

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

Flavie Lavoie-Cardinal (✉ flavie.lavoie-cardinal@cervo.ulaval.ca)

Université Laval

Anthony Bilodeau

CERVO Brain research center

Constantin Delmas

CERVO Brain research center

Martin Parent

CERVO Brain research center

Paul De Koninck

CERVO Brain Research Center <https://orcid.org/0000-0002-6436-1062>

Audrey Durand

Université Laval

Article

Keywords: High Throughput Quantitative Analysis, Image-level Annotations, Gradient Class Activation Maps, Expert Annotation Process

Posted Date: October 29th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-95613/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

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

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

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

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

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

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

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

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

Abstract

High throughput quantitative analysis of microscopy images presents a challenge due to the complexity of the image content and the difficulty to retrieve precisely annotated datasets. In this paper we introduce a weakly-supervised MICROscopy Analysis neural network (MICRA-Net) that can be trained on a simple main classification task using image-level annotations to solve multiple more complex auxiliary tasks, such as segmentation, detection, and enumeration. MICRA-Net relies on the latent information embedded within a trained model to achieve performances similar to state-of-the-art fully-supervised learning. This learnt information is extracted from the network using gradient class activation maps, which are combined to generate precise feature maps of the biological structures of interest. We demonstrate how MICRA-Net significantly alleviates the expert annotation process on various microscopy datasets and can be used for high-throughput quantitative analysis of microscopy images.

1 Introduction

The development of powerful microscopy techniques that allows to characterize biological structures with sub-cellular resolution and on large field of view tremendously increased the complexity of quantitative image analysis tasks [1]. The resulting images exhibit a wide range of structures that need to be identified, counted, precisely located, and segmented. Expert knowledge is commonly required to achieve successful identification and segmentation of the multiple structures of interest in microscopy images [2, 3]. These tasks can be tedious and time consuming especially for large databanks or for the comparison of multiple biological conditions. It was recently demonstrated that deep convolutional neural networks (CNN) are excellent feature extractors [4]. They were successfully applied to segmentation (e.g. whole cells, nuclei, dendritic spines), enumeration (e.g. cell counting), and classification (e.g. state of cell) of structures in microscopy images [5–9]. The most common deep learning (DL) approaches applied to microscopy and biomedical images are fully-supervised and require precisely annotated datasets. Hence, it is often a limiting step in the application of DL for quantitative analysis of biomedical imaging [3, 10, 11]. To alleviate the annotation process, weakly-supervised DL methods were introduced [11–14]. Bounding box annotations are commonly used for weakly-supervised segmentation tasks as they are simple, allow the task to be spatially

constrained [2, 13, 15–17], and were shown to decrease the annotation phase by 15-fold compared to precise identification of structure boundaries [18]. Methods for training with binary, image-level targets, reducing even further the complexity and duration of the annotation task, have been proposed when multiple instances are displayed on a single image [19]. Unfortunately, when applied to microscopy and biomedical image analysis, such weakly-supervised approaches using whole image annotations resulted in lower segmentation precision compared to approaches using precisely identified structures [20–22].

In this paper we propose MICRA-Net (MICROscopy Analysis Neural Network), a new approach relying only on image-level classification annotations for training that achieves performances similar to fully-supervised methods on multiple microscopy analysis tasks. MICRA-Net builds on *latent learning* [23], which refers to a model retaining information (*i.e.* latent space) that is not required for the task at hand in order to learn new auxiliary complementary tasks [23]. In this work, we leverage the information embedded within a trained classification network to solve multiple complementary, yet very different, tasks relevant to microscopy image analysis. The network uses binary classification targets as input to build a general representation of the specific dataset and generates precise feature maps from which specific tasks such as segmentation, detection, and classification can be addressed. Even further this showcases the potential of MICRA-Net for addressing various high-throughput microscopy analysis challenges, relying solely on weak image-level annotations for training.

2 Results

2.1 Multi-class segmentation of synthetic images

The generation of precisely annotated large datasets to train deep neural networks remains a challenge in the field of microscopy and biomedical imaging. MICRA-Net, a CNN-based method, addresses this challenge by using solely whole-image binary targets for training. The network was trained on a simple multi-class classification task and we evaluated its capacity to perform diverse auxiliary tasks (e.g. segmentation, enumeration, detection) relevant to high-throughput microscopy image analysis. Generation of an annotated training dataset for each task would have required complex and time-consuming annotations such as precise contours or localization of the biological structures. The training dataset for MICRA-Net requires the expert to identify class-specific positive and negative images with respect to the structures of interest. In contrast to the identification of the structure boundaries using precise or bounding box contours, image-level annotations do not require to specify the positions of the object in the field of view of the microscopy images (Figure 1a). This annotation scheme thereby reduces both the expert interactions and annotation times required for the generation of the training dataset (Figure 1b).

Figure 2a shows the architecture of MICRA-Net, which was designed around a CNN architecture, composed of 8 convolutional layers (L^1 to L^8) followed by a fully connected layer. The gradient class activated maps (Grad-CAM, see Methods) were extracted for each predicted class and at every layer of the network (Figure 2a,b). Thereafter, Rectified Linear Unit (ReLU) activation and thresholding on the Grad-CAM of the last convolutional layer (L^8) were applied to generate a coarse class-specific feature map [24]. To increase the information contained in the extracted feature map, local maps from layers L^{1-7} were concatenated, resulting in a class-specific 7-dimensions feature space (Figure 2b,c). We retrieved the first principal component of every pixel using principal component analysis (PCA) decomposition on the feature space to generate a single feature map that was used to solve different sets of specific auxiliary tasks (Figure 2c & Methods).

To characterize the performance of MICRA-Net for classification and segmentation tasks, we used a synthetic dataset containing N randomly sampled cluttered handwritten digits from the MNIST dataset [25] (Modified MNIST dataset, Figure 2c & Methods). Each image may contain several instances of digits (from 0 to 9), as well as variable levels of noise and signal to mimic variations observed in microscopy images. Our first step was to classify the digits appearing on each image to validate the representation capability of the network, which is confirmed by the obtained class-wise mean classification testing accuracy of $(98.9 \pm 0.5) \%$ (mean \pm std).

In addition to the classification task, MICRA-Net generates class-specific segmentation maps of the digits in the modified MNIST dataset. Using the information embedded in the Grad-CAMs of the hidden layers (L^{1-7}), to precisely locate each digit in the image, significantly increased the segmentation performance of the

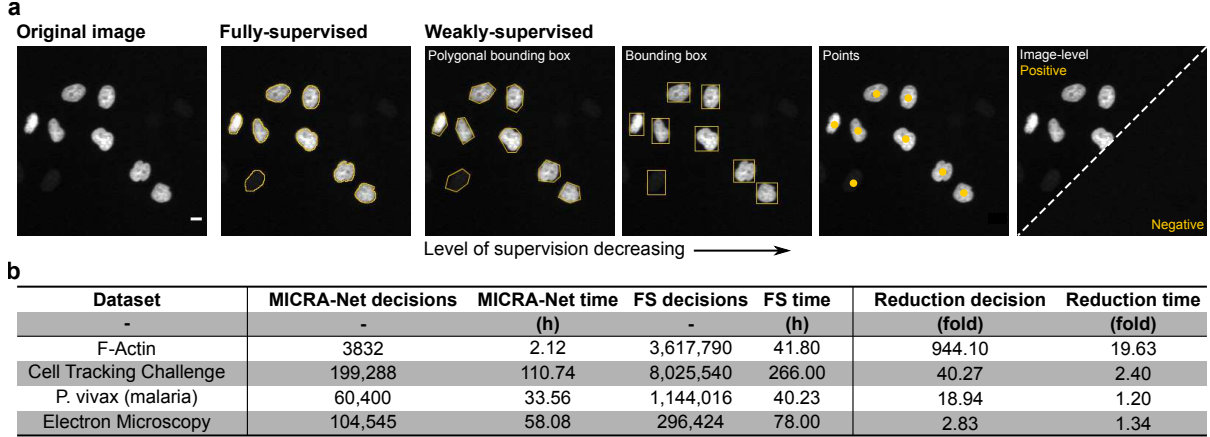


Figure 1: Various supervision levels can be employed for training a DL model to segment structures of interest in microscopy images. a) Representative image from the Cell Tracking Challenge dataset [8] overlaid with the corresponding fully- and weakly-supervised annotations. Annotated images are presented in decreasing spatial level of supervision and required annotation time (*from left to right*). b) 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 FS 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 (assignment of a positive or negative annotation) was on average 2 seconds for all datasets. For FS learning, the decision and annotation time for each dataset separately on a precisely annotated subset of images was evaluated (see Methods).

network when compared to the maps obtained from the Grad-CAMs of the last layer only (L^8) (Figure 2e & Supplementary Fig. 1. A U-Net [26] trained on the same dataset using a fully- and weakly-supervised training scheme was used as a baseline to better evaluate the performance of MICRA-Net. Fully-supervised learning consisted in training with the binary digits contours from MNIST, while weak contours were generated by a dilation of the digits with a square of size $\{5, 10, 25\}$ pixel as a structuring element (see Supplementary Note 1). Figure 2f shows that MICRA-Net achieves similar or superior segmentation performance compared to all weakly-supervised training instances of the U-Net and is only outperformed by fully-supervised training (Supplementary Fig. 2 & Supplementary Tab. 1).

2.2 Class-specific segmentation of super-resolution microscopy images

The next question that needed to be addressed was the applicability of our approach for super-resolution microscopy image segmentation, for which precisely annotated datasets are rarely available. The auxiliary task was the semantic segmentation of STimulated Emission Depletion (STED) microscopy images of two nanostructures of the F-actin cytoskeleton in neurons: 1) a periodical lattice structure and 2) longitudinal fibers (Figure 3a,b) [2]. The F-actin nanostructure segmentation task is challenging since the morphology of neurons is highly variable throughout the dataset, and there are many distractors around the structures of interest [2]. Figure 1 shows that image-level annotation reduced by more than 19 folds the time required by an expert to generate the training dataset compared to precise identification of the structure boundaries that would be required for fully-supervised DL approaches. This also corresponds to a reduction of the annotation time of more than 3 folds compared to the tracing of polygonal bounding boxes, which were recently used for weakly-supervised training of the U-Net architecture on this dataset [2].

On the main classification task, MICRA-Net achieves an accuracy of 75.2% and 83.7% on the testing dataset for the F-actin periodical lattice and longitudinal fibers, respectively. This is inline with a mean inter-expert classification accuracy of $(80 \pm 5) \%$ and $(75 \pm 7) \%$ for periodical lattice and longitudinal fibers respectively (calculated from 6 experts using a leave-one-out scheme from 50 images), confirming the model capability to handle data of this nature (Supplementary Fig. 3). As described in the previous section, an

