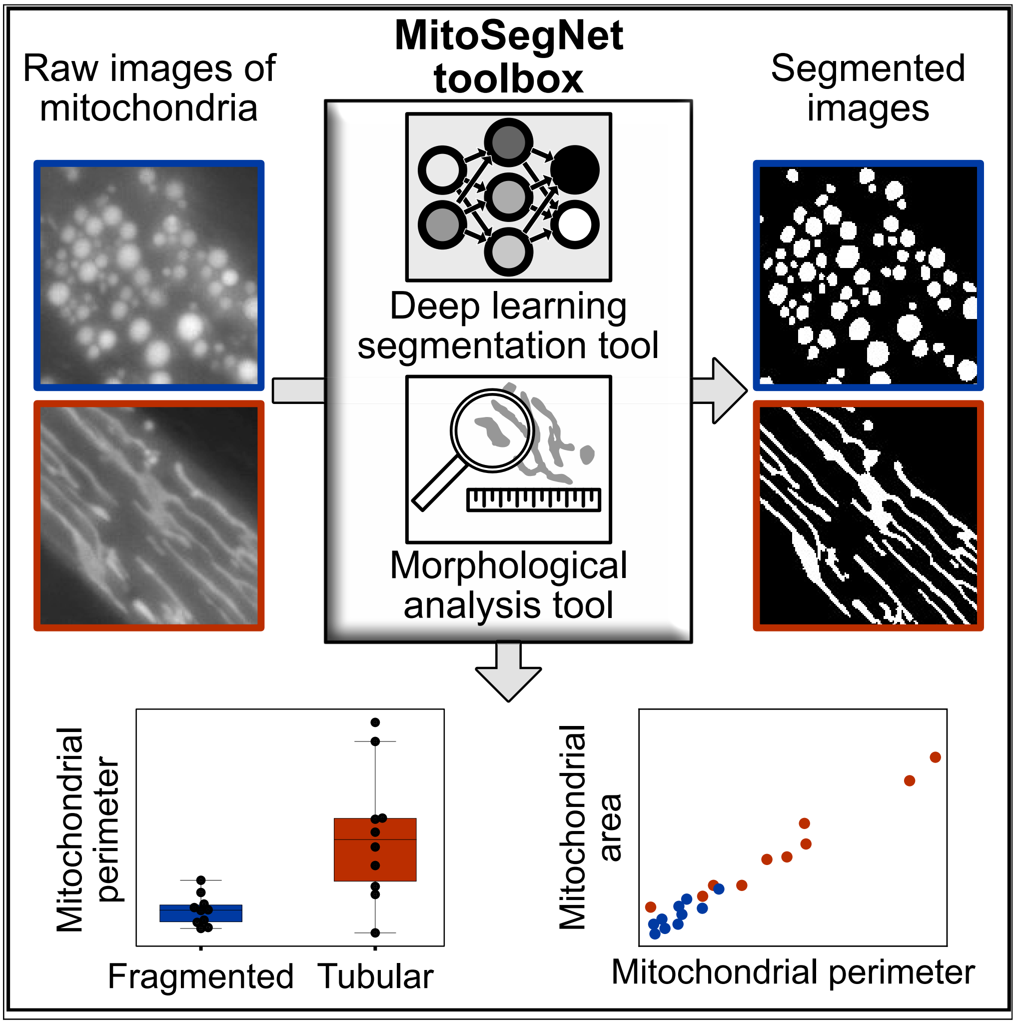
iScience

CellPress
OPEN ACCESS

Article

MitoSegNet: Easy-to-use Deep Learning Segmentation for Analyzing Mitochondrial Morphology



Christian A. Fischer, Laura Besora-Casals, Stéphane G. Rolland, ..., Michael Duchen, Barbara Conradt, Carsten Marr

b.conradt@ucl.ac.uk (B.C.)
carsten.marr@helmholtz-muenchen.de (C.M.)

