annotated testing dataset

207 was provided for the CTC, we precisely annotated 4 images for each cell line to evaluate the segmentation
208 performance of both approaches (*precisely annotated dataset*). We compared the achievable annotation
209 precision from participants to that of altered versions of the *precisely annotated dataset* (Figure 4a,b).
210 Figure 4b shows the distribution of IOU between associated objects (Object-IOU) of the User-Study (8
211 participants) and the altered versions of the dataset (8 repetitions) when comparing to the original *precisely*
212 *annotated dataset* testing set for each selected cell lines of the CTC. From Figure 4b we can conclude that the
213 distribution of the User-Study is similar to the distribution of ALT-5. Hence, training a DL architecture with
214 a training set obtained from multiple participants (e.g. crowd-sourced) should result in baseline performance
215 similar to the one trained with ALT-5.

216 To solve the semantic instance segmentation task for MICRA-Net, we trained MICRA-Net to predict both
217 the presence of a cell and the contact between cells, which was subtracted from the former (see Methods
218 & Supplementary Fig. 15). A binary segmentation map was obtained by using an Otsu threshold [35]
219 (Figure 4c). We used the F1-score detection as a function of the intersection over union (IOU) between
220 associated objects (see Methods) to quantify the results [27]. We extracted a single score from the curves by
221 calculating the normalized area under the curve (AUC). Figure 4d (left) reports the variation of MICRA-
222 Net in AUC from baselines trained with various level of supervision when pooling data from all selected cell
223 lines (see Methods and Supplementary Fig. 16-24). As shown in Figure 4d and Supplementary Fig 24, the
224 performance of baselines which were developed for fully-supervised datasets is affected when reducing the
225 supervision level (Supplementary Fig. 16-18). This is also depicted by the low classification accuracy of the
226 baselines compared to MICRA-Net (Supplementary Tab. 7). Strikingly, MICRA-Net achieves cumulative
227 similar performance to fully-supervised Mask R-CNN and Ilastik for the semantic instance segmentation on
228 the 5 cell lines. MICRA-Net is only significantly outperformed by U-Net when training is performed on the
229 fully-supervised or a slightly altered (ALT-2) dataset. Therefore, when no precisely annotated and proofed
230 dataset is available, or when the annotation error may be high, the performance of baseline architectures
231 cannot be guaranteed to achieve superior semantic instance segmentation performance on all cell lines (see
232 Supplementary Fig. 24, and Supplementary Tab. 8 and 9). The performance of the conducted User-Study
233 on the testing dataset were also compared to MICRA-Net (Figure 4d (right), Supplementary Fig. 2, 3).
234 A significant increase in performance is measured for MICRA-Net for bounding boxes and no significant
235 change is observed when comparing to precise annotations. Given the previous results, an approach like
236 MICRA-Net will perform similarly (or better) to the presented baselines for semantic instance segmentation
237 when no precisely annotated dataset is available. More importantly, MICRA-Net reduced by a factor of
238 40 the number of Expert decisions required to annotate the training dataset and by more than 150h the
239 necessary annotation time usually needed to complete this task while achieving precise human-level precision
240 (Figure 1f and Figure 4c).

241 2.6 Multi-device analysis

242 While DL approaches can be very powerful when tackling tasks on very similar images, challenges are often
243 encountered when the imaging conditions change over time (e.g. due to a new device) [37, 38]. To increase
244 the applicability of the proposed method to various experimental conditions, we investigated how MICRA-
245 Net could be fine-tuned on a new dataset that contains similar structures but acquired on a new device.
246 To address this, a brightfield microscopy dataset of Giemsa-stained [39] P. Vivax (malaria) infected human
247 blood smears was used (Figure 5a), for which the training and testing datasets had very distinct intensity
248 distributions (Figure 5a,b) [33, 36].

249 The first attempt to solve the classification task consisted in predicting the presence of infected smears in
250 a 256×256 pixel image. A mean testing classification accuracy of $(80 \pm 10)\%$ (mean \pm standard deviation,
251 calculated from 5 different instances of the network) was obtained. Since the testing images had a very
252 different pixel intensity distribution, we investigated whether the classification results could be improved
253 by adjusting for this. To this aim, we considered i) modifying the threshold of the linear layer and ii)
254 fine-tuning a model by training on $\{12, 24, 36\}$ sampled images from the test set using a k -fold training
255 scheme (see Supplementary Note 4 & Supplementary Fig. 25). We repeated the fine-tuning process 5 times
256 from each of the 5 naive instantiations (as starting points) while allowing i) linear layer [*Linear*], ii) linear
257 layer and depth 4 [*Linear + 4*], iii) linear layer and depths 3 and 4 [*Linear + 3, 4*], and iv) all [*All*] layers
258 to be updated (Figures 2a & 5c). A testing classification accuracy over 87% was obtained when updating