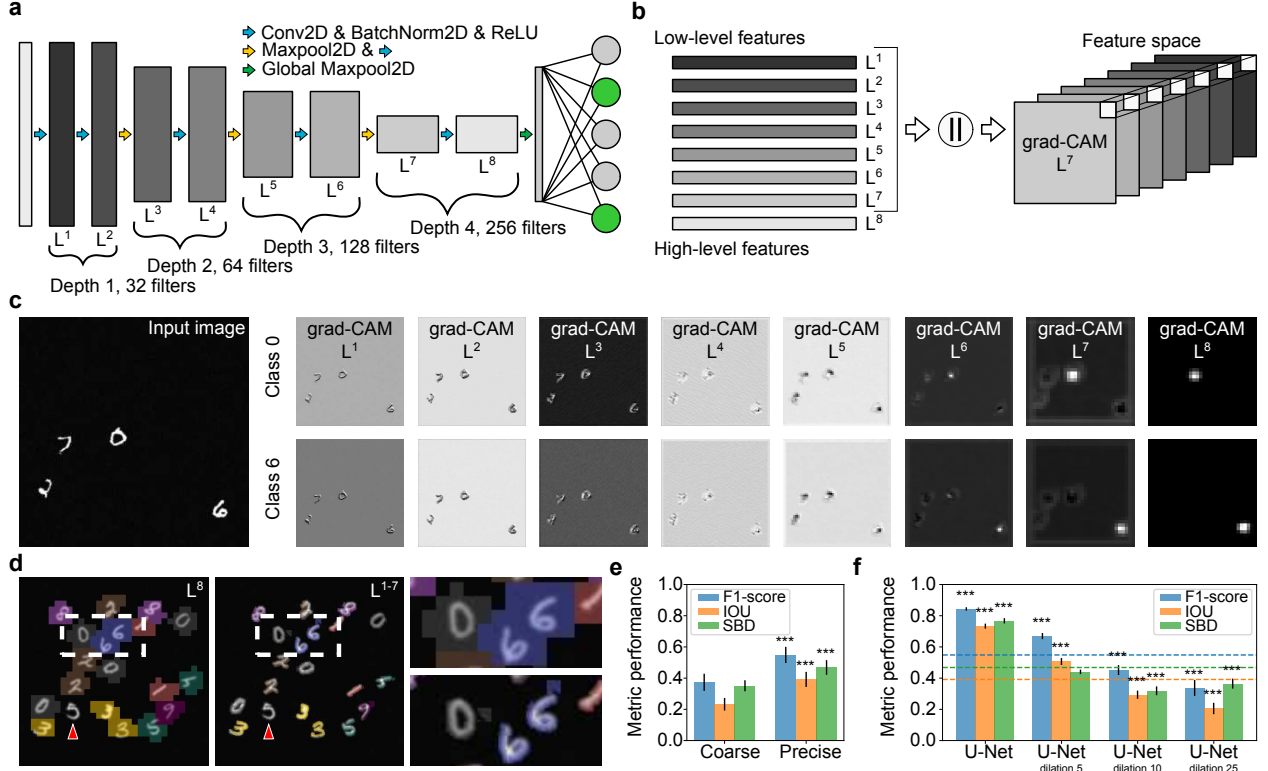


Figure 2: MICRA-Net architecture and experimental results on the modified MNIST dataset. a) MICRA-Net architecture (detailed in the Methods section). Each depth is composed of two sequential convolutional layers (Conv2D), batch normalization (BatchNorm2D), and Rectified Linear Unit activation (ReLU). 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. d) Coarse and precise segmentation maps of the digits of a representative image (256×256 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) Evaluation of the coarse and precise segmentation performance using the *i.e.* F1-score, intersection over union (IOU), and symmetric boundary dice (SBD) (see Methods). Reported here is the mean performance over the 10 classes (see Supplementary Fig. 1 for class-wise and density-wise performances). A significant increase in the segmentation performance is measured for precise over coarse segmentation (t-test, $p_{\text{F1-score}} = 1.6616 \times 10^{-6}$, $p_{\text{IOU}} = 5.2650 \times 10^{-7}$, $p_{\text{SBD}} = 1.1627 \times 10^{-5}$). f) 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. 2a. The MICRA-Net precise segmentation (color-coded dashed lines) surpasses the U-Net trained with 10 pixels dilation and is not statistically different from the U-Net trained with 5 pixels dilation. Only fully-supervised training outperforms MICRA-Net precise segmentation (one-way ANOVA followed by posthoc t-test, $p_{\text{F1-score}} = 2.0254 \times 10^{-12}$, $p_{\text{IOU}} = 8.7200 \times 10^{-14}$, $p_{\text{SBD}} = 5.7291 \times 10^{-13}$). All p-values are reported in Supplementary Tab. 1. Bar graphs show the mean values and standard deviation.

117 informative feature map was generated from the PCA decomposition of the combined L^{1-7} extracted features.
 118 Thresholding of this feature map resulted in precise binary masks that were used to solve the segmentation
 119 task. Relying on a *precisely annotated dataset* consisting of 25 images of each structure (Supplementary

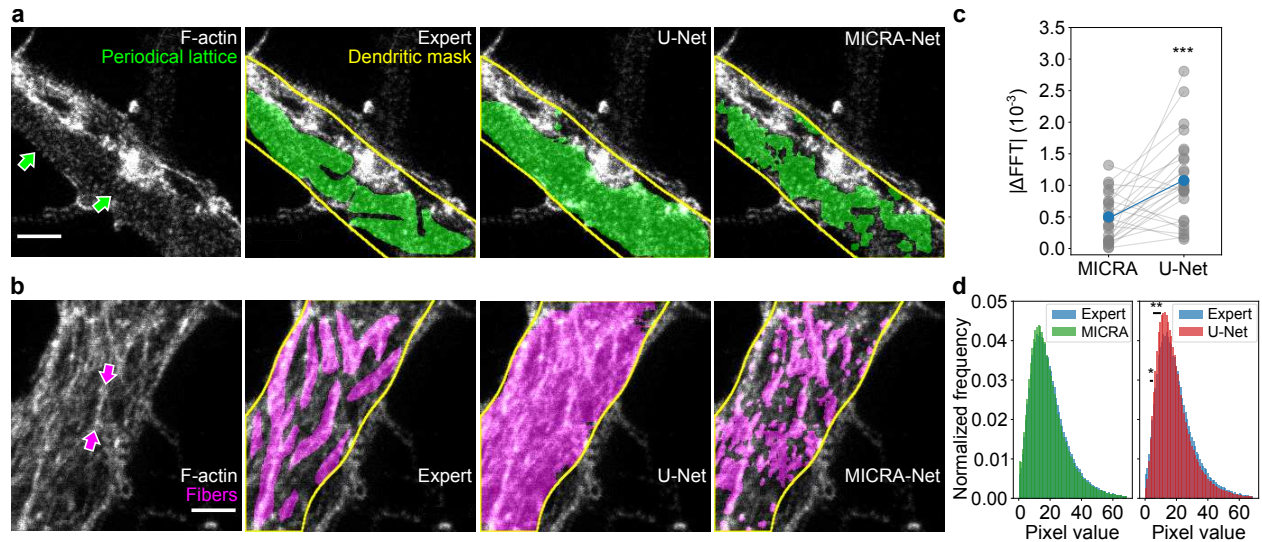


Figure 3: Semantic segmentation of F-actin nanostructures observed on super-resolution microscopy images. a,b) *Left*: Representative raw images from a dataset of STimulated Emission Depletion (STED) microscopy images of two F-actin nanostructures in fixed cultured hippocampal neurons: periodical lattice (a) and longitudinal fibers (b). Arrows point towards the periodical lattice (green) and longitudinal fibers (magenta). *Middle-Left*: precise expert contours of the nanostructure, *Middle-Right*: weakly-supervised U-Net segmentation masks and *Right* MICRA-Net segmentation masks. c, d) Performance evaluation of MICRA-Net and weakly-supervised U-Net segmentation on the precisely annotated testing dataset using custom metrics for rings (c) and fibers (d). The FFT (c) and intensity distribution (d) metrics evaluate the difference between the pixels found within the precise expert annotations and the two DL-based segmentation approaches (see Methods). An increased segmentation accuracy is measured with MICRA-Net compared to U-Net for rings (c) and fibers (d). No significant differences are measured between the precise expert annotations and MICRA-Net segmentation mask for both metrics. Statistical analysis: c) t-test, $p = 3.7342 \times 10^{-4}$, d) all p-values are shown in Supplementary Tab. 2. Performance evaluation was performed within the dendritic mask (a,b: yellow line). a,b) Scale bars: $1 \mu m$.

Fig. 4), each segmentation performance evaluation consisted of: i) MICRA-Net, ii) multi-expert bounding-box annotations (6 experts on 25 images of each structure: *user-study*), and iii) the U-Net trained with polygonal bounding boxes [2] (see Methods and Supplementary Note 2 for specific details). MICRA-Net achieved equivalent or superior segmentation performance on the *precisely annotated dataset* in comparison to both the *user-study* and the U-Net (Supplementary Figure 5). Thus, even if trained with weak image-level annotations, MICRA-Net can extract the necessary structural information to generate precise segmentation maps for both nanostructures.

A qualitative visual inspection of the segmentation masks suggested that MICRA-Net precise segmentation produced a finer detailed mask compared to the U-Net weakly-supervised segmentation [2], especially for fibers, for which it provides precise segmented contours of single fiber strains (Figure 3c, Fibers). Custom performance metrics that were adapted to the F-actin nanostructures were required to better characterize this observation. For the F-actin periodical lattice, we measured the Fourier Transform (FT) of the segmented areas for frequencies corresponding to the periodicity of the lattice (180-190 nm [27]) (Figure 3c & Methods). No significant difference with the *precisely annotated dataset* is measured with this FT-metric for MICRA-Net, while a significant increase is measured for the U-Net. This suggests an increased segmentation precision for our approach over U-Net segmentation. Similarly, a custom metric based on the pixel intensity distribution of the segmented areas was developed to evaluate the approaches on the fiber segmentation task. While no difference was observed for the regions identified with MICRA-Net compared to the regions from the *precisely annotated dataset*, a significant increase of the proportion of low-intensity pixels was obtained from the U-Net segmentation (Figure 3d, Supplementary Tab. 2 & Methods). This supports a higher accuracy to precisely identify the contours of individual fibers or periodical lattice regions of MICRA-Net over

weakly-supervised U-Net segmentation.

2.3 Single cell counting and segmentation

Cell counting and segmentation is a common challenge in high-throughput analysis of optical microscopy images [8, 9, 28]. Both fully- and weakly-supervised DL approaches were shown to be very powerful to assess these tasks on multiple cell lines [7, 22]. We characterized the performance of MICRA-Net for single cell detection and segmentation compared to a pre-trained fully-supervised U-Net [9] (Supplementary Fig. 7a). The first step consisted in validating the representation capability of the model on the classification task of 6 cultured cell lines of the Cell Tracking Challenge [8]. In contrast to the seminal implementation of Falk et al. [9], the U-Net was designed to output class-specific segmentation, hence comparing similar tasks between U-Net and MICRA-Net (see Supplementary Note 3). Both U-Net and MICRA-Net (see Methods) were trained with 256×256 pixel crops from the resampled images (with an effective pixel size of $0.5 \mu\text{m}$). MICRA-Net enabled the classification of the different cell lines with a testing accuracy of $(95.8 \pm 0.4) \%$ (calculated from 5 network instances) compared with $(96 \pm 1) \%$ for U-Net. Next, we aimed at solving the auxiliary tasks, *i.e.* single cell detection and segmentation. As expected for weakly-supervised learning from image-level targets, a sufficient amount of negative samples (images not containing the object of interest) is required to extract informative context from an image, *i.e.* to distinguish the cells in the field of view. Accordingly, MICRA-Net detection and segmentation performances are strongly reduced when no negative samples are provided (Supplementary Fig. 7b, DIC-C2DH-HeLa and Fluo-N2DH-GOWT1) despite having a high classification accuracy. It is therefore necessary to adapt the size of the training images that are provided to the network to the size of the structures of interest (Figure 4a), ensuring that enough images contain only background. For each cell type of the Cell Tracking Challenge, we resized the images in the dataset so that a single cell would be approximately half the size of a 128×128 pixel region (see Supplementary Tab. 3 for scaling factors).

Next, we compared the semantic segmentation and detection efficiency of MICRA-Net with the pre-trained fully-supervised U-Net from Falk et al. [9]. To obtain the MICRA-Net segmentation masks, it was necessary to threshold the resultant feature maps (dependent on the cell line) and apply a watershed transformation to separate connected cells (see Supplementary Note 3 for specific details). Since no precisely annotated testing dataset was provided for the *Cell Tracking Challenge dataset*, precise annotations of 4 images were generated for each cell line to evaluate the segmentation performance of both approaches. MICRA-Net and U-Net achieved comparable IOU for 3 cell lines for segmentation, while for the remaining cell lines the U-Net performance was superior (Figure 4b,d & Supplementary Tab. 4). In the detection task, MICRA-Net outperformed or is equivalent to the U-Net approach for 5 out of 6 cell lines (Figure 4c,d & Supplementary Tab. 5). Most importantly, MICRA-Net allowed the reduction by a factor of 40 the number of expert decisions required to generate the training dataset and by more than 150h the necessary annotation time usually needed to complete this task (Figure 1b).