HIGHLIGHTS

MitoSegNet is a deep learning model for quantifying mitochondrial morphology

MitoSegNet outperforms other mitochondria segmentation methods

MitoSegNet was used to identify previously unreported morphological differences

MitoSegNet successfully segments mitochondria in HeLa cells

Fischer et al., iScience *23*, 101601
October 23, 2020 © 2020 The Authors.
https://doi.org/10.1016/j.isci.2020.101601

iScience

CellPress OPEN ACCESS

Article

MitoSegNet: Easy-to-use Deep Learning Segmentation for Analyzing Mitochondrial Morphology

Christian A. Fischer,[1,2,3] Laura Besora-Casals,[1] Stéphane G. Rolland,[1] Simon Haeussler,[1] Kritarth Singh,[4] Michael Duchen,[4] Barbara Conradt,[1,2,4,*] and Carsten Marr[3,5,*]

SUMMARY

While the analysis of mitochondrial morphology has emerged as a key tool in the study of mitochondrial function, efficient quantification of mitochondrial microscopy images presents a challenging task and bottleneck for statistically robust conclusions. Here, we present Mitochondrial Segmentation Network (MitoSegNet), a pretrained deep learning segmentation model that enables researchers to easily exploit the power of deep learning for the quantification of mitochondrial morphology. We tested the performance of MitoSegNet against three feature-based segmentation algorithms and the machine-learning segmentation tool Ilastik. MitoSegNet outperformed all other methods in both pixelwise and morphological segmentation accuracy. We successfully applied MitoSegNet to unseen fluorescence microscopy images of mitoGFP expressing mitochondria in wild-type and *catp-6*$^{\mathrm{ATP13A2}}$ mutant *C. elegans* adults. Additionally, MitoSegNet was capable of accurately segmenting mitochondria in HeLa cells treated with fragmentation inducing reagents. We provide MitoSegNet in a toolbox for Windows and Linux operating systems that combines segmentation with morphological analysis.

INTRODUCTION

Cellular organelles are integral to eukaryotic cells and their functions. One organelle that has always been of particular interest is the mitochondrion, which plays an essential role in several metabolic pathways including that of Adenosine triphosphate (ATP). Mitochondria are often represented as static, bean-shaped organelles but actually form highly dynamic 'tubular' networks that often undergo changes in distribution and morphology (Tilokani et al., 2018; Chan, 2020). The steady-state morphology of mitochondria in a cell is a result of a balance between two opposing processes, mitochondrial fusion and fission. Changes in this balance result in changes in mitochondrial morphology. It has been shown that changes in morphology allow mitochondria to respond to metabolic or environmental stresses, while maintaining homeostasis (Tondera et al., 2009; Rolland et al., 2013; Wai and Langer, 2016). While partially damaged mitochondria can be rescued by exchanging their contents with functional mitochondria through mitochondrial fusion, mitochondrial fission enables the removal of damaged mitochondria and can also facilitate apoptosis during increased levels of cellular stress (Pernas and Scorrano, 2016). Mitochondrial fusion and fission are regulated by a conserved family of dynamin-related GTPases and have been well studied in *Caenorhabditis* elegans (van der Bliek et al., 2017). In *C. elegans*, the membrane anchored dynamin-related GTPases FZO-1$^{\mathrm{MFN}}$ and EAT-3$^{\mathrm{OPA1}}$ are required for the fusion of the outer- and inner mitochondrial membranes, respectively. The loss of function of either of these two proteins results in mitochondria with a 'fragmented' morphology (Breckenridge et al., 2008; Ichishita et al., 2008; Kanazawa et al., 2008; Tan et al., 2008; Rolland et al., 2009).

Mitochondrial fission in *C. elegans* is mediated by DRP-1$^{\mathrm{DRP-1}}$, a cytosolic dynamin-related GTPase. Depletion of DRP-1 has been shown to result in mitochondria with an 'elongated' morphology (Labrousse et al., 1999). Mutations in the human orthologs of the genes encoding these proteins have been associated with several diseases, including neurodegenerative diseases (Chan, 2020). For this reason, understanding mitochondrial fusion and fission is not only an important basic biological question but is critical for our ability to understand the pathology of these diseases and to develop novel therapeutics to treat them. However, such studies have been hindered by the fact that is difficult to assess mitochondrial morphology in different genetic backgrounds or physiological

[1] Fakultät für Biologie, Ludwig-Maximilians-Universität Munich, Planegg-Martinsried, Munich, 82152 Bavaria, Germany

[2] Centre for Integrated Protein Science, Ludwig-Maximilians-University, Planegg-Martinsried, Munich, 82152 Bavaria, Germany

[3] Institute of Computational Biology, Helmholtz Zentrum München – German Research Center for Environmental Health, Ingolstädter Landstr. 1, 85764 Neuherberg, Germany