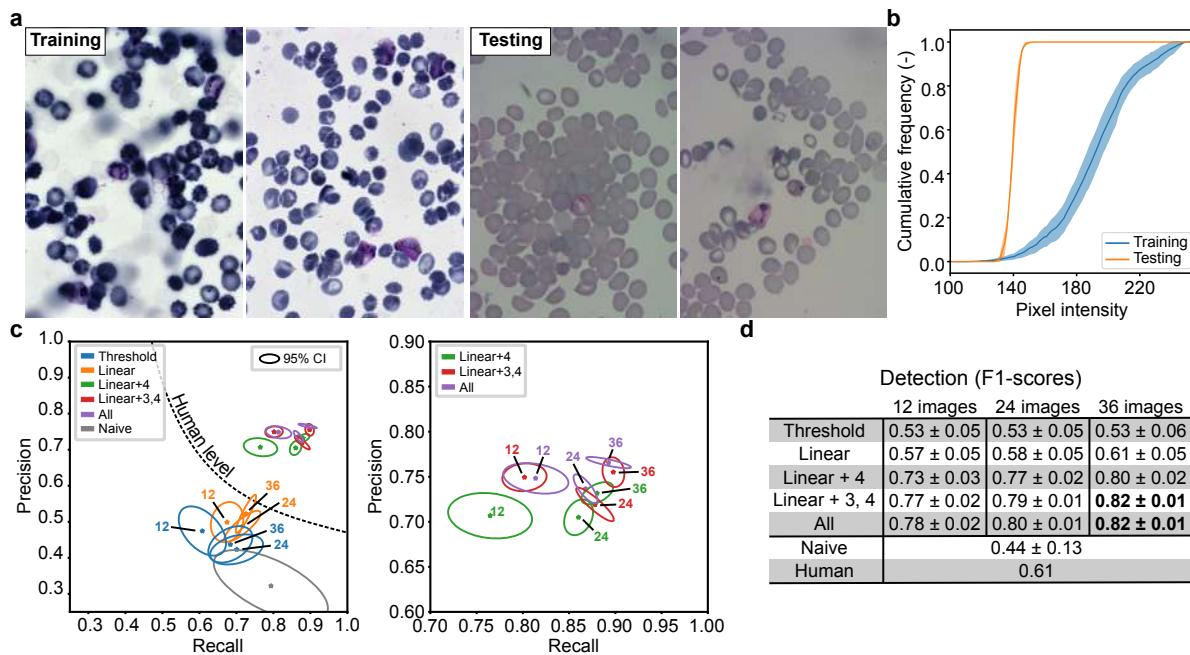


Figure 5: Segmentation of two different datasets of bright field microscopy images of Giemsa-stained red blood cells from [36]. a) Representative images from the training (2 left) and testing (2 right) datasets. The training dataset is composed of images taken from two different laboratories, while the testing images were acquired in a third laboratory. b) A change in the brightness and contrast is observed between the training and testing dataset. This results in a large difference in the mean pixel intensities (training: blue line, testing: orange line, with standard deviation: pale region) of the training and testing images. c, left) A precision-recall graph quantifies the detection performance of MICRA-Net on the testing dataset. Without fine-tuning, the performance on the testing dataset (Naive, grey ellipse) is characterized by a recall of 0.79, and a poor precision of 0.32. A variable number of images ($\{12, 24, 36\}$) from the testing dataset were used to adjust the detection threshold (Threshold, blue ellipse), which increased the precision but also reduced the recall by approximately 2 folds. Fine-tuning of the model on the sampled $\{12, 24, 36\}$ images from the testing set with different settings: i) allowing the linear layer (orange), and ii) different depths (depth 4: green; depth 3, 4: red) to be updated (see Supplementary Fig. 25 & Supplementary Note 4) resulted in precision-recall above human level detection. c, right) Zoomed region of the precision-recall performance of MICRA-Net. When the number of trainable parameters increases, the number of images required for a model with good generalization properties also increases. d) Detection efficiency (F1-score) of the various trained fine-tuned models. As a general tendency, increasing the number of images sampled from the testing set and allowing more layers to be updated resulted in better detection of infected red blood cells. The best detection accuracy (0.82 ± 0.01) was obtained by updating either *Linear + 3, 4* or *All* layers. This again demonstrates the capability of MICRA-Net to be fine-tuned and used across different microscopes. See Supplementary Tab. 17 for calculated *p*-values.

259 the threshold and over 88% for all fine-tuned models, demonstrating the capability of MICRA-Net to be
 260 fine-tuned on similar tasks performed on images acquired on different devices (Supplementary Table. 16 for
 261 detailed classification results).

262 In the context of parasite detection and stage determination for malaria, the most important task consists
 263 in the detection of infected cells [33]. When trained solely on the original training set, MICRA-Net performed
 264 worse on the detection task, obtaining a F1-score of 0.44 ± 0.13 (Figure 5c, d). However, with fine-tuning of
 265 at least the linear layer and the depth 4 of the architecture, the F1-score was significantly increased, beating
 266 the inter-expert accordance (0.61 [36]). Additionally, increasing the number of images sampled from the
 267 testing set can significantly increase the detection accuracy (Supplementary Tab. 17). The best detection
 268 accuracy (0.82 ± 0.01) was obtained by updating either *Linear + 3, 4* or *All* layers. This again demonstrates
 269 the capability of MICRA-Net to be fine-tuned and used across different microscopes.

270 We compared the segmentation results of MICRA-Net with Expert precise annotations. Due to the lack of
271 a precisely annotated dataset in the original publication by [33], we manually segmented all infected smears
272 from the test set (303 smears). In contrast to the results obtained for the detection accuracy, updating
273 more layers while fine-tuning (*Linear + 3*, *4* {12, 24, 36}, and *All* {12, 24}) significantly reduced the IOU
274 compared to only updating the linear layer (Supplementary Fig. 26 & Supplementary Table 18). Hence, a
275 trade-off should be made by the users according to their specific needs. For instance, with these P. Vivax
276 datasets, the best trade-off to maximize both detection and segmentation efficiency requires the fine-tuning
277 of at least the linear layer and depth 4.

278 2.7 Expert detection and segmentation assistance

279 The next step was to assess how MICRA-Net could be implemented as a tool to guide Experts in the
280 annotation of sparse and small structures in large images of an electron microscopy dataset. Our approach
281 was tested on a dataset of Scanning Electron Microscopy (SEM) images of ultrathin mouse brain sections
282 in which axons were genetically labeled with a small engineered peroxidase APEX2 [40] (referred to as Axon
283 DAB, see Methods). In the SEM dataset, 1-10 small axonal regions (with an averaged size of 113×113
284 pixel) needed to be identified in images of around $10\,000 \times 10\,000$ pixel (Figure 6a). Applied to this dataset,
285 MICRA-Net was used to suggest regions containing the Axon DAB marker and generate segmentation masks
286 of the structure in the regions that were accepted by the Expert.

287 An Expert identified Axon DAB positive regions on the training (158 images) and testing (44 images)
288 sets using point annotations (see Methods). To train MICRA-Net, all positive regions (1024×1024 pixel
289 i.e. $5.12 \times 5.12 \mu\text{m}^2$) centered on the detected Axon DAB were extracted from the original images (image
290 size of $10\,240 \times 10\,240$). As previously stated, MICRA-Net requires negative crops (not containing Axon
291 DAB) for training. Therefore, all negative 1024×1024 pixel crops without overlap (Figure 6a, Methods &
292 Supplementary Note 5) were also included in the dataset.

293 In the context of very sparse detections, positive-unlabeled (PU) learning can improve the performance of
294 a given architecture [41]. On the main classification task, an accuracy between 83% and 90% was obtained for
295 all PU ratios (Supplementary Tab. 19). We next investigated how PU learning could improve the detection
296 rate of Axon DAB in the SEM images and obtained best performances for a PU ratio between 1:5 and 1:16
297 (Figure 6b & Supplementary Tab. 20). The usage of MICRA-Net for this sparse detection task resulted
298 in an increase of the measured recall above the inter-expert accordance (0.791, Supplementary Fig. 27),
299 while requiring from an Expert to proof only 3.13% of a newly acquired image. Accordingly, the area that
300 was inspected by the Expert and consequently the annotation time were reduced by 30 folds. Additionally,
301 MICRA-Net allowed the Expert to detect 57 new Axon DAB regions in the test set (representing 25% more
302 detections) that had been missed by the Expert during the initial image annotation process (Figure 6c).
303 This demonstrates the potential of MICRA-Net as a tool to assist Experts in the analysis of newly acquired
304 images, not only reducing the manual annotation time, but also increasing the recall above the inter-expert
305 variability. An attempt was made at comparing the detection results with Ilastik as a baseline trained
306 on positive pixels obtained from points annotations with constant size. Ilastik achieved a classification
307 accuracy of 8% resulting in an almost complete annotation of a new image (Supplementary Fig. 28). We also
308 inspected how MICRA-Net performed on a second auxiliary task: the segmentation of Axon DAB regions
309 (Supplementary Fig. 29a). For this purpose, an Expert carefully highlighted the boundaries of 170 positive
310 Axon DAB regions sampled from the testing set. As in the detection task, MICRA-Net had the same
311 tendency of achieving better performance with PU ratios above 1:2 and could obtain a maximal IOU score
312 of 0.62 ± 0.03 with the 1:5 ratio (Supplementary Fig. 29 & Supplementary Tab. 21). Application of MICRA-
313 Net to this electron microscopy annotation task was thus successful to reduce the burden of generating the
314 training dataset, while also significantly increasing the discovery of regions of interest that were missed by
315 the manual Expert annotation.