2.4 Multi-device analysis

While DL approaches can be very powerful when tackling tasks on very similar images, challenges are often encountered when the imaging conditions change over time (e.g. due to a new device) [30, 31]. To increase the applicability of the proposed method to various experimental conditions, we investigated how MICRA-Net could be fine-tuned on a new dataset that contains similar structures but acquired on a new device. To address this, a brightfield microscopy dataset of Giemsa-stained [32] *P. vivax* (malaria) infected human blood smears was used (Figure 5a), for which the training and testing datasets had very distinct intensity distributions (Figure 5a,b) [28, 29].

The first attempt to solve the classification task consisted in predicting the presence of infected smears in a 256×256 pixel image. A mean testing classification accuracy of $(78 \pm 13) \%$ (mean \pm standard deviation, calculated from 5 different instances of the network) was obtained. Since the testing images had a very different pixel intensity distribution, we investigated whether the classification results could be improved by adjusting for this. To this aim, we considered i) modifying the threshold of the linear layer and ii) fine-tuning a model by training on $\{12, 24, 36\}$ sampled images from the test set using a k -fold training scheme (see Supplementary Note 4 & Supplementary Fig. 8). We repeated the fine-tuning process 5 times

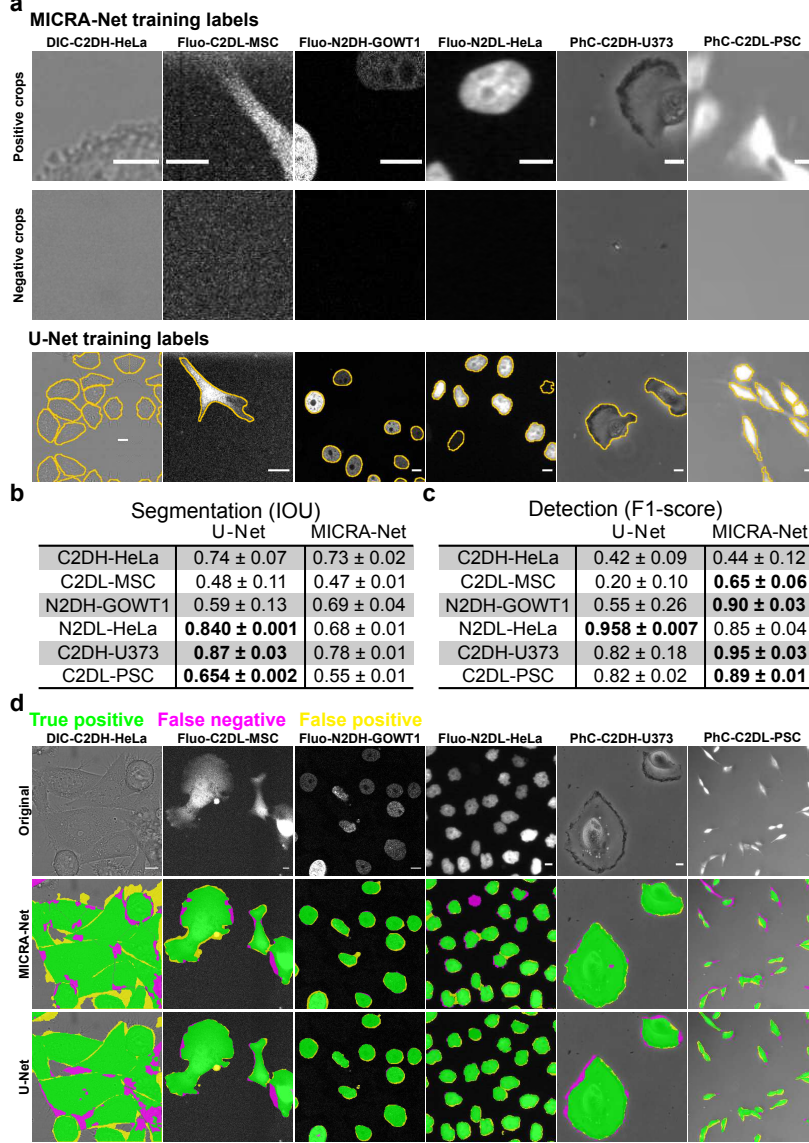


Figure 4: Cell counting and segmentation on 6 cell lines of the *Cell Tracking Challenge dataset*. a) Representative annotated images for each cell line used for MICRA-Net (top rows) and U-Net (bottom row) training. We compared the segmentation and detection of the MICRA-Net architecture with the pre-trained U-Net from Falk et al. [9]. Bootstrapped mean and 95% confidence interval are reported from 5 instances of network with random initialisation weights. d) Example images (*top row*) of each dataset comparing the MICRA-Net (*middle row*) and U-Net (*bottom row*) segmentation. On the segmentation maps, true positives are depicted in green, false positives in yellow and false negatives in magenta. All scale bars are 10 μm .

from each of the 5 naive instantiations (as starting points) while allowing i) linear layer [*Linear*], ii) linear layer and depth 4 [*Linear + 4*], iii) linear layer and depths 3 and 4 [*Linear + 3, 4*], and iv) all [*All*] layers to be updated (Figures 2a, 5c). A testing classification accuracy over 87% was obtained when updating the threshold and over 88% for all fine-tuned models, demonstrating the capability of MICRA-Net to be fine-tuned on similar tasks performed on images acquired on different devices (see Supplementary Table. 6 for detailed classification results).

In the context of parasite detection and stage determination for malaria, the most important task consists in the detection of infected cells [28]. When trained solely on the original training set, MICRA-Net performed worse on the detection task, obtaining a F1-score of 0.44 ± 0.13 (Figure 5c, d). However, with fine-tuning of

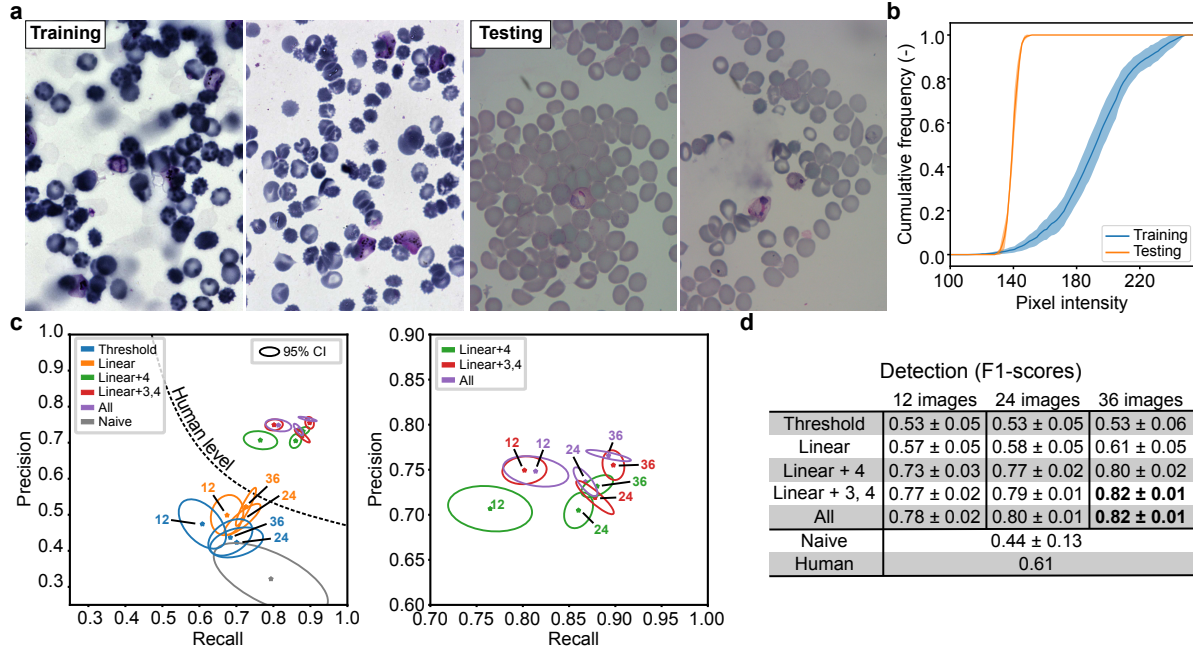


Figure 5: Segmentation of two different datasets of bright field microscopy images of Giemsa-stained red blood cells from Hung and Carpenter [29]. a) Representative images from the training (2 left) and testing (2 right) datasets. The training/validation dataset is composed of images taken from two different 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. 8 & 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 of all trained models is highlighted in bold. See Supplementary Tab. 7 for calculated p -values.

at least the linear layer and the depth 4 of the architecture, the F1-score was significantly increased, beating the inter-expert concordance (0.61 [29]). Additionally, increasing the number of images sampled from the testing set can significantly increase the detection accuracy (see Supplementary Table 7). 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.

We compared the segmentation results of MICRA-Net with expert precise annotations. Due to the lack of a precisely annotated dataset in the original publication by Ljosa et al. [28], we manually segmented all infected smears from the test set (303 smears). In contrast to the results obtained for the detection accuracy, updating more layers while fine-tuning (*Linear + 3, 4* $\{12, 24, 36\}$, and *All* $\{12, 24\}$) significantly reduced the IOU compared to only updating the linear layer (see Supplementary Fig. 9 & Supplementary Table 8). Hence, a trade-off should be made by the users according to their specific needs. For instance, with these

P. vivax datasets, the best trade-off to maximize both detection and segmentation efficiency requires the fine-tuning of at least the linear layer and depth 4.

2.5 Expert detection and segmentation assistance

The next step was to assess how MICRA-Net could be implemented as a tool to guide experts in the annotation of sparse and small structures in large images of an electron microscopy dataset. Our approach was tested on a dataset of Scanning Electron Microscopy (SEM) images of ultrathin mouse brain sections in which axons were genetically labeled with a small engineered peroxidase APEX2 [33] (referred to as Axon DAB, see Methods). In the SEM dataset, 1-10 small axonal regions (with an averaged size of 113×113 pixel) needed to be identified in images of around $10\,000 \times 10\,000$ pixel (Figure 6a). Applied to this dataset, MICRA-Net was used to suggest regions containing the Axon DAB marker and generate segmentation masks of the structure in the regions that were accepted by the expert.

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

In the context of very sparse detections, positive-unlabeled (PU) learning can improve the performance of a given architecture [34]. On the main classification task, an accuracy between 83% and 90% was obtained for all PU ratios (see Supplementary Tab. 9). We next investigated how PU learning could improve the detection rate of Axon DAB in the SEM images and obtained best performances for a PU ratio between 1:5 and 1:16 (Figure 6b & Supplementary Fig. 10). The usage of MICRA-Net for this sparse detection task resulted in an increase of the measured recall above the inter-expert concordance (0.79), while requiring from an expert to proof only 3.13% of a newly acquired image. Accordingly, the area that was inspected by the expert and consequently the annotation time were reduced by 30 folds. Additionally, MICRA-Net allowed the expert to detect 57 new Axon DAB regions in the test set (representing 25% more detections) that had been missed by the expert during the initial image annotation process (Figure 6c). This demonstrates the potential of MICRA-Net as a tool to assist experts in the analysis of newly acquired images, not only reducing the manual annotation time by 30 folds, but also increasing the recall above the inter-expert variability. We also inspected how it performed on a second auxiliary task: the segmentation of Axon DAB regions (Supplementary Fig. 10a). For this purpose, an expert carefully highlighted the boundaries of 170 positive Axon DAB regions sampled from the testing set. As in the detection task, MICRA-Net had the same tendency of achieving better performance with PU ratios above 1:2 and could obtain a maximal IOU score of 0.62 ± 0.03 with the 1:5 ratio (see Supplementary Fig. 10 & Supplementary Tab. 11). Application of MICRA-Net to this electron microscopy annotation task was thus successful to reduce the burden of generating the training dataset, while also significantly increasing the discovery of regions of interest that were missed by the manual expert annotation.