[4] Department of Cell and Developmental Biology, Division of Biosciences, University College London, London WC1E 6AP, UK

[5] Lead Contact

*Correspondence: b.conradt@ucl.ac.uk (B.C.), carsten.marr@helmholtz-muenchen.de (C.M.)

https://doi.org/10.1016/j.isci.2020.101601

conditions in an unbiased and quantitative manner. Specifically, the diversity of shapes among mitochondria (elongated, fragmented, tubular, as well as 'mixed' morphologies) poses a challenge to the automated quantification of mitochondrial morphology. For this reason, researchers often resorted to the use of a simple qualitative assessment of mitochondrial morphology. As a result, subtle differences in morphology and, hence, phenotypes are often not detected. To analyze mitochondrial morphology, for example, in *C. elegans*, mitochondria are labeled using either a mitochondria-specific fluorescent dye (such as TMRE) or a transgene expressing a mitochondrial-targeted GFP (mitoGFP) (Regmi and Rolland, 2017) and a fluorescent microscopy image is acquired. In order to quantify the mitochondrial morphology in an automated and unbiased manner, the next critical step is to segment the objects in the fluorescent image. The simplest form of image segmentation is thresholding, which is only successful if features are well separated and their intensities vary considerably from the background (Torborg and Feller, 2004). These requirements are seldom met in live cell imaging due to autofluorescence, noise or fluctuating intensities. Thresholding segmentation can be improved through the prior application of feature enhancement algorithms based on intensity derived features, such as the Difference-of-Gaussians (DoG), Determinant of Hessian (Sato et al., 1998) or Laplacian-of-Gaussian (LoG), which are also known as blob detectors. DoG is used to enhance the visibility of edges by removing high frequency information but at the cost of reducing the overall image contrast, while LoG is useful for detecting edges that appear at different image scales or degrees of focus (Marr and Hildreth, 1980; Lindeberg, 1994). Curvilinear structures (such as nerve fibers or blood vessels) can be segmented using the eigenvalues of a Hessian matrix, with which one can calculate the object curvature. There are a many other methods available used in segmentation workflows, such as morphological filtering (dilation, erosion, etc.), region accumulation (watershed transform), deformable model fitting (active contour model) and machine learning (k-means clustering, random forest, etc.) (Meijering, 2012). Most of these methods can now be implemented by biologists through free and opensource tools such as Fiji (Schindelin et al., 2012), CellProfiler (McQuin et al., 2018) or Ilastik (Berg et al., 2019).

All of these segmentation methods have shown varying degrees of success depending on the images they were supposed to segment (de Boer et al., 2015; Li et al., 2015; Akram et al., 2017; Berg et al., 2019). However, with rising image complexity as well as a decreased signal-to-noise ratio, most of the methods perform poorly. For such cases, the only option in the past was manual segmentation, which is highly laborious and introduces a varying degree of bias on each labeled image. With the recent emergence of deep learning and in particular the development of convolutional neural networks (CNNs) (LeCun et al., 1989; Krizhevsky et al., 2012) automated approaches that perform these tasks with human accuracy have become available. CNNs were inspired by the research of Hubel and Wiesel on the primary visual cortex of cats (Wiesel and Hubel, 1963). CNN's can classify data based on convolution and pooling operations. Convolution describes the extraction of features from an image by sliding filters across the image and generating feature maps. Pooling reduces the dimensionality of each feature map, while retaining the most important information. It also reduces the number of network parameters, prevents overfitting, and makes the network invariant to small distortions in the input image (Scherer et al., 2010). Through the successive and repetitive application of convolution and pooling, CNNs are capable of classifying highly complex images with great accuracy (Szegedy et al., 2014). To perform semantic segmentation, which is the assignment of a class label to each pixel, one must use a fully convolutional neural network (FCNN) (Long et al., 2014). A popular FCNN in the biological community is the U-Net that was specifically developed for biomedical image segmentation (Ronneberger et al., 2015). It has been successfully applied to many different biomedical image segmentation tasks and yields good results with only a few hand-segmented images (Chlebus et al., 2018; De Fauw et al., 2018; Stember et al., 2018).