316 3 Discussion

317 While pixel-wise metrics and ground-truth annotations are well established in the field of DL and computer
318 vision with natural images, retrieval of ground truth annotations in biomedical imaging is a laborious process,
319 requires highly-trained Experts, and annotation imprecision often occurs [3, 42] (Figure 1). This stresses

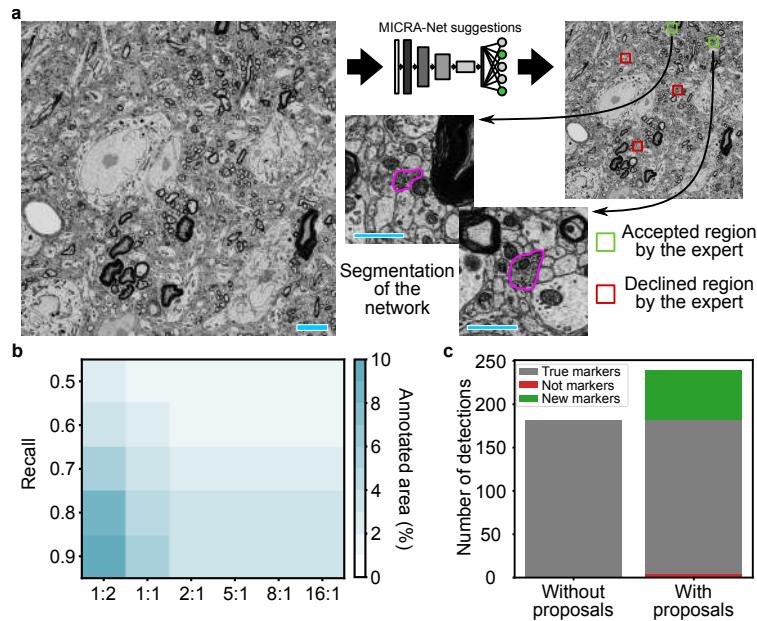


Figure 6: MICRA-Net is used as a tool to assist Experts in the detection of sparse Axon DAB markers in large SEM images of ultrathin mouse brain sections. a) Schematic representation of the proposed approach. MICRA-Net is first swept over the entire field of view with a 75% overlap in both directions to output the probability of presence of an axonal DAB markers. The probability of overlapping crops are then averaged to generate a probabilistic map of positions. The plausible regions are then viewed by the Expert who can accept or decline it. For each accepted region, MICRA-Net generates a segmentation map of the Axon DAB. b) The total percentage of annotated area is color-coded as a function of the positive-unlabeled (PU) ratio at the inter-expert for different recall. Using MICRA-Net trained with a PU ratio of 1:5 as an assisting tool results in the validation of approximately 3% of an image which would require an Expert less than 15 minutes to validate the complete testing set (44 images) and result in a recall of 0.9. The annotated area as a function of the recall for each PU ratio is shown in Supplementary Fig. 27. c) Total number of detections from the testing dataset with and without assistance from MICRA-Net. Using MICRA-Net the Expert could identify 57 new Axon DAB positive regions which correspond to an increase of 25% in the total number of detections. The scale bar is 5 μm for the full field of view and is 1 μm for extracted crops.

320 the need for weakly-supervised DL approaches that do not rely on spatially precise annotations of the
 321 structure of interest, but rather on annotations that are easier and faster to retrieve. MICRA-Net, a CNN-
 322 based method, relies on the information embedded in the latent space of a main simple task, in our case
 323 classification, to learn multiple complementary tasks without the need to generate task-specific precisely
 324 annotated training sets. We designed multiple experiments to challenge MICRA-Net at solving common
 325 microscopy tasks (segmentation, enumeration, or localization) relevant to high-throughput microscopy image
 326 analysis [3, 9]. Unlike multi-task learning [43], MICRA-Net does not combine auxiliary tasks to increase the
 327 learning performance of a main task, nor requires more annotations from the dataset for each task [44, 45].
 328 Hence, the use of MICRA-Net should significantly reduce the burden of task-specific annotation of bioimaging
 329 datasets thereby increasing the accessibility of such deep learning based microscopy image analysis.

330 Our results show that MICRA-Net can be applied to various microscopy modalities and biological con-
 331 texts, while significantly reducing the number of required Expert decisions to generate the training dataset
 332 (Figure 1b). While fully-supervised DL approaches (e.g. based on U-Net or Mask R-CNN architectures) have
 333 the drawback of being costly to train, they can benefit from pre-training [9, 46, 47] given the image space
 334 is similar [48], and have access to precise information about the structure boundaries. On the other hand,
 335 MICRA-Net leverages on the extraction of spatial features from the hidden layers of the network to generate
 336 detailed feature maps using solely, easy to retrieve, binary image-level annotations for training. Considering
 337 the observed reduction of the inter-expert variability when diminishing the complexity of the annotations,

338 this will be an important aspect for future applications leveraging on crowd-sourced annotations for training.
339 MICRA-Net provides similar or even superior performance on multiple tasks to the state-of-the art weakly-
340 and fully-supervised learning approaches, thus making it an unprecedented alternative to address bioimaging
341 analysis challenges for which large and precisely annotated datasets are not available.

342 Additionally we demonstrated that MICRA-Net could be fine-tuned when facing strong variations in
343 the quality of the available datasets, for example when images were acquired on two different microscopes.
344 Fine-tuning of the architecture on few images from another microscopy system was sufficient to achieve
345 better detection efficiency than inter-expert agreement. This is of particular interest for large-scale studies,
346 conducted on multiple sites, that require analysis framework to be easily adaptable to new experimental
347 conditions [33, 49, 50]. Future work on fine-tuning of such approaches to new structures of interest and
348 analysis task will be an important step to increase their accessibility to a larger network of researchers.

349 Lastly, MICRA-Net was used to assist an Expert to perform a complex annotation task, that is the
350 detection of small sparse objects (sections of genetically-labeled axons) in large fields of view of brain sections
351 imaged with Scanning Electron Microscopy. Originally, this task was prone to identification errors and
352 fatigue, limiting the performance of the Experts, and increasing inter-expert variability. MICRA-Net was
353 successfully applied to assist the Experts at finding possible positive regions in the images. Instead of
354 screening the whole field of view, Experts could focus their attention on less than 5% of the image and
355 quickly decline or accept the proposed regions. This allowed an increase in the total number of detected
356 regions of interest (genetically-tagged axons) by 25% while reducing the required annotation time for newly
357 acquired images by 30 folds.

358 Precise annotations, even if obtained from trained Experts, are associated with inter-participant variabil-
359 ity, especially when defining the boundaries (Figure 1). This variability needs to be assessed to characterize
360 the annotated dataset and the precision of the neural network precision [3, 51]. We observed that image-level
361 binary annotations can help to increase the consistency among Experts by reducing the complexity of the
362 annotation task. By alleviating the annotation burden, an approach such as MICRA-Net can help increasing
363 the accessibility of deep learning assisted quantitative image analysis in microscopy. As a whole, it can be
364 used in multi-class detection, segmentation, counting, and classification tasks in bioimaging, for which a
365 precisely annotated dataset is not available or tedious to obtain.

366 Acknowledgments

367 Laurence Emond for F-Actin sample preparation and immunocytochemistry. Francine Nault, Charleen
368 Salesse and Laurence Emond for the neuronal cell culture. Jonathan Marek and Renaud Bernatchez for
369 the development of a custom Python annotation application. Thibault Dhellemmes for inter-expert axon
370 DAB annotations in electron microscopy images. Christian Gagné and Marc-André Gardner for preliminary
371 discussion on semantic segmentation. Annette Schwerdtfeger and Ana Gabela for careful proofreading of the
372 manuscript. Funding was provided by grants from the Natural Sciences and Engineering Research Council of
373 Canada (P.D.K. and F.L.C.), Canadian Institutes of Health Research (P.D.K.), Neuronex Initiative (National
374 Science Foundation and Fond de recherche du Québec - Santé) (P.D.K., F.L.C.), CERVO Brain Research
375 Center Foundation (F.L.C.), the Canadian Foundation for Innovation (P.D.K.). F.L.C. is a Canada Research
376 Chair Tier II, Audrey Durand is a CIFAR AI Chair, and A.B. is supported by a PhD scholarship from the
377 Fonds de Recherche Nature et Technologie Quebec (FRQNT) and an excellence scholarship from the FRQNT
378 strategic cluster UNIQUE.