3 Discussion

While pixel-wise metrics and ground-truth annotations are well established in the field of DL and computer vision with natural images, retrieval of ground truth annotations in biomedical imaging is a laborious process, requires highly-trained experts, and annotation imprecision often occurs [3, 35]. This stresses the need for weakly-supervised DL approaches that do not rely on spatially precise annotations of the structure of interest, but rather on annotations that are easier and faster to retrieve. MICRA-Net, a CNN-based method, relies on the information embedded in the latent space of a main simple task, in our case classification, to learn multiple complementary tasks without the need to generate task-specific precisely annotated training sets. We designed multiple experiments to challenge MICRA-Net at solving common microscopy tasks (segmentation, enumeration or localization) relevant to high-throughput microscopy image analysis [3, 9]. Unlike multi-task learning [36], MICRA-Net does not combine auxiliary tasks to increase the learning performance of a main task, nor requires more annotations from the dataset for each task [37, 38]. Hence, the

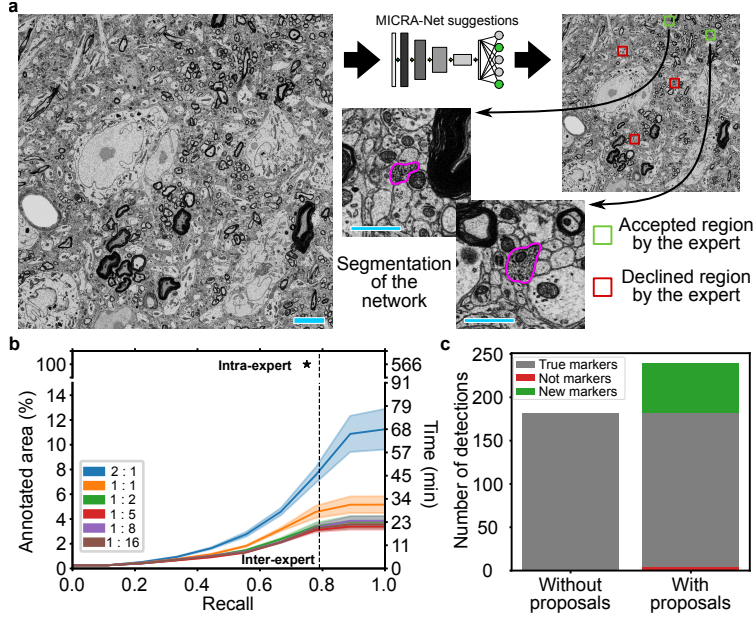


Figure 6: MICRA-Net is used as a tool to assist experts in the detection of sparse Axon DAB markers in large SEM images of ultrathin mouse brain sections. a) Schematic representation of the proposed approach. MICRA-Net is first swept over the entire field of view to output suggestions of plausible positions for axonal DAB markers. The suggestions 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 region. b) Validation of MICRA-Net suggestions accuracy. Two experts with variable levels of experience (high: [A], and intermediate: [B]) annotated all positive detections from MICRA-Net with a detection threshold set to have a recall of 1, *i.e.* all the original identifications (from expert [A], generated *without* MICRA-Net assistance) are detected by MICRA-Net. An intra-expert (expert [A]) recall of 0.756 was calculated for the visualisation of 100% of the field of view (without MICRA-Net assistance, star marker). Comparison of the decisions of [B] with [A] resulted in a inter-expert recall of 0.791 (dashed line). Using MICRA-Net as expert annotation assistance reduces the detection time by 25 folds (right vertical axis) while maintaining intra- and inter-experts recall levels. For constant recall, positive-unlabeled ratios of 1:2 and above allow a reduction of the annotation time compared to the 2:1 and 1:1 ratios, showing the importance of negative instances in the training dataset. The solid lines and associated pale regions are the bootstrapped mean and 95% confidence interval of 5 networks instantiated randomly for each condition. 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.

use of MICRA-Net should significantly reduce the burden of task-specific annotation of bioimaging datasets thereby increasing the accessibility of such deep learning based microscopy image analysis.

Our results show that MICRA-Net can be applied to various microscopy modalities and biological contexts, while significantly reducing the number of required expert decisions to generate the training dataset (Figure 1b). While fully-supervised DL approaches (e.g. based on a U-Net architecture) have the drawback of being costly to train, they can benefit from pre-training [9, 39, 40], and have access to precise information about the structure boundaries. On the other hand, MICRA-Net leverages on the extraction of spatial features from the hidden layers of the network to generate precise feature maps using solely, easy to retrieve, binary image-level annotations for training. MICRA-Net provides similar or even superior performance on multiple tasks to the state-of-the-art weakly- and fully-supervised learning approaches, thus making it an unprecedented alternative to address bioimaging analysis challenges for which large and precisely annotated datasets are not available.

Additionally we demonstrated that MICRA-Net could be fine-tuned when facing strong variations in

the quality of the available datasets, for example when images were acquired on two different microscopes. Fine-tuning of the architecture on few images from another microscopy system was sufficient to achieve better detection efficiency than inter-expert agreement. This is of particular interest for large-scale studies, conducted on multiple sites, that require analysis framework to be easily adaptable to new experimental conditions [28, 41, 42]. Future work on fine-tuning of such approaches to new structures of interest and analysis task will be an important step to increase their accessibility to a larger network of researchers.

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

Precise annotations, even if obtained from trained experts, are associated with inter-expert variability, especially when defining the boundaries. This variability needs to be assessed to characterize the annotated dataset and the precision of the neural network precision [3, 43]. Nevertheless, image-level binary levels can help to increase the consistency among experts by reducing the complexity of the annotation task. As a whole, MICRA-Net can be used in multi-class detection, segmentation, counting, and classification tasks in bioimaging, for which a precise annotated dataset is not available and can be used to reduce the identification variability between experts. The approach presented will simplify the development of deep learning based analysis strategies for microscopy and will open up new possibilities for quantitative bioimaging.

Acknowledgments

Laurence Emond for F-Actin sample preparation and immunocytochemistry. Francine Nault, Charleen Salesse and Laurence Emond for the neuronal cell culture. Jonathan Marek and Renaud Bernatchez for the development of a custom Python annotation application. Thibault Dhellemmes for inter-expert axon DAB annotations in electron microscopy images. Christian Gagné and Marc-André Gardner for preliminary discussion on semantic segmentation. Annette Schwerdtfeger and Ana Gabela for careful proofreading of the manuscript. Funding was provided by grants from the Natural Sciences and Engineering Research Council of Canada (P.D.K. and F.L.C.), Canadian Institutes of Health Research (P.D.K.), CERVO Brain Research Center Foundation (F.L.C.), the Canadian Foundation for Innovation (P.D.K.). F.L.C. is a Canada Research Chair Tier II, Audrey Durand is a CIFAR AI Chair, and A.B. is supported by a PhD scholarship from the Fonds de Recherche Nature et Technologie.

Author contributions

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

Competing interests

The authors declare no competing interest.

Data and code availability

The datasets included in this study are available from the corresponding author upon reasonable request. Open source code for the MICRA-Net approach will be available online upon publication of the article: <https://github.com/FLClab/MICRA-Net>

References

- [1] Lothar Schermelleh, Alexia Ferrand, Thomas Huser, Christian Eggeling, Markus Sauer, Oliver Biehlmaier, and Gregor PC Drummen. Super-resolution microscopy demystified. *Nature Cell Biology*, 21(1):72, 2019.
- [2] Flavie Lavoie-Cardinal, Anthony Bilodeau, Mado Lemieux, Marc-André Gardner, Theresa Wiesner, Gabrielle Laramée, Christian Gagné, and Paul De Koninck. Neuronal activity remodels the f-actin based submembrane lattice in dendrites but not axons of hippocampal neurons. *Scientific Reports (Nature Publisher Group)*, 10(1), 2020.
- [3] Thomas Schlegl, Philipp Seeböck, Sebastian M Waldstein, Georg Langs, and Ursula Schmidt-Erfurth. f-anogan: Fast unsupervised anomaly detection with generative adversarial networks. *Medical image analysis*, 54:30–44, 2019.
- [4] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 521(7553):436, 2015.
- [5] Anindya Gupta, Philip J Harrison, Håkan Wieslander, Nicolas Pielawski, Kimmo Kartasalo, Gabriele Partel, Leslie Solorzano, Amit Suveer, Anna H Klemm, Ola Spjuth, et al. Deep learning in image cytometry: a review. *Cytometry Part A*, 95(4):366–380, 2019.
- [6] Juan C. Caicedo, Allen Goodman, Kyle W. Karhohs, Beth A. Cimini, Jeanelle Ackerman, Marzieh Haghighi, CherKeng Heng, Tim Becker, Minh Doan, Claire McQuin, Mohammad Rohban, Shantanu Singh, and Anne E. Carpenter. Nucleus segmentation across imaging experiments: the 2018 Data Science Bowl. *Nature Methods*, 16(12):1247–1253, December 2019. ISSN 1548-7091, 1548-7105. doi: 10.1038/s41592-019-0612-7.
- [7] Erick Moen, Dylan Bannon, Takamasa Kudo, William Graf, Markus Covert, and David Van Valen. Deep learning for cellular image analysis. *Nature methods*, pages 1–14, 2019.
- [8] Vladimír Ulman, Martin Maška, Klas EG Magnusson, Olaf Ronneberger, Carsten Haubold, Nathalie Harder, Pavel Matula, Petr Matula, David Svoboda, Miroslav Radojevic, et al. An objective comparison of cell-tracking algorithms. *Nature methods*, 14(12):1141–1152, 2017.
- [9] Thorsten Falk, Dominic Mai, Robert Bensch, Özgün Çiçek, Ahmed Abdulkadir, Yassine Marrakchi, Anton Böhm, Jan Deubner, Zoe Jäckel, Katharina Seiwald, et al. U-net: deep learning for cell counting, detection, and morphometry. *Nature Methods*, 16(1):67, 2019.
- [10] Benjamin G Wilhelm, Sunit Mandad, Sven Truckenbrodt, Katharina Kröhnert, Christina Schäfer, Burkhard Rammner, Seong Joo Koo, Gala A Claßen, Michael Krauss, Volker Haucke, et al. Composition of isolated synaptic boutons reveals the amounts of vesicle trafficking proteins. *Science*, 344(6187):1023–1028, 2014.
- [11] Veronika Cheplygina, Marleen de Bruijne, and Josien P. W. Pluim. Not-so-supervised: A survey of semi-supervised, multi-instance, and transfer learning in medical image analysis. *Medical Image Analysis*, 54:280–296, May 2019. ISSN 1361-8415. doi: 10.1016/j.media.2019.03.009.
- [12] George Papandreou, Liang-Chieh Chen, Kevin P Murphy, and Alan L Yuille. Weakly-and semi-supervised learning of a deep convolutional network for semantic image segmentation. In *Proceedings of the IEEE international conference on computer vision*, pages 1742–1750, 2015.
- [13] Anna Khoreva, Rodrigo Benenson, Jan Hosang, Matthias Hein, and Bernt Schiele. Simple does it: Weakly supervised instance and semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 876–885, 2017.
- [14] Jia Xu, Alexander G Schwing, and Raquel Urtasun. Tell me what you see and i will show you where it is. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3190–3197, 2014.