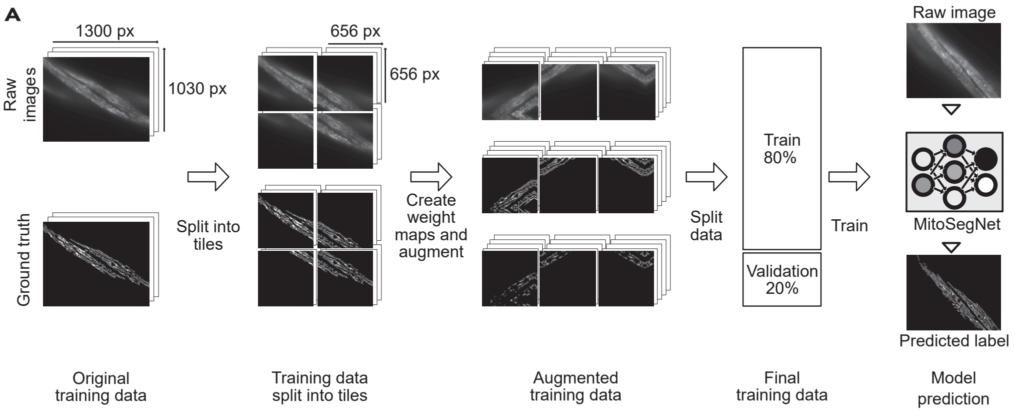
In this study, we trained a U-Net, which we named Mitochondrial Segmentation Network (MitoSegNet), to learn how to segment mitochondria in adult *C. elegans* body wall muscle cells, compared its performance and tested its generalizability in biologically relevant applications that demonstrate *C. elegans* animals carrying a loss-of-function mutation in the gene *catp-6*^{ATP13A2} exhibit a previously unreported mitochondrial morphology phenotype. We also show that MitoSegNet can be successfully used to analyze mitochondrial morphology in HeLa cells.

RESULTS

The MitoSegNet Model

The MitoSegNet model was generated by training a modified U-Net with a training set of 12 1300 × 1030 pixel fluorescent microscopy, maximum-intensity projection images, depicting mitochondria in body wall muscle cells of adult *C. elegans* worms (mitochondria were visualized using a transgene expressing mitochondrial matrix-targeted GFP under the control of a body wall muscle-specific promoter (P_{myo3}::*mitoGFP*))

A



B

Training session 1	Training session 2	Training session 3		Training session 12
Test img 1	Train img 1	Train img 1		Train img 1
Train img 2	Test img 2	Train img 2		Train img 2
Train img 3	Train img 3	Test img 3		Train img 3
⋮	⋮	⋮	…	⋮
Train img 11	Train img 11	Train img 11		Train img 11
Train img 12	Train img 12	Train img 12		Test img 12

Figure 1. Training the MitoSegNet Model and Using It with the MitoS Tool

(A) The original training data is comprised 12 raw images and the appendant hand-generated ground truth images. Each image is split into 4 overlapping tiles of equal length. For each tile, a weight map is generated and subsequently all three set of tiles (raw images, ground truth, and weight maps) are augmented 80 times, increasing the size of the training data to 3,840 image tiles. Prior to training, the augmented training data is split into training (80%) and validation data (20%). The pretrained MitoSegNet model can now be used to segment new images of mitochondria.

(B) We performed a cross validation for which 12 separate MitoSegNet models were trained each with 11 images, excluding one image that was later used to test the prediction accuracy against other segmentation methods.

(Figure 1A and Methods). Our U-Net modification entails the removal of dropout layers at the end of the contracting pathway and instead placing batch normalization layers after every convolutional layer prior to ReLU activation in the contracting pathway. This modification decreased the amount of necessary training time. Each image was split into 4 overlapping tiles. For each tile, 80 augmented copies were generated for training the model. A cross validation was performed to estimate the performance of the MitoSegNet on an unseen test set and to compare it against other segmentation methods (Figure 1B).