379 Author contributions

380 A.B. and F.L.C. designed the approach. A.B. implemented the neuronal network architectures, generated
381 the modified MNIST dataset, created the annotation application for the user study and performed all deep
382 learning experiments. A.B., A.D. and F.L.C analysed the results. F.L.C. acquired and annotated the F-actin
383 dataset. C.V.L.D. and M.P. generated and provided the annotated electron microscopy dataset. F.L.C., A.D.
384 and P.D.K. supervised the project. F.L.C, A.D. and A.B. wrote the manuscript.

385 Competing interests

386 The authors declare no competing interest.

387 Data and code availability

388 The MNIST, Cell Tracking Challenge, and P. Vivax datasets are all publicly available online. The F-actin
389 dataset is available at the following website: <https://s3.valeria.science/flclab-micranel/index.html>. The Electron Microscopy dataset included in this study is available from the corresponding author
390 upon reasonable request. All relevant material related to this paper is available on the following website:
391 <https://s3.valeria.science/flclab-micranel/index.html>. Open source code for the MICRA-Net ap-
392 proach is available online: <https://github.com/FLClab/MICRA-Net>
393

394 References

395 [1] Schermelleh, L. *et al.* Super-resolution microscopy demystified. *Nature Cell Biology* **21**, 72 (2019).

396 [2] Lavoie-Cardinal, F. *et al.* Neuronal activity remodels the f-actin based submembrane lattice in dendrites
397 but not axons of hippocampal neurons. *Scientific Reports (Nature Publisher Group)* **10** (2020).

398 [3] Schlegl, T., Seeböck, P., Waldstein, S. M., Langs, G. & Schmidt-Erfurth, U. f-anogan: Fast unsupervised
399 anomaly detection with generative adversarial networks. *Medical image analysis* **54**, 30–44 (2019).

400 [4] LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436 (2015).

401 [5] Gupta, A. *et al.* Deep learning in image cytometry: a review. *Cytometry Part A* **95**, 366–380 (2019).

402 [6] Caicedo, J. C. *et al.* Nucleus segmentation across imaging experiments: the 2018 Data Science Bowl.
403 *Nature Methods* **16**, 1247–1253 (2019).

404 [7] Moen, E. *et al.* Deep learning for cellular image analysis. *Nature methods* 1–14 (2019).

405 [8] Ulman, V. *et al.* An objective comparison of cell-tracking algorithms. *Nature methods* **14**, 1141–1152
406 (2017).

407 [9] Falk, T. *et al.* U-net: deep learning for cell counting, detection, and morphometry. *Nature Methods* **16**,
408 67 (2019).

409 [10] He, K., Gkioxari, G., Dollár, P. & Girshick, R. Mask R-CNN. *arXiv:1703.06870 [cs]* (2018). 1703.06870.

410 [11] Kromp, F. *et al.* An annotated fluorescence image dataset for training nuclear segmentation methods.
Scientific Data **7**, 262 (2020).

412 [12] Stringer, C., Wang, T., Michaelos, M. & Pachitariu, M. Cellpose: A generalist algorithm for cellular
413 segmentation. *Nature Methods* **18**, 100–106 (2021).

414 [13] Wilhelm, B. G. *et al.* Composition of isolated synaptic boutons reveals the amounts of vesicle trafficking
415 proteins. *Science* **344**, 1023–1028 (2014).

416 [14] Cheplygina, V., de Bruijne, M. & Pluim, J. P. W. Not-so-supervised: A survey of semi-supervised,
417 multi-instance, and transfer learning in medical image analysis. *Medical Image Analysis* **54**, 280–296
418 (2019).

419 [15] Papandreou, G., Chen, L.-C., Murphy, K. P. & Yuille, A. L. Weakly-and semi-supervised learning of a
420 deep convolutional network for semantic image segmentation. In *Proceedings of the IEEE international
421 conference on computer vision*, 1742–1750 (2015).

422 [16] Khoreva, A., Benenson, R., Hosang, J., Hein, M. & Schiele, B. Simple does it: Weakly supervised
423 instance and semantic segmentation. In *Proceedings of the IEEE conference on computer vision and
424 pattern recognition*, 876–885 (2017).

425 [17] Xu, J., Schwing, A. G. & Urtasun, R. Tell me what you see and i will show you where it is. In *Proceedings*
426 *of the IEEE conference on computer vision and pattern recognition*, 3190–3197 (2014).

427 [18] Pesce, E. *et al.* Learning to detect chest radiographs containing pulmonary lesions using visual attention
428 networks. *Medical image analysis* **53**, 26–38 (2019).

429 [19] Rajchl, M. *et al.* Deepcut: Object segmentation from bounding box annotations using convolutional
430 neural networks. *IEEE transactions on medical imaging* **36**, 674–683 (2016).

431 [20] Yang, L. *et al.* Boxnet: Deep learning based biomedical image segmentation using boxes only annotation.
432 *arXiv preprint arXiv:1806.00593* (2018).

433 [21] Lin, T.-Y. *et al.* Microsoft coco: Common objects in context. In *European conference on computer*
434 *vision*, 740–755 (Springer, 2014).

435 [22] Vezhnevets, A., Ferrari, V. & Buhmann, J. M. Weakly supervised structured output learning for semantic
436 segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*,
437 845–852 (IEEE, 2012).

438 [23] Dubost, F. *et al.* Weakly supervised object detection with 2d and 3d regression neural networks. *arXiv*
439 *preprint arXiv:1906.01891* (2019).

440 [24] Li, J. *et al.* An em-based semi-supervised deep learning approach for semantic segmentation of
441 histopathological images from radical prostatectomies. *Computerized Medical Imaging and Graphics*
442 **69**, 125–133 (2018).

443 [25] Kraus, O. Z., Ba, J. L. & Frey, B. J. Classifying and segmenting microscopy images with deep multiple
444 instance learning. *Bioinformatics* **32**, i52–i59 (2016).

445 [26] Chatterjee, B. & Poullis, C. Semantic segmentation from remote sensor data and the exploitation of
446 latent learning for classification of auxiliary tasks. *arXiv preprint arXiv:1912.09216* (2019).

447 [27] Caicedo, J. C. *et al.* Evaluation of Deep Learning Strategies for Nucleus Segmentation in Fluorescence
448 Images. *Cytometry Part A* **95**, 952–965 (2019).

449 [28] Selvaraju, R. R. *et al.* Grad-cam: Visual explanations from deep networks via gradient-based localization.
450 In *Proceedings of the IEEE International Conference on Computer Vision*, 618–626 (2017).

451 [29] Berg, S. *et al.* ilastik: interactive machine learning for (bio)image analysis. *Nature Methods* (2019).
452 URL <https://doi.org/10.1038/s41592-019-0582-9>.

453 [30] LeCun, Y., Bottou, L., Bengio, Y., Haffner, P. *et al.* Gradient-based learning applied to document
454 recognition. *Proceedings of the IEEE* **86**, 2278–2324 (1998).

455 [31] Ronneberger, O., Fischer, P. & Brox, T. U-net: Convolutional networks for biomedical image segmentation.
456 In *International Conference on Medical image computing and computer-assisted intervention*,
457 234–241 (Springer, 2015).

458 [32] Xu, K., Zhong, G. & Zhuang, X. Actin, spectrin, and associated proteins form a periodic cytoskeletal
459 structure in axons. *Science* **339**, 452–456 (2013).

460 [33] Ljosa, V., Sokolnicki, K. L. & Carpenter, A. E. Annotated high-throughput microscopy image sets for
461 validation. *Nature methods* **9**, 637–637 (2012).

462 [34] Kromp, F. *et al.* Deep Learning architectures for generalized immunofluorescence based nuclear image
463 segmentation. *arXiv:1907.12975 [cs, q-bio]* (2019). 1907.12975.

464 [35] Otsu, N. A threshold selection method from gray-level histograms. *IEEE Transactions on Systems,*
465 *Man, and Cybernetics* **9**, 62–66 (1979).