- [15] Emanuele Pesce, Samuel Joseph Withey, Petros-Pavlos Ypsilantis, Robert Bakewell, Vicky Goh, and Giovanni Montana. Learning to detect chest radiographs containing pulmonary lesions using visual attention networks. *Medical image analysis*, 53:26–38, 2019.
- [16] Martin Rajchl, Matthew CH Lee, Ozan Oktay, Konstantinos Kamnitsas, Jonathan Passerat-Palmbach, Wenjia Bai, Mellisa Damodaram, Mary A Rutherford, Joseph V Hajnal, Bernhard Kainz, et al. Deepcut: Object segmentation from bounding box annotations using convolutional neural networks. *IEEE transactions on medical imaging*, 36(2):674–683, 2016.
- [17] Lin Yang, Yizhe Zhang, Zhuo Zhao, Hao Zheng, Peixian Liang, Michael TC Ying, Anil T Ahuja, and Danny Z Chen. Boxnet: Deep learning based biomedical image segmentation using boxes only annotation. *arXiv preprint arXiv:1806.00593*, 2018.
- [18] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco: Common objects in context. In *European conference on computer vision*, pages 740–755. Springer, 2014.
- [19] Alexander Vezhnevets, Vittorio Ferrari, and Joachim M Buhmann. Weakly supervised structured output learning for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 845–852. IEEE, 2012.
- [20] Florian Dubost, Hieab Adams, Pinar Yilmaz, Gerda Bortsova, Gijs van Tulder, M Arfan Ikram, Wiro Niessen, Meike Vernooij, and Marleen de Bruijne. Weakly supervised object detection with 2d and 3d regression neural networks. *arXiv preprint arXiv:1906.01891*, 2019.
- [21] Jiayun Li, William Speier, King Chung Ho, Karthik V Sarma, Arkadiusz Gertych, Beatrice S Knudsen, and Corey W Arnold. An em-based semi-supervised deep learning approach for semantic segmentation of histopathological images from radical prostatectomies. *Computerized Medical Imaging and Graphics*, 69:125–133, 2018.
- [22] Oren Z Kraus, Jimmy Lei Ba, and Brendan J Frey. Classifying and segmenting microscopy images with deep multiple instance learning. *Bioinformatics*, 32(12):i52–i59, 2016.
- [23] Bodhiswatta Chatterjee and Charalambos Poullis. Semantic segmentation from remote sensor data and the exploitation of latent learning for classification of auxiliary tasks. *arXiv preprint arXiv:1912.09216*, 2019.
- [24] Ramprasaath R Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 618–626, 2017.
- [25] Yann LeCun, Léon Bottou, Yoshua Bengio, Patrick Haffner, et al. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.
- [26] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.
- [27] Ke Xu, Guisheng Zhong, and Xiaowei Zhuang. Actin, spectrin, and associated proteins form a periodic cytoskeletal structure in axons. *Science*, 339(6118):452–456, 2013.
- [28] Vebjorn Ljosa, Katherine L Sokolnicki, and Anne E Carpenter. Annotated high-throughput microscopy image sets for validation. *Nature methods*, 9(7):637–637, 2012.
- [29] Jane Hung and Anne Carpenter. Applying faster r-cnn for object detection on malaria images. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 56–61, 2017.
- [30] Chinmay Belthangady and Loic A Royer. Applications, promises, and pitfalls of deep learning for fluorescence image reconstruction. *Nature methods*, pages 1–11, 2019.

- [31] Martin Weigert, Uwe Schmidt, Tobias Boothe, Andreas Müller, Alexandr Dibrov, Akanksha Jain, Benjamin Wilhelm, Deborah Schmidt, Coleman Broaddus, Siân Culley, et al. Content-aware image restoration: pushing the limits of fluorescence microscopy. *Nature methods*, 15(12):1090–1097, 2018.
- [32] Juan José Barcia. The giemsa stain: its history and applications. *International journal of surgical pathology*, 15(3):292–296, 2007.
- [33] Stephanie S Lam, Jeffrey D Martell, Kimberli J Kamer, Thomas J Deerinck, Mark H Ellisman, Vamsi K Mootha, and Alice Y Ting. Directed evolution of apex2 for electron microscopy and proximity labeling. *Nature methods*, 12(1):51–54, 2015.
- [34] Jessa Bekker and Jesse Davis. Learning from positive and unlabeled data: a survey. *Mach. Learn.*, 109(4):719–760, 2020.
- [35] Eric M Christiansen, Samuel J Yang, D Michael Ando, Ashkan Javaherian, Gaia Skibinski, Scott Lipnick, Elliot Mount, Alison O’Neil, Kevan Shah, Alicia K Lee, et al. In silico labeling: predicting fluorescent labels in unlabeled images. *Cell*, 173(3):792–803, 2018.
- [36] Rich Caruana. Multitask learning. *Machine learning*, 28(1):41–75, 1997.
- [37] Ross Girshick. Fast r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 1440–1448, 2015.
- [38] Sebastian Ruder. An overview of multi-task learning in deep neural networks. *arXiv preprint arXiv:1706.05098*, 2017.
- [39] Alexander Mathis, Pranav Mamidanna, Kevin M Cury, Taiga Abe, Venkatesh N Murthy, Mackenzie Weygandt Mathis, and Matthias Bethge. Deeplabcut: markerless pose estimation of user-defined body parts with deep learning. *Nature neuroscience*, 21(9):1281–1289, 2018.
- [40] Kaiming He, Ross Girshick, and Piotr Dollár. Rethinking imagenet pre-training. In *Proceedings of the IEEE international conference on computer vision*, pages 4918–4927, 2019.
- [41] Kevin W. Eliceiri, Michael R. Berthold, Ilya G. Goldberg, Luis Ibáñez, B. S. Manjunath, Maryann E. Martone, Robert F. Murphy, Hanchuan Peng, Anne L. Plant, Badrinath Roysam, Nico Stuurman, Jason R. Swedlow, Pavel Tomancak, and Anne E. Carpenter. Biological imaging software tools. *Nature Methods*, 9(7):697–710, July 2012. ISSN 1548-7105. doi: 10.1038/nmeth.2084.
- [42] Wei Ouyang, Casper F. Winsnes, Martin Hjelmare, Anthony J. Cesnik, Lovisa Åkesson, Hao Xu, Devin P. Sullivan, Shubin Dai, Jun Lan, Park Jinmo, Shaikat M. Galib, Christof Henkel, Kevin Hwang, Dmytro Poplavskiy, Bojan Tunguz, Russel D. Wolfinger, Yinzhen Gu, Chuanpeng Li, Jinbin Xie, Dmitry Buslov, Sergei Fironov, Alexander Kiselev, Dmytro Panchenko, Xuan Cao, Runmin Wei, Yuanhao Wu, Xun Zhu, Kuan-Lun Tseng, Zhifeng Gao, Cheng Ju, Xiaohan Yi, Hongdong Zheng, Constantin Kappel, and Emma Lundberg. Analysis of the Human Protein Atlas Image Classification competition. *Nature Methods*, 16(12):1254–1261, December 2019. ISSN 1548-7105. doi: 10.1038/s41592-019-0658-6.
- [43] Gloria P. Mazzara, Robert P. Velthuizen, James L. Pearlman, Harvey M. Greenberg, and Henry Wagner. Brain tumor target volume determination for radiation treatment planning through automated MRI segmentation. *International Journal of Radiation Oncology, Biology, Physics*, 59(1):300–312, May 2004. ISSN 0360-3016. doi: 10.1016/j.ijrobp.2004.01.026.
- [44] Harold Hotelling. Analysis of a complex of statistical variables into principal components. *Journal of educational psychology*, 24(6):417, 1933.
- [45] Adam Paszke, Sam Gross, Soumith Chintala, Gregory Chanan, Edward Yang, Zachary DeVito, Zeming Lin, Alban Desmaison, Luca Antiga, and Adam Lerer. Automatic differentiation in pytorch. In *31st Conference on Neural Information Processing Systems*, 2017.
- [46] Robert L Cook. Stochastic sampling in computer graphics. *ACM Transactions on Graphics (TOG)*, 5(1):51–72, 1986.

- [47] Nobuyuki Otsu. A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, 9(1):62–66, 1979.
- [48] Stefan Van der Walt, Johannes L Schönberger, Juan Nunez-Iglesias, François Boulogne, Joshua D Warner, Neil Yager, Emmanuelle Gouillart, and Tony Yu. scikit-image: image processing in python. *PeerJ*, 2:e453, 2014.
- [49] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.
- [50] Varduhi Yeghiazaryan and Irina D Voiculescu. Family of boundary overlap metrics for the evaluation of medical image segmentation. *Journal of Medical Imaging*, 5(1):015006, 2018.
- [51] Harold W Kuhn. The hungarian method for the assignment problem. *Naval research logistics quarterly*, 2(1-2):83–97, 1955.
- [52] Michael M Scott, Christi J Wylie, Jessica K Lerch, RoxAnne Murphy, Katherine Lobur, Stefan Herlitze, Weihong Jiang, Ron A Conlon, Ben W Strowbridge, and Evan S Deneris. A genetic approach to access serotonin neurons for in vivo and in vitro studies. *Proceedings of the National Academy of Sciences*, 102(45):16472–16477, 2005.
- [53] Phillip I. Good. *Resampling Methods*. Birkhäuser Basel, 3 edition, 2006. ISBN 978-0-8176-4444-4.

4 Methods

4.1 MICRA-Net architecture

Figure 2a shows the schematic representation of the MICRA-Net architecture. Each depth of the network contains two blocks of convolutions (kernel size of 3) followed by batch normalization, and ReLU activation. The number of filters in the convolutional layers is doubled after maxpooling (stride and kernel size of 2) to increase the richness of the representation. The number of filters for each layer is $\{32, 64, 128, 256\}$. Global maxpooling on the output layer allows a reduction of the dimensionality and a fully connected layer (FCL) is used to provide a classification prediction. Dropout (probability of 0.5) is applied on the input features of the FCL.

At inference, MICRA-Net predicts a whole image target from a given sample. Then, from each activated class c , a local map L^l is calculated from the weighted combination of the activation map $A^{l,k}$ and the mean gradient $\alpha_{l,k}^c$ of each l layer [24]. The mean gradient $\alpha_{l,k}^c$ is calculated from the backpropagated class activation y^c

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

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

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

Since MICRA-Net produced spatially reduced feature maps, local maps were upsampled using nearest neighbor interpolation to match the input image size of 256×256 pixel. These images were then normalized in the range $[0, 1]$ using a min-max scaling. ReLU activation is applied on the last layer (L^8) of the network, as in the seminal implementation of Grad-CAM [24], to be used for the coarse segmentation. Local maps from layers L^{1-7} (Figure 2a-c) were concatenated into a feature space and retrieved the first principal component of every pixel using principal component analysis (PCA) [44] decomposition to retain prominent information from the feature space. The network was built and trained with the PyTorch library [45].

4.2 Datasets

4.2.1 Modified MNIST dataset

We generated the modified MNIST training dataset by randomly sampling N digits from the original MNIST training dataset and randomly distributed them on a 256×256 pixel field of view. To avoid overlap between digits we used a random Poisson disc sampling algorithm with a radius size of 25 pixels [46]. The number of digits N was uniformly sampled from $\{1, 2, 3, 4, 5, 10, 15, 20, \text{Max}\}$, where **Max** corresponds to the maximum number of digits that can be placed without overlap. A rotation of $\pm 30^\circ$ uniformly sampled was applied to the digits before placement on the image. We applied, in a random order, a Gaussian blur with sigma uniformly sampled in $[0, 2[$ and artificial normalized Poisson noise with $\lambda = \frac{\sqrt{255}}{2}$. The resulting image intensities were clipped to lie in $[0, 1]$. Using this technique, we generated 2000 and 1000 images for training and validation respectively.

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

4.2.2 F-actin dataset

The F-actin dataset was generated by using a sliding window of size 256×256 pixel with a stride of 192 pixels over 260 complete images with an approximate size of 1000×1000 pixel. Since the super-resolution microscopy images used are mostly composed of background, we set out to keep the crops containing at least 10% of dendritic area thereby reducing the number of crops to identify. The dendritic mask was obtained from the foreground detection on the confocal image of the dendritic marker MAP2 using a global Otsu thresholding on the normalized Gaussian blurred image [47]. The sigma parameter of the Gaussian blur was set to 20 pixels as it provided suitable dendrite detection over a wide range of images. We next annotated each generated crop as being positive to the presence of the F-actin periodical lattice or longitudinal fibers. The resulting training dataset contained 3832 crops (256×256 pixel, 897 images positive to the periodical lattice and 1456 positive to the longitudinal fibers), the validation dataset contained 1287 crops (405 positive to periodical lattice and 377 positive to fibers), and the testing dataset contained 416 crops (83 positive to 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 which we calculated the median in addition to 3 standard deviation. The minimum value was calculated as the median of minimas (**min**). To ensure a proper scaling of the images we also added a scaling factor of 0.8

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

To evaluate the segmentation performance of the trained models, an expert precisely highlighted the contours of the structures in 50 images (25 images positive to periodical lattice and 25 images positive to fibers) from the testing set randomly sampled. This small segmentation dataset only served to compare the segmentation performance from the MICRA-Net, weakly-supervised U-Net, and user-study.