Visual Comparison of Segmentation Performance

To qualitatively evaluate the performance of the MitoSegNet, we compared the predicted segmentations against manually segmented ground truth in an unseen test set. The same procedure was repeated for four other segmentation methods. We considered three classical feature enhancement methods (Gaussian, Hessian, and Laplacian) followed by different thresholding algorithms, all implemented in ImageJ/Fiji. The fourth method is the machine-learning segmentation tool Ilastik (Kreshuk and Zhang, 2019). The Gaussian, Hessian, Laplacian, and Ilastik methods failed to consistently prevent false positive and/or false negative segmentation on all phenotypes (Figure 2). The Gaussian segmentation produced large sections of false positive predictions in the mixed and tubular phenotype (indicated by yellow arrows). The Hessian and Laplacian segmentation largely avoided false positive predictions but instead often failed to recognize mitochondria, resulting in false negative segmentations in the elongated, mixed, and tubular phenotype (and fragmented for the Laplacian segmentation) (Figure 2). The Ilastik-based segmentation produced only very little false negative predictions but like the Gaussian segmentation, predicted large amounts of false positives in all but the fragmented phenotype. The MitoSegNet segmentation drastically reduces the amount of false negative or false positive segmentation when compared to the other methods and yielded consistent results across all different phenotypes (Figure 2).

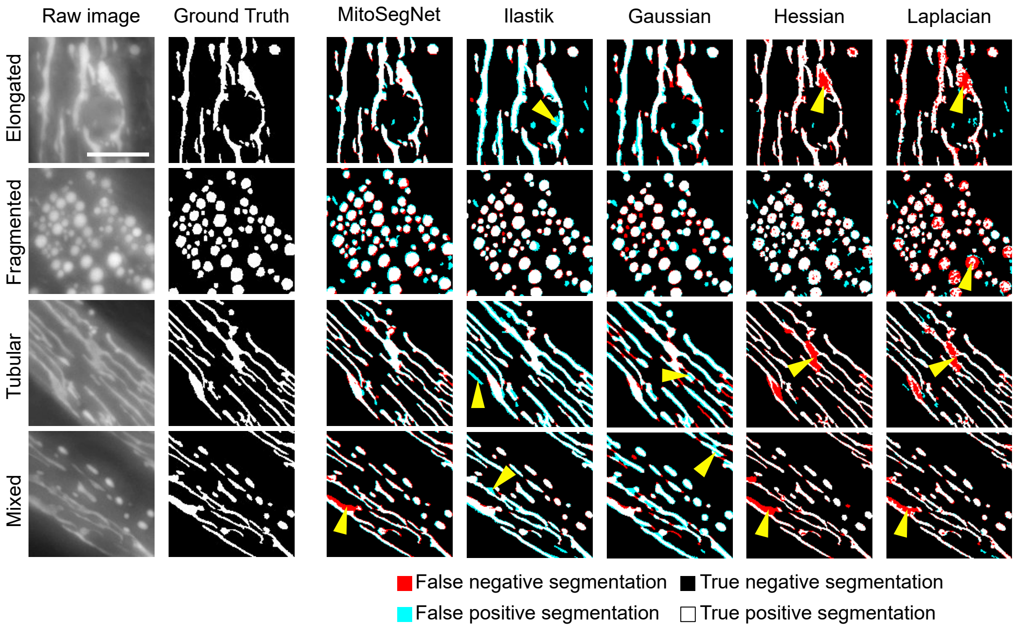


Figure 2. MitoSegNet Visually Outperforms Feature-Based Segmentation Approaches

On the left side, four images of elongated, fragmented, tubular, and mixed *C. elegans* mitochondria and their respective ground truth are shown. The masks on the right show the results of MitoSegNet and the four segmentation methods applied to each image, displaying the false negative segmentation in red, the false positive segmentation in turquoise, the true negative segmentation as black, and the true positive segmentation as white. The yellow arrows indicate areas in which false segmentation occured. The scale bar is 5 μm.