466 [36] Hung, J. & Carpenter, A. Applying faster r-cnn for object detection on malaria images. In *Proceedings*
467 *of the IEEE conference on computer vision and pattern recognition workshops*, 56–61 (2017).

468 [37] Belthangady, C. & Royer, L. A. Applications, promises, and pitfalls of deep learning for fluorescence
469 image reconstruction. *Nature methods* 1–11 (2019).

470 [38] Weigert, M. *et al.* Content-aware image restoration: pushing the limits of fluorescence microscopy.
471 *Nature methods* **15**, 1090–1097 (2018).

472 [39] Barcia, J. J. The giemsa stain: its history and applications. *International journal of surgical pathology*
473 **15**, 292–296 (2007).

474 [40] Lam, S. S. *et al.* Directed evolution of apex2 for electron microscopy and proximity labeling. *Nature*
475 *methods* **12**, 51–54 (2015).

476 [41] Bekker, J. & Davis, J. Learning from positive and unlabeled data: a survey. *Mach. Learn.* **109**, 719–760
477 (2020).

478 [42] Christiansen, E. M. *et al.* In silico labeling: predicting fluorescent labels in unlabeled images. *Cell* **173**,
479 792–803 (2018).

480 [43] Caruana, R. Multitask learning. *Machine learning* **28**, 41–75 (1997).

481 [44] Girshick, R. Fast r-cnn. In *Proceedings of the IEEE international conference on computer vision*,
482 1440–1448 (2015).

483 [45] Ruder, S. An overview of multi-task learning in deep neural networks. *arXiv preprint arXiv:1706.05098*
484 (2017).

485 [46] Mathis, A. *et al.* Deeplabcut: markerless pose estimation of user-defined body parts with deep learning.
486 *Nature neuroscience* **21**, 1281–1289 (2018).

487 [47] He, K., Girshick, R. & Dollár, P. Rethinking imagenet pre-training. In *Proceedings of the IEEE*
488 *international conference on computer vision*, 4918–4927 (2019).

489 [48] Raghu, M., Zhang, C., Kleinberg, J. & Bengio, S. Transfusion: Understanding Transfer Learning for
490 Medical Imaging. In Wallach, H. *et al.* (eds.) *Advances in Neural Information Processing Systems 32*,
491 3347–3357 (Curran Associates, Inc., 2019).

492 [49] Eliceiri, K. W. *et al.* Biological imaging software tools. *Nature Methods* **9**, 697–710 (2012).

493 [50] Ouyang, W. *et al.* Analysis of the Human Protein Atlas Image Classification competition. *Nature*
494 *Methods* **16**, 1254–1261 (2019).

495 [51] Mazzara, G. P., Velthuizen, R. P., Pearlman, J. L., Greenberg, H. M. & Wagner, H. Brain tumor
496 target volume determination for radiation treatment planning through automated MRI segmentation.
497 *International Journal of Radiation Oncology, Biology, Physics* **59**, 300–312 (2004).

498 [52] Hotelling, H. Analysis of a complex of statistical variables into principal components. *Journal of*
499 *educational psychology* **24**, 417 (1933).

500 [53] Paszke, A. *et al.* Automatic differentiation in pytorch. In *31st Conference on Neural Information*
501 *Processing Systems* (2017).

502 [54] Kingma, D. P. & Ba, J. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*
503 (2014).

504 [55] Cook, R. L. Stochastic sampling in computer graphics. *ACM Transactions on Graphics (TOG)* **5**, 51–72
505 (1986).

506 [56] Van der Walt, S. *et al.* scikit-image: image processing in python. *PeerJ* **2**, e453 (2014).

507 [57] Kuhn, H. W. The hungarian method for the assignment problem. *Naval research logistics quarterly* **2**,
508 83–97 (1955).

509 [58] Yeghiazaryan, V. & Voiculescu, I. D. Family of boundary overlap metrics for the evaluation of medical
510 image segmentation. *Journal of Medical Imaging* **5**, 015006 (2018).

511 [59] Scott, M. M. *et al.* A genetic approach to access serotonin neurons for in vivo and in vitro studies.
512 *Proceedings of the National Academy of Sciences* **102**, 16472–16477 (2005).

513 [60] Good, P. I. *Resampling Methods* (Birkhäuser Basel, 2006), 3 edn.

514 4 Methods

515 4.1 MICRA-Net

516 4.1.1 Architecture

517 Figure 2a shows the schematic representation of the MICRA-Net architecture. MICRA-Net is based on
518 the encoder part of a U-Net [31]. The rationale is that U-Net is an established method to solve different
519 analysis tasks (e.g. segmentation, localization, detection) on biomedical datasets. Each depth of the network
520 contains two blocks of convolutions (kernel size of 3) followed by batch normalization, and ReLU activation.
521 The number of filters in the convolutional layers is doubled after maxpooling (stride and kernel size of 2) to
522 increase the richness of the representation. The number of filters for each layer is {32, 64, 128, 256}. Global
523 maxpooling on the output layer allows a reduction of the dimensionality and a fully connected layer (FCL)
524 is used to provide a classification prediction. Dropout (probability of 0.5) is applied on the input features of
525 the FCL.

526 At inference, MICRA-Net predicts a whole image target from a given sample. Then, from each activated
527 class c , a local map L^l is calculated from the weighted combination of the activation map $A^{l,k}$ and the
528 mean gradient $\alpha_{l,k}^c$ of each l layer [28]. The mean gradient $\alpha_{l,k}^c$ is calculated from the backpropagated class
529 activation y^c

$$\alpha_{l,k}^c = \frac{1}{Z} \sum_i \sum_j \underbrace{\frac{\partial y^c}{\partial A_{i,j}^{l,k}}}_{\text{gradients via backprop}} . \quad (1)$$

530 The local map L^l is calculated as the linear combination of the activation map and the mean gradient of
531 each layer of convolutions in the network

$$L^l = \sum_k \alpha_{l,k}^c A^{l,k} . \quad (2)$$

532 Since MICRA-Net produced spatially reduced feature maps, local maps were upsampled using nearest
533 neighbor interpolation to match the input image size of 256×256 pixel. These images were then normalized in
534 the range [0, 1] using a min-max scaling. ReLU activation is applied on the last layer (L^8) of the network, as
535 in the seminal implementation of Grad-CAM [28], to be used for the coarse segmentation. Local maps from
536 layers L^{1-7} (Figure 2a-c) were concatenated into a feature space and retrieved the first principal component
537 of every pixel using principal component analysis (PCA) [52] decomposition to retain prominent information
538 from the feature space. The network was built and trained with the PyTorch library [53].

539 To facilitate the analysis of new images using MICRA-Net, a graphical user interface (GUI) is provided
540 to qualitatively analyse the influence of each local map (Supplementary Fig. 30). While the implementation
541 of MICRA-Net uses layers L^{1-7} with a PCA decomposition of the resultant feature space, the GUI allows to
542 arbitrarily combine different local maps of the MICRA-Net architecture and threshold the resultant detailed
543 feature map.

544 4.1.2 Training procedure

545 The general training procedure of the MICRA-Net architecture are reported within this section. For specific
546 training details for each dataset, see Supplementary Notes 1-5. MICRA-Net was trained using the Adam
547 optimizer with a learning rate specific to each dataset and other default parameters [54]. A learning rate
548 scheduler was used to reduce the learning rate of the optimizer with a minimal possible learning rate of
549 1×10^{-5} . The number of training epochs was adapted to the specific dataset (Supplementary Tab. 22-26).
550 Early stopping was used to reduce overfitting. Unless otherwise specified, we used binary cross entropy with
551 logits loss. We kept the model with the best generalization properties on the validation set (calculated from
552 the objective loss function).