4.2.3 Cell Tracking Challenge dataset

We selected 6 cell line datasets from the Cell Tracking Challenge (CTC) [8]: the DIC-C2DH-HeLa dataset which was acquired using differential interferometry contrast microscopy, three non-synthetic fluorescence microscopy datasets (Fluo-C2DL-MSC, Fluo-N2DH-GOWT1, and Fluo-N2DL-HeLa) and two phase contrast microscopy datasets (PhC-C2DH-U373, and PhC-C2DL-PSC). All original images were rescaled in the $[0, 1]$ range using a per image min-max scale. We then resized each image and associated precise annotations according to the specific needs using bi-linear interpolation and nearest neighbors respectively with the **Scikit-Image** [48] Python library (see Supplementary Table 3 for scaling factors). We used a sliding window of size 128×128 pixel or 256×256 pixel with a 75% overlap between crops in both directions. Using this sliding window technique yielded a total of 27,106 positive crops and 3,364 negative crops for the 256×256 pixel crops resized to have an effective pixel size of $0.5 \mu\text{m}$. The sliding window with size 128×128 pixel crops and resized to have single cells in the field of view yielded a total of 121,699 positive crops and

77,589 negative crops. We generated weak annotations from the precise contours of the cells provided in the original CTC dataset by identifying an image crop as positive if the corresponding annotated crop contained at least 1% of annotated area, and negative otherwise. To evaluate the segmentation and detection tasks, we manually segmented 4 images randomly sampled per cell line in the testing set.

4.2.4 P. vivax dataset

We used image set BBBC041v1, available from the Broad Bioimage Benchmark Collection [28]. The complete dataset contained 1327 3-channel images and was already split into a training (1207 images) and testing (120 images) set. The dataset is composed of blood smears that were stained with Giemsa reagent [32] and acquired on three different brightfield microscopes from three different laboratories. All blood smears (infected or uninfected) were annotated using bounding boxes. The blood smears were later classified as infected (gametocytes, rings, trophozoites, and schizonts) or uninfected (red blood cells, and leukocytes) by an expert. The task was to differentiate infected from uninfected blood smears. The dataset is highly unbalanced towards red blood cells which composes over 95% of the annotated cells.

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

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

4.2.5 Scanning Electron Microscopy dataset

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

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

4.3 MICRA-Net training procedure

Here follows the general training procedure of the MICRA-Net architecture. For specific training details for each dataset, see Supplementary Notes 1-5. MICRA-Net was trained using the Adam optimizer with a learning rate specific to each dataset and other default parameters [49]. A learning rate scheduler was used to reduce the learning rate of the optimizer with a minimal possible learning rate of 1×10^{-5} . The number of training epochs was adapted to the specific dataset (see Supplementary Tab. 12-16). Early stopping was used to reduce overfitting. Unless otherwise specified, we used binary cross entropy with logits loss. We kept the model generalizing the most on the validation set during the training phase.

Data augmentation was used to increase the performance of the network. Refer to Supplementary Tab. 12-16 for a detailed data augmentation procedure for each dataset. All operations were applied in a random order with a probability of 50%.

4.4 Evaluation procedure

4.4.1 Segmentation

The binary segmentation masks were obtained from MICRA-Net using the dataset specific procedures described in Supplementary Notes 1-5. The segmentation performances of the trained models was evaluated using three common evaluation metrics: F1-score, Intersection Over Union (IOU), and the Symmetric Boundary Dice (SBD) [50]. If multiple instances of a model were trained on the same task, we bootstrapped the average of the trained models to show the bootstrapped mean and 95% confidence interval (10 000 repetitions).

4.4.2 Detection

The centroid of each detected object was obtained from MICRA-Net by using the dataset specific procedures detailed in Supplementary Notes 1-5. Each detected centroid was associated with the centroid of objects in the ground truth mask using the Hungarian algorithm [51] with a maximal distance of N pixels, where N is approximately the object radius. In this context, an associated detected object is considered as a true positive, a non-associated detected object is a false positive, and a missed ground truth object is a false negative. To evaluate the detection capability of MICRA-Net, we reported the F1-score.

For a quantitative comparison, we repeated the evaluation for each trained model. We then bootstrapped the average of the trained models to show the bootstrapped mean and 95% confidence interval (10 000 repetitions).

4.4.3 Classification

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

4.4.4 Custom performance metrics

The F-actin periodical lattice is detected as an oscillating pattern between high- and low-intensity stripes with 180-190 nm periodicity [27]. We designed a metric that would take this periodicity into account to evaluate the MICRA-Net precise segmentation performance. We computed, as a baseline, the Fourier transform (FT) of the original image (FT_b) and the FT of the segmented regions: for the expert (FT_e), and for the predicted segmentation masks (FT_{pred}). The variation from the baseline was computed as the difference in the FT spectrum, for spatial frequencies in the range $[170, 200[$ nm, between $FT_{e,pred}$ and FT_b over the sum of FT_b . A smaller absolute difference between the variation of the expert and the variation of the predicted mask implies more similar segmentation.

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

4.4.5 User-study for F-Actin segmentation

We performed a user-study in which six experts highlighted the contours of the F-actin periodical lattice and longitudinal fibres on a small dataset of 50 images using polygonal bounding boxes. We used polygonal bounding boxes as this annotation method reduces the time required by an expert by more than 3 folds compared to precisely identifying the boundaries of the structures (Supplementary Fig. 6). We used our own annotation application that was optimized for this type of task. Annotation of the full dataset required

approximately 40 minutes for the expert. The averaged performance of the six experts was compared to MICRA-Net using F1-score, IOU, and SBD.

4.5 In-house datasets acquisition

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

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

4.5.2 Animals and stereotaxic injections for scanning electron microscopy dataset

This study was carried out on 3-month-old mice, weighing 25-35g. Animals were housed under a 12h light-dark cycle with water and food ad libitum. All procedures were approved by the *Comité de Protection des Animaux de l'Université Laval*, in accordance with the Canadian Council on Animal Care's Guide to the Care and Use of Experimental Animals (Ed2), and with the ARRIVE guidelines. Maximum efforts were made to minimize the number of animals used. Transgenic e-Pet Cre mice expressing Cre recombinase under the control of Fev promoter, known to be specific for serotonin (5-HT) neurons [52], were injected in the dorsal raphe nucleus (DRN) with 1 μ l of AAV9-CAG-DIO-APEX2NES-WPRE. Stereotaxic injections were done using a 30° angle along the frontal plane at AP: -4.78; ML: +2.00 and DV: -3.20. In these injected transgenic mice, the small engineered peroxidase APEX2 [33] is specifically expressed in the cytosol/cytoplasm of 5-HT-infected neurons of the DRN and is used, in presence with hydrogen peroxide, to oxidize 3,3 Diaminobenzidine (DAB) chromogen that can readily be visible at the light and electron microscope levels.

4.5.3 Tissue preparation for scanning electron microscopy dataset

After a period of 21 days following stereotaxic injection, mice were anesthetized with a mixture of ketamine (100 mg/kg) and xylazine (10 mg/kg) and transcardially perfused with 50 ml of phosphate-buffered-saline (PBS: 50 mM at pH 7.4) followed by 150 ml of 4% paraformaldehyde (PFA) and 1% glutaraldehyde diluted in phosphate buffer (PB; 100 mM at pH 7.4). Brains were dissected out, post-fixed for 24h in the same fixative solution and cut with a vibratome (model VT1200; Leica, Germany) into 50 μ m-thick frontal sections, which were serially collected in sodium phosphate buffer saline (PBS, 100 mM, pH 7.4). Frontal brain sections at the level of the subthalamic nucleus (STN) were processed to reveal the presence of APEX2 in axons arising from DRN-infected neurons using 3,3'diaminobenzidine (DAB; catalog no. D5637; Sigma-Aldrich) as the chromogen. Briefly, selected 50 μ m-thick sections were washed 3 times in PBS and then twice in Tris. Sections were then incubated for 1h in 0.05% DAB solution diluted in Tris, then for 1h in 0.05% DAB solution containing 0.015% hydrogen peroxide (H_2O_2). Sections were then rinsed twice in Tris and 3 times in PBS. Sections were temporally mounted in PBS and coverslipped for light microscope examination. STN sections containing DAB-labeled axons were selected for further processing. These sections were washed 3 times in PB, then incubated during 1h in 2% osmium tetroxide diluted in 1.5% potassium ferrocyanide solution. They were then washed 3 times in ddH₂O, incubated for 20 min in 1% thiocarbonylhydrazide (TCH) solution and washed again 3 times in ddH₂O. Sections were placed 30 min in 2% osmium tetroxide and washed 3 times in ddH₂O. Sections were then dehydrated in ethanol and propylene oxide and flat-embedded in Durcupan (Electron microscopy Science). Areas of interest were cut from embedded sections and glued to the tip of resin blocks. Blocks were cut with an ultramicrotome (Leica EM UC7) in ultrathin sections (80 nm), which were serially collected on silicon-coated 10 x 10 mm chip wafer (Ted Pella, Inc; #16006).

4.5.4 Scanning electron microscopy (SEM)

Serial sections were imaged in a SEM (Zeiss Gemini 540) with the help of the ATLAS acquisition software. Images were acquired at a resolution of 5 nm/pixel, using acceleration voltage of 1.4 kV and current of 1.2 nA. Serial sections acquisitions produced a stack of 38 rectangle images of 25370 x 25633 pixel (126.850 x 128.165 microns) taken out of 38 ultrathin sections. In addition, a large single section acquisition was acquired and produced a single trapezoidal image of 31065 pixels for the small base (155.329 microns), 91393 pixel for the large base (456.967 microns) and 53161 pixels for the height (265.809 microns). All acquired images were subdivided into overlapping square tiles of 10240×10240 pixel (51.2 x 51.2 microns).

4.6 Statistical assessment using resampling

Resampling was used as a statistical test to verify the statistical difference between two groups [53]. Statistical analysis was performed using a randomization test with the null hypothesis being that the different conditions (A, B) belong to the same distribution. The absolute difference between mean values of A and B was calculated ($D_{\text{gt}} = |\mu_A - \mu_B|$). For the randomization test, each value belonging to A and B was randomly reassigned to A' and B', with the sizes of A' and B' being N_A and N_B , respectively. The absolute difference between the mean values of A' and B' was determined ($D_{\text{rand}} = |\mu_{A'} - \mu_{B'}|$) and the randomization test was repeated 10 000 times. The obtained distribution was compared with the absolute difference of the mean of A and B (D_{gt}) to verify the null hypothesis.

When the number of groups was greater than 2, the F-statistic was sampled from each group using a resampling method. The F-statistic was calculated from all groups (A, B, C, etc.) as a ground truth (F_{gt}). Each value was randomly re-assigned to new groups (A', B', C', etc.) where group X' has the same size as group X. The F-statistic of newly formed groups (F_{rand}) was calculated and this process was repeated 10 000 times. We compared F_{rand} with F_{gt} to confirm the null hypothesis that the groups have the same mean distribution. When the null hypothesis was rejected, *i.e.* at least one group did not have the same mean distribution, we compared each group in a one-to-one manner using the randomization test described above. In all cases, a confidence level of 0.05 was used to reject the null hypothesis.

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

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

Figures

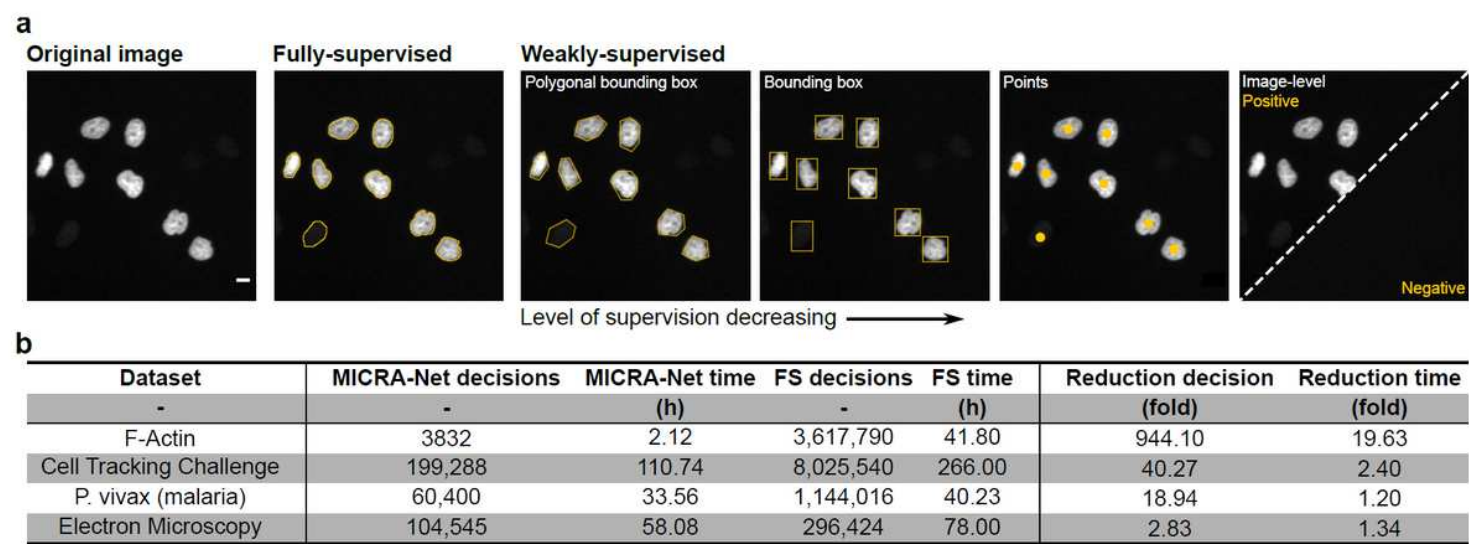


Figure 1

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] overlaid 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) 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 FS 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 FS learning, the decision and annotation time for each dataset separately on a precisely annotated subset of images was evaluated (see Methods).