Quantitative Comparison of Segmentation Performance

To compare the methods' performance more quantitatively, we evaluated the pixelwise segmentation accuracy using the dice coefficient (Taha and Hanbury, 2015). The MitoSegNet outperforms the feature-based and Ilastik-generated segmentations (Figure 3A) with a median dice coefficient of 0.89 and a lower and upper 95% confidence interval of 0.87 and 0.91 ($N = 12$) significantly ($p = 5.11{*}10^{-5}$, Kruskal-Wallis test). However, pixelwise accuracy as measured by the dice coefficient does not necessarily guarantee correct prediction of morphology (Figure 3B). Because segmented images in biology are often used for morphological quantification (de Boer et al., 2015; Abdolhoseini et al., 2019; Orozco-Fuentes et al., 2019), we assessed the morphological accuracy with two other approaches. The single object shape deviation per object was measured for five shape descriptors (area, eccentricity, aspect ratio, perimeter, and solidity) and averaged over 12 images (Figure 3C). The MitoSegNet with a median average fold deviation of 1.09 and a lower and upper 95% confidence interval of 1.07 and 1.12 ($N = 60$) outperforms all other methods in the accurate prediction of single object morphology ($p = 7.4{*}10^{-10}$, Kruskal-Wallis test) (Figure 3D). Because the single object shape deviation method does not consider false negative predictions, we also compared all segmented objects in ground truth and prediction. For each image and each of the five object descriptors, the energy distance between the ground truth and predicted distributions was calculated (Figure 3E). Due to the different value ranges among the descriptors the values were normalized prior statistical analysis. The MitoSegNet segmentation achieves a median normalized energy distance of 0.20 with a lower and upper 95% confidence interval of 0.16 and 0.23 ($N = 60$) and again statistically outperforms all other non-deep learning segmentation methods ($p = 3.3{*}10^{-18}$, Kruskal-Wallis test) (Figure 3F).

Comparison of Mitochondrial Morphology between *Catp-6*^{ATP13A2} Mutant and Wild-type

To evaluate the applicability of the MitoSegNet on a different, unseen set of images, we decided to use the MitoSegNet to determine whether a loss-of-function mutation of the gene *catp-6*^{ATP13A2}, *ok3473* (hereafter referred to as *catp-6*(lf)), causes a mitochondrial morphology phenotype. *catp-6*^{ATP13A2} encodes a member of the family of P-type ATPases, which transport various compounds across membranes using ATP hydrolysis as energy source (Moller et al., 1996; Lambie et al., 2013; Anand et al., 2020). In addition, *catp-6*^{ATP13A2}

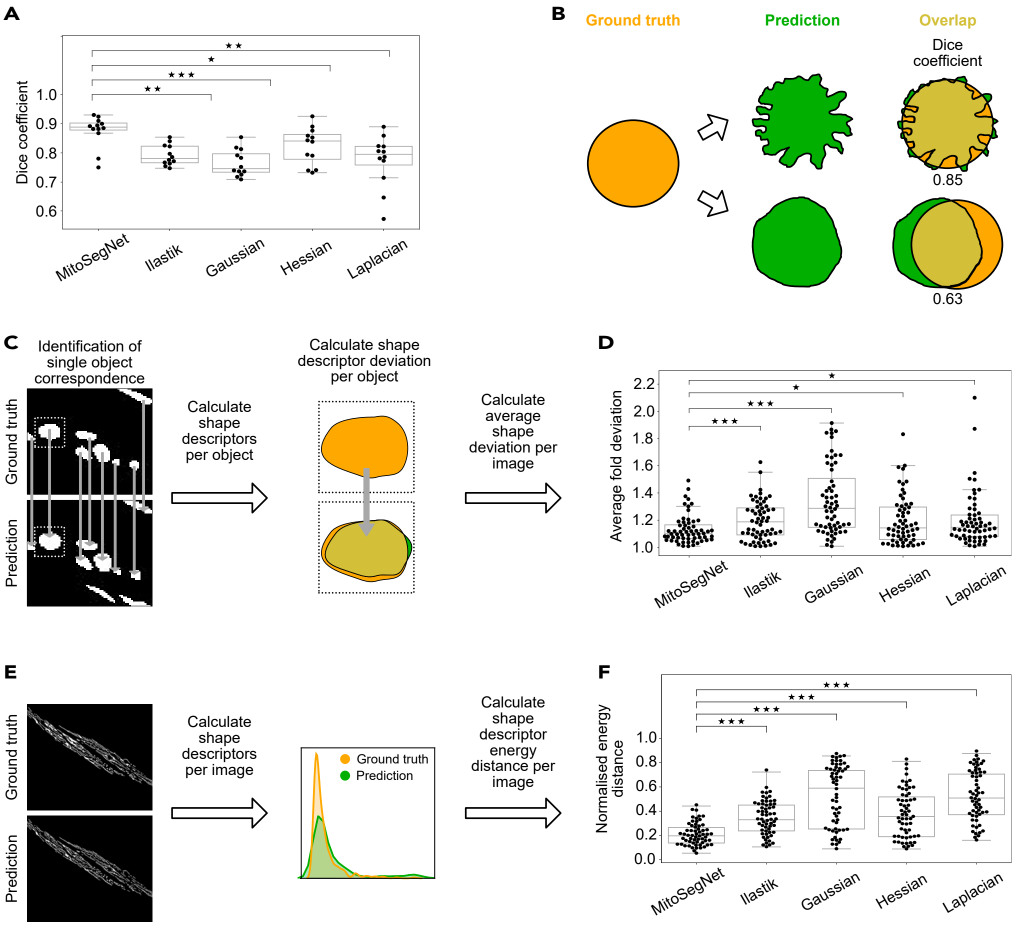


Figure 3. MitoSegNet's Pixelwise Accuracy Outperforms Non-Deep Learning Segmentation Methods