553 Data augmentation was used to increase the performance of the network. Refer to Supplementary Tab. 22-
554 26 for a detailed data augmentation procedure for each dataset. All operations were applied in a random
555 order with a probability of 50%.

556 4.1.3 Auxiliary tasks

557 This section presents how MICRA-Net can be used to solve the common auxiliary tasks in microscopy images.

558 **Classification.** The classification task is used on all presented dataset in the paper. It serves as a
559 guideline to validate the representation capability of MICRA-Net. The classification task is solved by design
560 using MICRA-Net since it is trained using a classification task. The prediction from MICRA-Net are mapped
561 in the $[0, 1]$ range using a sigmoid function.

562 **Semantic segmentation.** The semantic segmentation task is solved on all presented dataset in the
563 paper. This task is solved by first extracting a detailed semantic feature map as described in Section 4.1.1.
564 The semantic segmentation masks are obtained by thresholding the resultant semantic feature map using
565 common thresholding algorithm (e.g. Otsu or percentile thresholding). The dataset specific thresholding is
566 detailed in Supplementary Notes 1-5.

567 **Detection.** The detection task on the P. Vivax and EM microscopy dataset is solved by predicting the
568 probability of presence of an object on all extracted crops. The overlap between the crops is of 75% in both
569 directions. Overlapping crops are averaged and reassigned to an output feature map of the same shape as
570 the image. The detection threshold is inferred from the validation set using a precision-recall curve.

571 **Semantic instance segmentation.** The semantic instance segmentation task is required on the Cell
572 Tracking Challenge dataset. MICRA-Net is required to predict i) the presence of an object and ii) the contact
573 between objects. The grad-CAMs of the activated objects are extracted from the architecture and combined
574 using a principal component analysis (PCA) as presented in Section 4.1.1. If a contact is predicted on an
575 image, the grad-CAM from L^8 which contains the prominent information of the contact is extracted. The
576 contact feature map is subtracted from the object feature map as in some fully-supervised techniques [27].
577 An Otsu threshold is used to generate the semantic segmentation masks of the instances.

578 4.2 Datasets

579 4.2.1 Modified MNIST dataset

580 We generated the modified MNIST training dataset by randomly sampling N digits from the original MNIST
581 training dataset and randomly distributed them on a 256×256 pixel field of view. To avoid overlap between
582 digits we used a random Poisson disc sampling algorithm with a radius size of 25 pixels [55]. The number of
583 digits N was uniformly sampled from $\{1, 2, 3, 4, 5, 10, 15, 20, \text{Max}\}$, where Max corresponds to the maximum
584 number of digits that can be placed without overlap. A rotation of $\pm 30^\circ$ uniformly sampled was applied
585 to the digits before placement on the image. We applied, in a random order, a Gaussian blur with sigma
586 uniformly sampled in $[0, 2]$ and artificial normalized Poisson noise with $\lambda = \frac{\sqrt{255}}{2}$. The resulting image
587 intensities were clipped to lie in $[0, 1]$. Using this technique, we generated 2000 and 1000 images for training
588 and validation respectively.

589 The modified MNIST testing dataset consists of 1000 images of handwritten digits sampled from the
590 original MNIST testing dataset. As for the training dataset, we also applied, in a random order, Gaussian
591 blur and artificial normalized Poisson noise sampled as before.

592 4.2.2 F-actin dataset

593 The F-actin dataset was generated by using a sliding window of size 256×256 pixel with a stride of 192
594 pixels over 260 complete images with an approximate size of 1000×1000 pixel. Since the super-resolution
595 microscopy images used are mostly composed of background, we set out to keep the crops containing at least
596 10% of dendritic area thereby reducing the number of crops to identify. The dendritic mask was obtained
597 from the foreground detection on the confocal imaged of the dendritic marker MAP2 using a global Otsu
598 thresholding on the normalized Gaussian blurred image [2, 35]. The sigma parameter of the Gaussian blur
599 was set to 20 pixels as it provided suitable dendrite detection over a wide range of images. We next annotated
600 each generated crop as being positive to the presence of the F-actin periodical lattice or longitudinal fibers.
601 The resulting training dataset contained 3832 crops (256×256 pixel, 897 images positive to the periodical
602 lattice and 1456 positive to the longitudinal fibers), the validation dataset contained 1287 crops (405 positive
603 to periodical lattice and 377 positive to fibers), and the testing dataset contained 416 crops (83 positive to
604 periodical lattice and 132 positive to fibers). The images were rescaled to lie in the $[0, 1]$ interval. The maximum value for scaling (**max**) was obtained by sampling the maximal value of all training images from
605 which we calculated the median in addition to 3 standard deviation. The minimum value was calculated as
606 the median of minimas (**min**). To ensure a proper scaling of the images we also added a scaling factor of 0.8
607

$$x' = \frac{x - \min}{0.8(\max - \min)}. \quad (3)$$

608 To evaluate the segmentation performance of the trained models, an Expert precisely highlighted the
609 contours of the structures in 50 images (25 images positive to periodical lattice and 25 images positive to
610 fibers) randomly sampled from the testing set. This small segmentation dataset only served to compare
611 the segmentation performance from the MICRA-Net, weakly-supervised baselines (U-Net, Mask-R-CNN,
612 Ilastik), and User-Study.

613 4.2.3 Cell Tracking Challenge dataset

614 We selected 6 cell line datasets from the Cell Tracking Challenge (CTC) [8]: the DIC-C2DH-HeLa dataset
615 which was acquired using differential interferometry contrast microscopy, three non-synthetic fluorescence
616 microscopy datasets (Fluo-C2DL-MSC, Fluo-N2DH-GOWT1, and Fluo-N2DL-HeLa) and two phase contrast
617 microscopy datasets (PhC-C2DH-U373, and PhC-C2DL-PSC). All original images were rescaled in the $[0,$
618 $1]$ range using a per image min-max scale. We then resized each image and associated precise annotations
619 according to the specific needs using bi-linear interpolation and nearest neighbors respectively with the
620 **Scikit-Image** [56] Python library (Supplementary Table 6 for scaling factors). We used a sliding window of
621 size 128×128 pixel or 256×256 pixel with a 25% overlap between crops in both directions. Using this sliding
622 window technique yielded a total of 27,106 positive crops and 3,364 negative crops for the 256×256 pixel
623 crops resized to have an effective pixel size of $0.5 \mu\text{m}$. The sliding window with size 128×128 pixel crops
624 and resized to have single cells in the field of view yielded a total of 66,466 positive crops (20,724 positive
625 to contact) and 88,722 negative crops for training and 17,621 positive crops (5,606 positive to contact) and
626 22,279 negative crops for validation. We simulated weak annotations from the precise contours of the cells
627 provided in the original CTC dataset by identifying an image crop as positive if the corresponding annotated
628 crop contained at least the size of the average annotated cell, and negative otherwise. To evaluate the
629 segmentation and detection tasks, we manually segmented 4 images randomly sampled per cell line in the
630 testing set.

631 4.2.4 P. Vivax dataset

632 We used image set BBBC041v1, available from the Broad Bioimage Benchmark Collection [33]. The complete
633 dataset contained 1327 3-channel images and was already split into a training (1207 images) and testing
634 (120 images) set. The dataset is composed of blood smears that were stained with Giemsa reagent [39]
635 and acquired on three different brightfield microscopes from three different laboratories. All blood smears
636 (infected or uninfected) were annotated using bounding boxes. The blood smears were later classified as
637 infected (gametocytes, rings, trophozoites, and schizonts) or uninfected (red blood cells, and leukocytes)

638 by an Expert. The task was to differentiate infected from uninfected blood smears. The dataset is highly
639 unbalanced towards red blood cells which composes over 95% of the annotated cells.