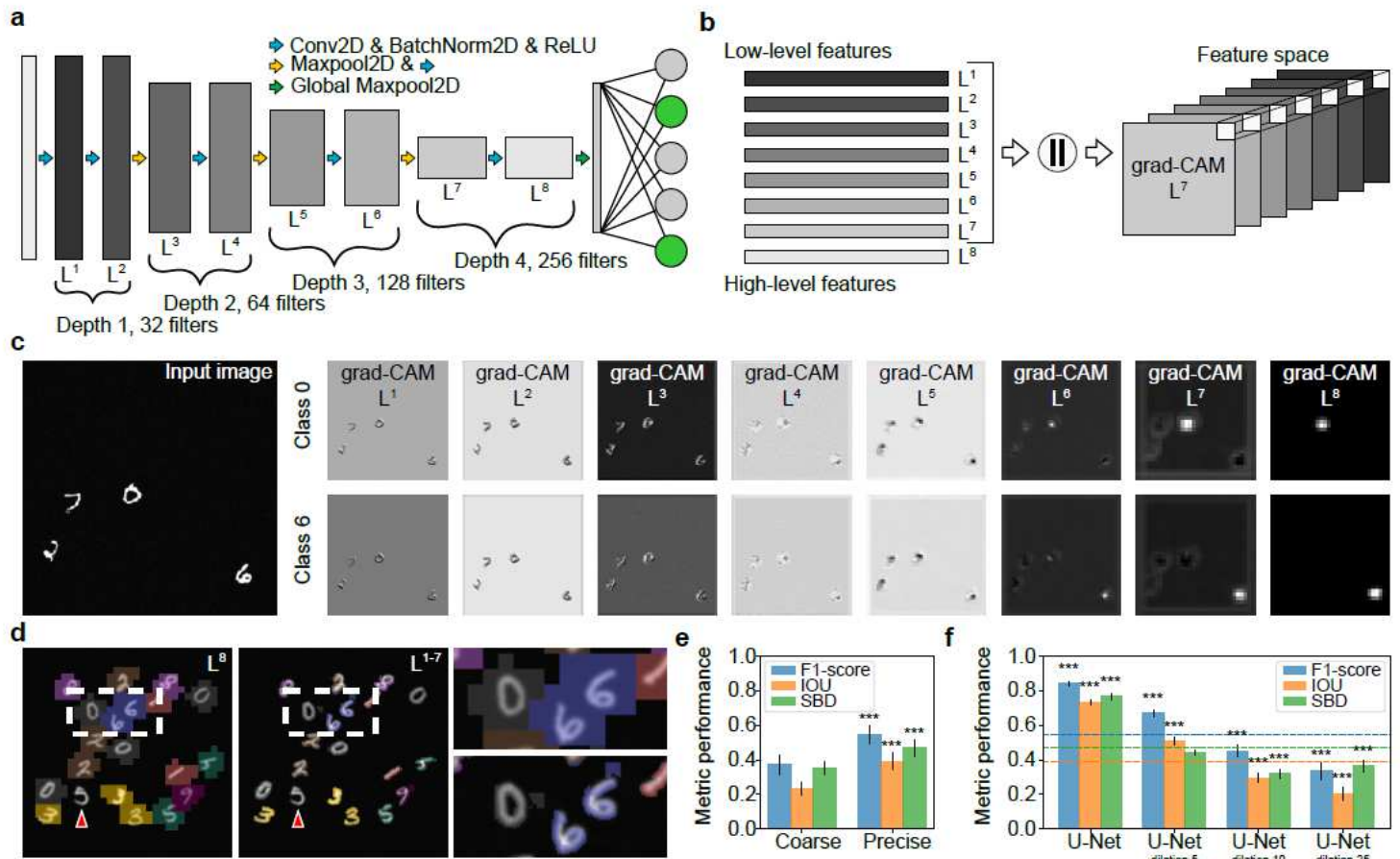


Figure 2

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 L8 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 class 6 on the modified MNIST dataset. Each activated class is backpropagated through the network and a local map for each layer of the network (L1–8) is computed. d) Coarse and precise segmentation maps of the digits of a representative image (256 × 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) Evaluation of the coarse and precise segmentation performance using the i.e. F1-score, intersection over union (IOU), and symmetric boundary dice (SBD) (see Methods). Reported here is the mean performance over the 10 classes (see Supplementary Fig. 1 for class-wise and density-wise performances). A significant increase in the segmentation performance is measured for precise over coarse segmentation (t-test, $p_{\text{F1-score}} = 1.6616 \times 10^{-6}$, $p_{\text{IOU}} = 5.2650 \times 10^{-7}$, $p_{\text{SBD}} = 1.1627 \times 10^{-5}$). f) Mean performance over the 10 classes obtained with the U-Net trained with and without dilation across different depths.

without dilation of the ground truth contours. The segmentation maps are presented in Supplementary Fig. 2a. The MICRA-Net precise segmentation (color-coded dashed lines) surpasses the U-Net trained with 10 pixels dilation and is not statistically different from the U-Net trained with 5 pixels dilation. Only fully-supervised training outperforms MICRA-Net precise segmentation (one-way ANOVA followed by posthoc t-test, $pF1\text{-score} = 2.0254 \times 10^{-12}$, $pIOU = 8.7200 \times 10^{-14}$, $pSBD = 5.7291 \times 10^{-13}$). All p-values are reported in Supplementary Tab. 1. Bar graphs show the mean values and standard deviation.

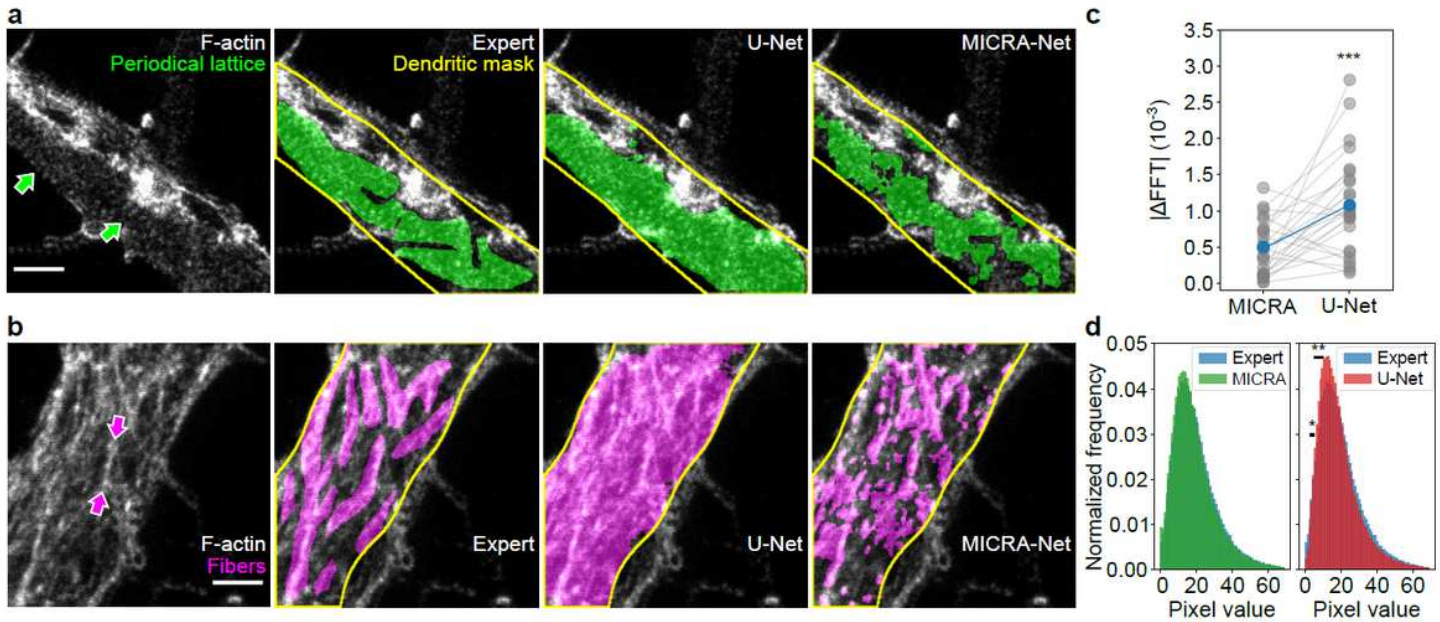


Figure 3

Semantic segmentation of F-actin nanostructures observed on super-resolution microscopy images. a,b) Left: 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 (b). Arrows point towards the periodical lattice (green) and longitudinal fibers (magenta). Middle-Left: precise expert contours of the nanostructure, Middle-Right: weakly-supervised U-Net segmentation masks and Right MICRA-Net segmentation masks. c, d) Performance evaluation of MICRA-Net and weakly-supervised U-Net segmentation on the precisely annotated testing dataset using custom metrics for rings (c) and fibers (d). The FFT (c) and intensity distribution (d) metrics evaluate the difference between the pixels found within the precise expert annotations and the two DL-based segmentation approaches (see Methods). An increased segmentation accuracy is measured with MICRA-Net compared to U-Net for rings (c) and fibers (d). No significant differences are measured between the precise expert annotations and MICRA-Net segmentation mask for both metrics. Statistical analysis: c) t-test, $p = 3.7342 \times 10^{-4}$, d) all p-values are shown in Supplementary Tab. 2. Performance evaluation was performed within the dendritic mask (a,b: yellow line). a,b) Scale bars: 1 μm .