(A) The average dice coefficient achieved with the MitoSegNet is significantly higher than with the four other segmentation approaches. The data was statistically evaluated by using the Kruskal-Wallis test followed by a Dunn's multiple comparisons test.

(B) The dice coefficient is limited as a predictor of morphological segmentation performance: A dice coefficient close to 1.0 does not guarantee correct prediction of morphology. Contrary, a low dice coefficient does not rule out an accurate shape prediction. Ground truth segmentation is shown in orange, and the predicted segmentation in green.

(C) To gain insight into how accurately the shape of ground truth objects is predicted, five object shape descriptors (area, eccentricity, aspect ratio, perimeter, and solidity) are compared by calculating the fold deviation. Predicted objects that correspond to more than one ground truth object (or vice versa) are excluded from this analysis.

(D) The MitoSegNet shows the lowest average fold deviation between predicted and ground truth object shape descriptors. The data were statistically evaluated by first testing for normality using D'Agostino's K-squared test and then subsequently using the Kruskal-Wallis test followed by a Dunn's multiple comparisons test. N = 60.

(E) To determine the total morphological prediction accuracy, the same five shape descriptors were measured for each image. The descriptor distributions in the ground truth and predicted images were statistically evaluated for differences by calculating the energy distances between predicted and ground truth distribution. The energy distances for each shape descriptor and image were normalized prior to statistical analysis.

(F) The MitoSegNet shows the lowest normalized energy distance, statistically outperforming all other segmentation approaches. The data were first tested for normality using the D'Agostino's K-squared. After determining that all distributions were non-parametric, a Kruskal-Wallis test was used followed by a Dunn's multiple comparisons test. N = 60. $^*p < 0.05$, $^{**}0.001 < p < 0.01$, $^{***}p < 0.001$ for (A), (D) and (F).

is the human ortholog of ATP13A2, mutations of which leads to Kufor-Rakeb syndrome, a form of inherited juvenile-onset Parkinsons disease (Ramirez et al., 2006; Di Fonzo, Chien et al., 2007). No abnormal differences in mitochondrial morphology has so far been reported for the *catp-6*(lf). Consistent with this, upon brief visual inspection, no obvious differences in mitochondrial morphology are noticeable (Figure 4A). We applied the MitoSegNet (Figures 4A) to 19 fluorescence microscopy images of each genotype