640 For training and testing, we applied a whitening normalization (null mean and standard deviation of 1)
641 to each image (and channel) to minimize the impact of a very different intensity distribution. The binary
642 targets for training were generated using the provided bounding boxes. A crop was considered as positive if
643 it contained at least 5% of overlap with an infected cell, otherwise as negative. The crops were 256×256
644 pixel.

645 We manually extracted and precisely annotated all infected cells in the testing set resulting in 303 small
646 crops of size 256×256 pixel centered on the cell of interest.

647 4.2.5 Scanning Electron Microscopy dataset

648 The dataset contained 92 images of $10,240 \times 10,240$ pixel for training, 66 for validation, and 44 for testing.
649 An Expert annotated the images using positional markers to locate the Axon DAB markers. On average the
650 large fields of view contained 3 small detections (113×113 pixel, between 1 and 10 detections per image).
651 This resulted in an annotation time of approximately 30 minutes per field of view. Training and inference
652 was performed on 512×512 pixel size crops. The dataset contained all positive crops (1024×1024 pixel,
653 centered on the Axon DAB markers), and all negative crops (without overlap). To manually annotate the
654 images the Expert inverted the acquired images. Hence, we provided MICRA-Net with the inverted image
655 to mimic the Expert task. We rescaled the provided 8-bit depth images in the $[0, 1]$ range by dividing by a
656 scalar value of 255.

657 All Axon DAB markers were extracted from the testing set (170 positive markers) and an Expert carefully
658 identified their contours.

659 4.3 Evaluation procedure

660 4.3.1 Classification

661 The classification accuracy of MICRA-Net was evaluated by inferring the testing images. To quantitatively
662 assess the performances, the classification accuracy was calculated for each trained model. We reported the
663 mean \pm standard deviation of the trained models.

664 4.3.2 Detection

665 The centroid of each detected object was obtained from MICRA-Net by using the dataset specific procedures
666 detailed in Supplementary Notes 1-5. Each detected centroid was associated with the centroid of objects in
667 the ground truth mask using the Hungarian algorithm [57] with a maximal distance of N pixels, where N
668 is approximately the object radius. In this context, an associated detected object is considered as a true
669 positive, a non-associated detected object is a false positive, and a missed ground truth object is a false
670 negative. To evaluate the detection capability of MICRA-Net, we reported the F1-score. For a quantitative
671 comparison, we repeated the evaluation for each trained model. We then bootstrapped the average of the
672 trained models to show the bootstrapped mean and 95% confidence interval (10 000 repetitions).

673 4.3.3 Segmentation

674 The segmentation performance of the trained models was evaluated using three common evaluation metrics:
675 F1-score, Intersection Over Union (IOU), and the Symmetric Boundary Dice (SBD) [58]. If multiple instances
676 of a model were trained on the same task, we bootstrapped the average of the trained models to show the
677 bootstrapped mean and 95% confidence interval (10 000 repetitions).

678 4.3.4 Instance segmentation

679 Prior to evaluation, we removed small objects ($<20 \times 20$ pixels) from the segmentation mask and filled holes
680 for all trained models. All segmentation masks were resized to the baseline scale (Supplementary Table 6)
681 for proper comparison. The instance segmentation performance were evaluated using the method proposed
682 by [27] (Supplementary Figures 19-22). Briefly, this method evaluates the detection and failures of the

683 architecture dependant on the IOU. [27] used a minimal IOU of 0.5 to avoid multiple predicted objects to
684 be associated with a ground truth object. The goal is to maximize the F1-score vs. IOU, while the failure
685 modes should be minimized. We on the other hand solved the association between the ground truth and
686 predicted objects using the Hungarian algorithm [57], which allowed to report the performance and failure
687 modes across the entire range of IOU. Using a broader range of IOU allows to report the performance
688 in instance detection and segmentation. The normalized area under the resultant curves for each trained
689 model is bootstrapped to obtain the mean and 95% confidence interval (10 000 repetition) and is reported
690 in Figure 4.

691 4.3.5 Custom performance metrics

692 The F-actin periodical lattice is detected as an oscillating pattern between high- and low-intensity stripes
693 with 180-190 nm periodicity [32]. We designed a metric that would take this periodicity into account to
694 evaluate the MICRA-Net detailed segmentation performance. We computed, as a baseline, the Fourier
695 transform (FT) of the original image (FT_b) and the FT of the segmented regions: for the Expert (FT_e),
696 and for the predicted segmentation masks (FT_{pred}). The variation from the baseline was computed as the
697 difference in the FT spectrum, for spatial frequencies in the range [170, 200] nm, between $FT_{e,pred}$ and FT_b
698 over the sum of FT_b . A smaller absolute difference between the variation of the Expert and the variation of
699 the predicted mask implies more similar segmentation.

700 Since F-actin fibers are contiguous and have a high intensity on the dendrites, we designed a metric that
701 would use the distribution of pixels under a segmented mask. The rational behind this metric is that the
702 F-actin nanostructures on dendrites are composed of both high- and low-intensity pixels. Since F-actin fibers
703 have high intensities, a detailed segmentation of fibers would imply few low intensity pixels annotated, while
704 a coarse segmentation would introduce more low-intensity identified pixels. Hence, we considered a pixel
705 within the segmentation mask as part of a fiber if its value was superior to a given threshold. We calculated
706 this threshold by first measuring the 25th percentile of pixel intensities outside of the Expert mask for all
707 images. We then extracted the 90th percentile intensity values from all images containing F-actin fibers.
708 This resulted in a threshold between high- and low-intensity pixels within the dendritic mask of 9.

709 4.4 User-Study

710 We conducted two different User-Study in this paper, one for the F-actin nanostructure segmentation and one
711 for the instance segmentation on the Cell Tracking Challenge. All participants were familiar with bio-medical
712 images.

713 4.4.1 F-Actin segmentation

714 We performed a User-Study in which six participants highlighted the contours of the F-actin periodical
715 lattice and longitudinal fibres on a small dataset of 50 images using polygonal bounding boxes. We used
716 polygonal bounding boxes as this annotation method reduces the time required by a participant by more
717 than 3 folds compared to precisely identifying the boundaries of the structures (Supplementary Fig. 11). We
718 used our own annotation application that was optimized for this type of task. Annotation of the full dataset
719 required approximately 40 minutes for the participants. The averaged performance of the six participants
720 was compared to MICRA-Net using F1-score, IOU, and SBD.

721 4.4.2 Cell Tracking Challenge instance segmentation

722 A User-Study was conducted using the Cell Tracking Challenge to analyse the required time per cells and
723 the achievable performance of inter-participant annotation for such task. The User-Study consisted in the
724 annotation the 24 testing image using different level of supervision (precise, bounding boxes, and points).
725 For each level of supervision, the participants were asked to annotate a quarter of the testing image, which
726 was the same for all participants. The image intensity scale was set at a constant value for all participants.
727 The participants used the Fiji software to annotate the images. The median of the participant scores on
728 the testing set are reported, as well as the inter-participant scores. The time required by the participant to
729 annotate each image was recorded, which allowed to calculate the time per cell for each cell-line.

730 4.5 In-house datasets acquisition

731 4.5.1 Cell culture, Immunostaining and STED imaging for F-actin imaging

732 Before dissection of hippocampi, neonatal Sprague Dawley rats were sacrificed by decapitation, in accordance
733 to the procedures approved by the animal care committee of Université Laval. Dissociated cells were plated
734 on poly-d-lysine coated glass coverslips, fixed and immunostained as described previously [2]. F-Actin was
735 stained with Phalloidin-STAR635 (Abberior GmbH, Germany). Dendrites Microtubule-Associated-Protein
736 (MAP2) [2]. STED images of the F-Actin nanostructures were acquired on a 4 color Abberior Expert-Line
737 STED microscope (Abberior Instruments GmbH, Germany), equiped with a 100x 1.4 NA oil objective and
738 using pulsed (40 MHz) excitation (640 nm) and depletion (775 nm) lasers. Fluorescence was detected with
739 an Avalanche Photodiode (APD) and a ET685/70 (Chroma, USA) fluorescence filter. Pixel size was set to
740 20 nm.