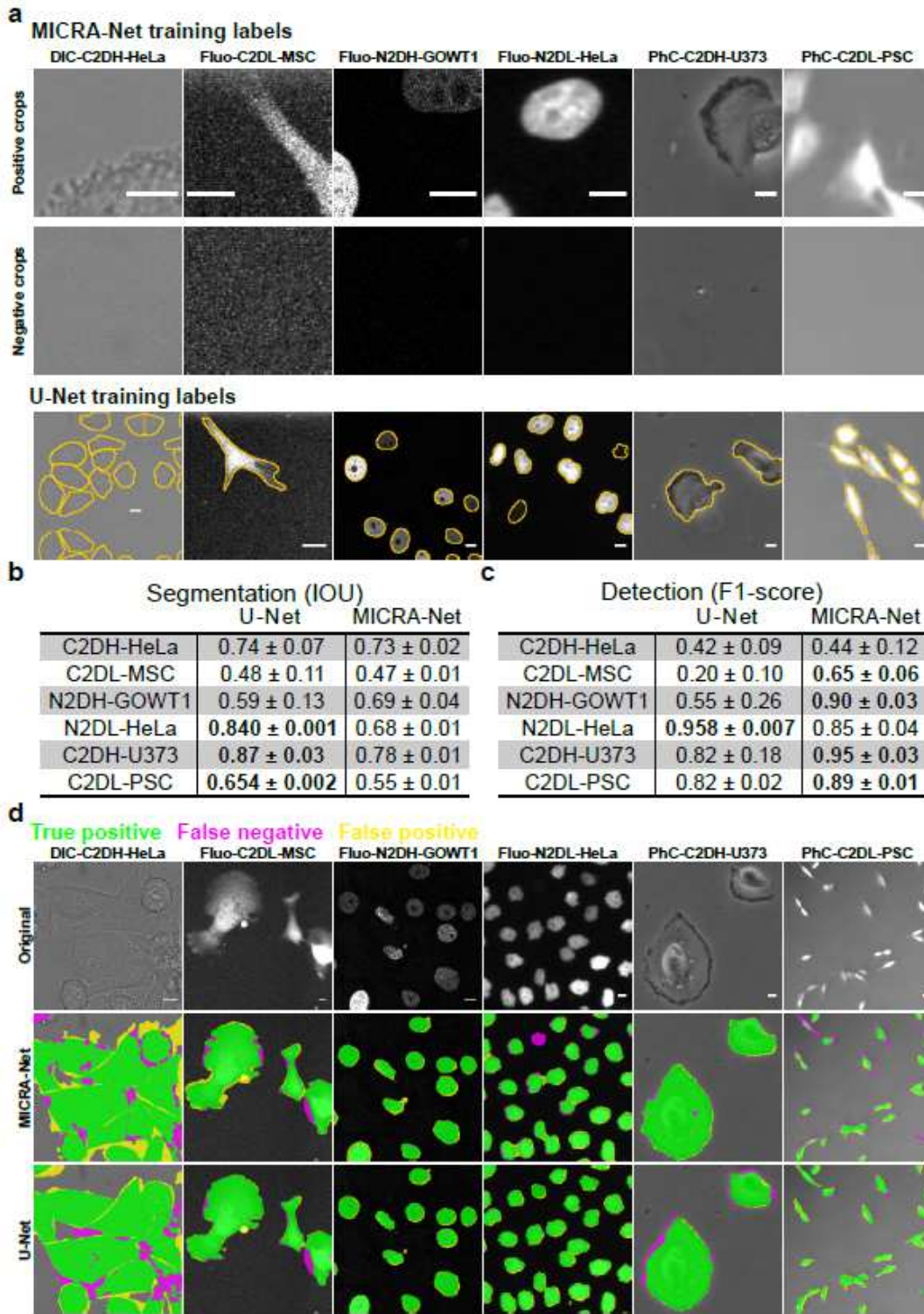


Figure 4

Cell counting and segmentation on 6 cell lines of the Cell Tracking Challenge dataset. a) Representative annotated images for each cell line used for MICRA-Net (top rows) and U-Net (bottom row) training. We compared the segmentation and detection of the MICRA-Net architecture with the pre-trained U-Net from Falk et al. [9]. Bootstrapped mean and 95% confidence interval are reported from 5 instances of network with random initialisation weights. d) Example images (top row) of each dataset comparing the MICRA-

Net (middle row) and U-Net (bottom row) segmentation. On the segmentation maps, true positives are depicted in green, false positives in yellow and false negatives in magenta. All scale bars are 10 μm .

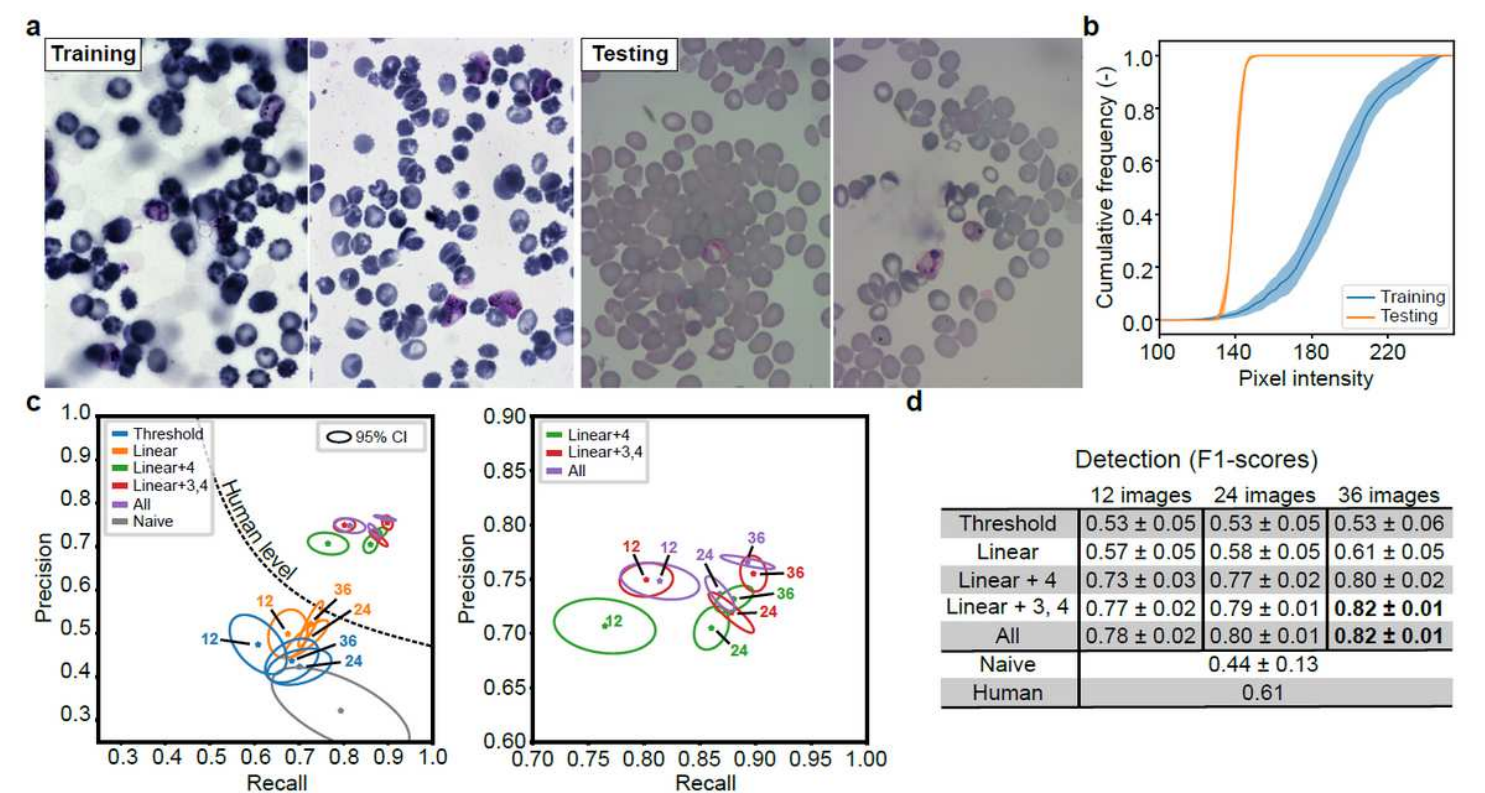


Figure 5

Segmentation of two different datasets of bright field microscopy images of Giemsa-stained red blood cells from Hung and Carpenter [29]. a) Representative images from the training (2 left) and testing (2 right) datasets. The training/validation 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. 8 & 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 of all trained models is highlighted in bold. See Supplementary Tab. 7 for calculated p-values.

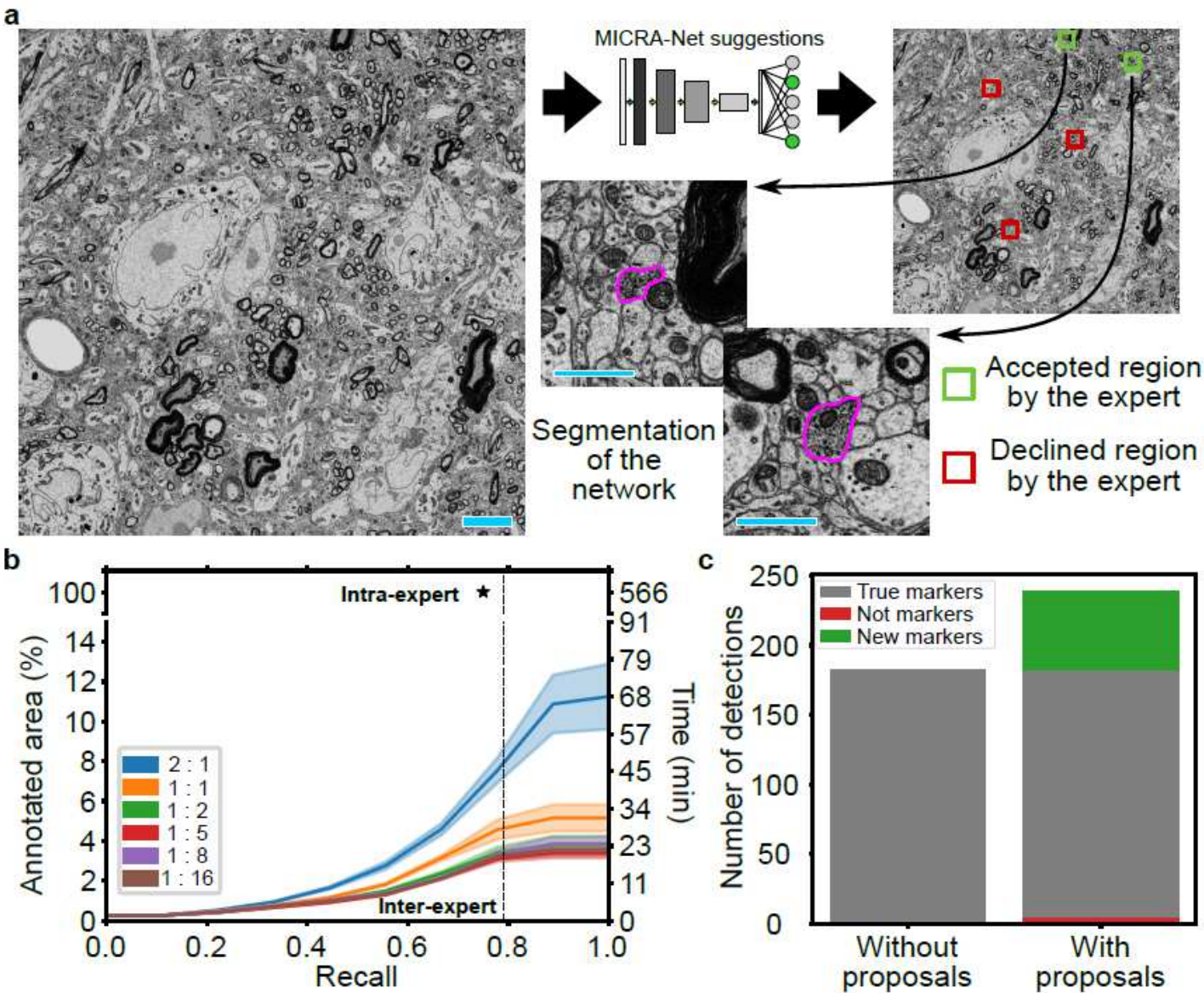


Figure 6

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 to output suggestions of plausible positions for axonal DAB markers. The suggestions 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 region. b) Validation of MICRA-Net suggestions accuracy. Two experts with variable levels of experience (high: [A], and intermediate: [B]) annotated all positive detections from MICRA-Net with a detection threshold set to have a recall of 1, i.e. all the original identifications (from expert [A], generated without MICRA-Net assistance) are detected by MICRA-Net. An intra-expert (expert [A]) recall of 0.756 was calculated for the visualisation of 100% of the field of view (without MICRA-Net assistance, star marker). Comparison of the decisions of [B] with [A]

resulted in a inter-expert recall of 0.791 (dashed line). Using MICRA-Net as expert annotation assistance reduces the detection time by 25 folds (right vertical axis) while maintaining intra- and inter-experts recall levels. For constant recall, positive-unlabeled ratios of 1:2 and above allow a reduction of the annotation time compared to the 2:1 and 1:1 ratios, showing the importance of negative instances in the training dataset. The solid lines and associated pale regions are the bootstrapped mean and 95% confidence interval of 5 networks instantiated randomly for each condition. 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.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Bilodeau2020SupMat09102020.pdf](#)