741 4.5.2 Animals and stereotaxic injections for scanning electron microscopy dataset

742 This study was carried out on 3-month-old mice, weighing 25-35g. Animals were housed under a 12h light-
743 dark cycle with water and food ad libitum. All procedures were approved by the *Comité de Protection des*
744 *Animaux de l'Université Laval*, in accordance with the Canadian Council on Animal Care's Guide to the Care
745 and Use of Experimental Animals (Ed2), and with the ARRIVE guidelines. Maximum efforts were made
746 to minimize the number of animals used. Transgenic e-Pet Cre mice expressing Cre recombinase under the
747 control of Fev promoter, known to be specific for serotonin (5-HT) neurons [59], were injected in the dorsal
748 raphe nucleus (DRN) with 1 μ l of AAV9-CAG-DIO-APEX2NES-WPRE. Stereotaxic injections were done
749 using a 30° angle along the frontal plane at AP: -4.78; ML: +2.00 and DV: -3.20. In these injected transgenic
750 mice, the small engineered peroxidase APEX2 [40] is specifically expressed in the cytosol/cytoplasm of 5-HT-
751 infected neurons of the DRN and is used, in presence with hydrogen peroxide, to oxidize 3,3 Diaminobenzidine
752 (DAB) chromogen that can readily be visible at the light and electron microscope levels.

753 4.5.3 Tissue preparation for scanning electron microscopy dataset

754 After a period of 21 days following stereotaxic injection, mice were anesthetized with a mixture of ketamine
755 (100 mg/kg) and xylazine (10 mg/kg) and transcardially perfused with 50 ml of phosphate-buffered-saline
756 (PBS: 50 mM at pH 7.4) followed by 150 ml of 4% paraformaldehyde (PFA) and 1% glutaraldehyde diluted
757 in phosphate buffer (PB; 100 mM at pH 7.4). Brains were dissected out, post-fixed for 24h in the same fixative
758 solution and cut with a vibratome (model VT1200; Leica, Germany) into 50 μ m-thick frontal sections, which
759 were serially collected in sodium phosphate buffer saline (PBS, 100 mM, pH 7.4). Frontal brain sections at
760 the level of the subthalamic nucleus (STN) were processed to reveal the presence of APEX2 in axons arising
761 from DRN-infected neurons using 3,3'diaminobenzidine (DAB; catalog no. D5637; Sigma-Aldrich) as the
762 chromogen. Briefly, selected 50 μ m-thick sections were washed 3 times in PBS and then twice in Tris.
763 Sections were then incubated for 1h in 0.05% DAB solution diluted in Tris, then for 1h in 0.05% DAB
764 solution containing 0.015% hydrogen peroxide (H₂O₂). Sections were then rinsed twice in Tris and 3 times
765 in PBS. Sections were temporally mounted in PBS and coverslipped for light microscope examination. STN
766 sections containing DAB-labeled axons were selected for further processing. These sections were washed
767 3 times in PB, then incubated during 1h in 2% osmium tetroxide diluted in 1.5% potassium ferrocyanide
768 solution. They were then washed 3 times in ddH₂O, incubated for 20 min in 1% thiocarbohydrazide (TCH)
769 solution and washed again 3 times in ddH₂O. Sections were placed 30 min in 2% osmium tetroxide and
770 washed 3 times in ddH₂O. Sections were then dehydrated in ethanol and propylene oxide and flat-embedded
771 in Durcupan (Electron microscopy Science). Areas of interest were cut from embedded sections and glued
772 to the tip of resin blocks. Blocks were cut with an ultramicrotome (Leica EM UC7) in ultrathin sections
773 (80 nm), which were serially collected on silicon-coated 10 x 10 mm chip wafer (Ted Pella, Inc; #16006).

774 4.5.4 Scanning electron microscopy (SEM)

775 Serial sections were imaged in a SEM (Zeiss Gemini 540) with the help of the ATLAS acquisition software.
776 Images were acquired at a resolution of 5 nm/pixel, using acceleration voltage of 1.4 kV and current of 1.2 nA.
777 Serial sections acquisitions produced a stack of 38 rectangle images of 25370 x 25633 pixel (126.850 x 128.165

778 microns) taken out of 38 ultrathin sections. In addition, a large single section acquisition was acquired and
779 produced a single trapezoidal image of 31065 pixels for the small base (155.329 microns), 91393 pixel for the
780 large base (456.967 microns) and 53161 pixels for the height (265.809 microns). All acquired images were
781 subdivided into overlapping square tiles of 10240 \times 10240 pixel (51.2 x 51.2 microns).

782 4.6 Statistical assessment using resampling

783 Resampling was used as a statistical test to verify the statistical difference between two groups [60]. Statistical
784 analysis was performed using a randomization test with the null hypothesis being that the different conditions
785 (A, B) belong to the same distribution. The absolute difference between mean values of A and B was
786 calculated ($D_{gt} = |\mu_A - \mu_B|$). For the randomization test, each value belonging to A and B was randomly
787 reassigned to A' and B', with the sizes of A' and B' being N_A and N_B , respectively. The absolute difference
788 between the mean values of A' and B' was determined ($D_{rand} = |\mu_{A'} - \mu_{B'}|$) and the randomization test was
789 repeated 10 000 times. The obtained distribution was compared with the absolute difference of the mean of
790 A and B (D_{gt}) to verify the null hypothesis.

791 When the number of groups was greater than 2, the F-statistic was sampled from each group using a
792 resampling method. The F-statistic was calculated from all groups (A, B, C, etc.) as a ground truth (F_{gt}).
793 Each value was randomly re-assigned to new groups (A', B', C', etc.) where group X' has the same size
794 as group X. The F-statistic of newly formed groups (F_{rand}) was calculated and this process was repeated
795 10 000 times. We compared F_{rand} with F_{gt} to confirm the null hypothesis that the groups have the same
796 mean distribution. When the null hypothesis was rejected, *i.e.* at least one group did not have the same
797 mean distribution, we compared each group in a one-to-one manner using the randomization test described
798 above. In all cases, a confidence level of 0.05 was used to reject the null hypothesis. Since the precision of
799 the calculation of the *p*-value is limited to $\frac{1}{N}$, where N in the number of repetitions, we report a *p*-value of
800 $<1.0000 \times 10^{-4}$ instead of 0.

801 4.7 Evaluation of required decisions and time for fully-supervised training

802 *F-actin*: The number of decisions for a fully-supervised training dataset was estimated as the mean number of
803 edge pixels in the 50 precisely annotated images multiplied by the total number of positive crops. The mean
804 annotation time per crop was calculated using the precisely annotated dataset. *Cell Tracking Challenge*:
805 The mean image annotation time of 900 seconds was obtained from the precise annotation of each image of
806 the testing set. *P. Vivax*: The annotation time for fully-supervised annotations was estimated at 2 minutes
807 per image from the precise annotation of 10 images. *Electron Microscopy*: The required annotation time
808 was calculated as the average time required by the Expert per image (30 minutes per image, 156 images)
809 to detect all axon DAB markers. We added 14 seconds (calculated from highlighting the contours of the
810 Axon DAB regions on the testing set) for each positive detection (537 detections) to account for precise
811 annotation.