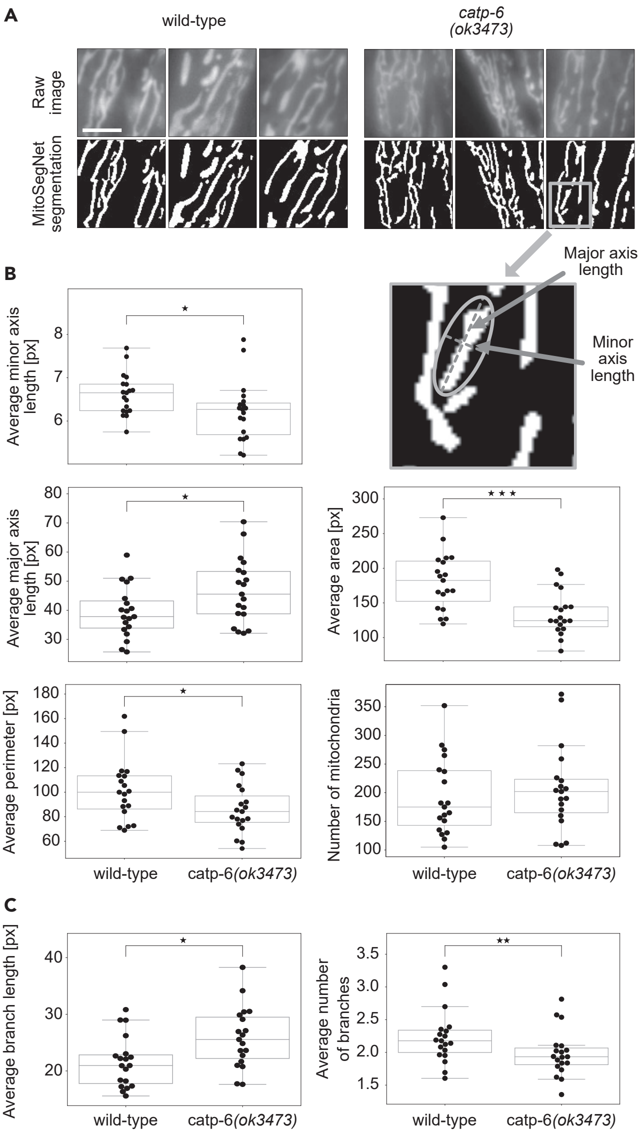


Figure 4. Wild-type and *Catp-6* Mutant Mitochondria Show Significant Morphological Differences

(A) Visual comparison of *catp-6* mutant and wild-type mitochondrial morphology. Raw images are at the top and MitoSegNet model segmentations at the bottom. The scale bar is 3 μm.

(B) Mitochondrial shape descriptor comparison.

(C) Mitochondrial branch descriptor comparison. Average area, minor and major axis length (see scheme for explanation), perimeter, number of mitochondria, average branch length, and number of branches were measured in segmented images of wild-type and *catp-6* mutant mitochondria. $^{*}p < 0.05$, $^{**}0.001 < p < 0.01$, $^{***}p < 0.001$ using the Mann-Whitney U test. N = 19.

and subsequently analyzed the data with the MitoA tool (Figure 4B). Segmentation masks (Figure 4A) visually matched the raw images closely and subsequent quantification revealed a statistically significant morphological difference between mitochondria in wild-type and *catp-6*(lf) mutants. Compared to wild-type, mitochondria are significantly thinner and longer in *catp-6*(lf) mutants, as determined by the average minor ($p = 0.047$, independent two-sample t test) and major axis length ($p = 0.029$, independent two-sample t test) (Figure 4B). Furthermore, the average mitochondrial area is larger ($p = 0.00039$, independent two-sample t test) and the perimeter is longer ($p = 0.043$, independent two-sample t test) in wild-type compared to *catp-6*(lf) mutants (Figure 5B). Excessive mitochondrial fission (i.e. mitochondrial fragmentation) as a cause for these observations can be excluded since the numbers of mitochondria in wild-type and *catp-6*(lf) mutants are similar ($p = 0.56$, independent two-sample t test) (Figure 4B). Differences were also found

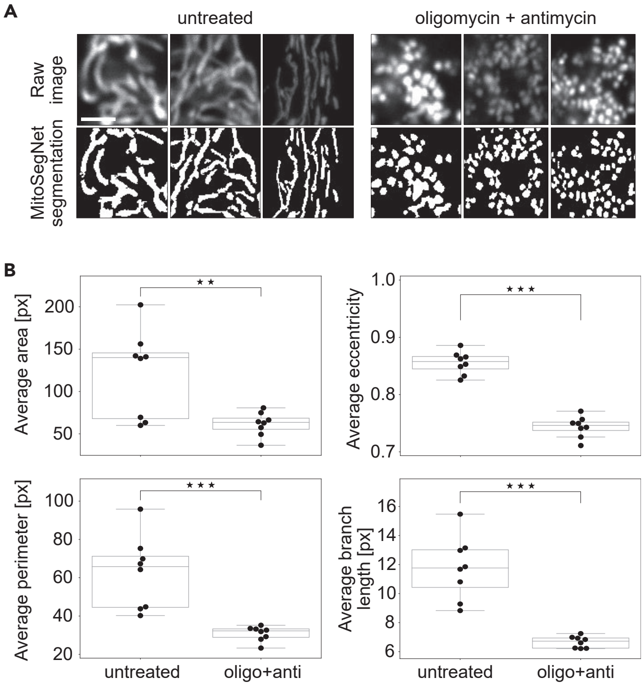


Figure 5. MitoSegNet Model Segmentation and Morphological Quantification Can Be Applied to Mitochondria of Mammalian Cells

Comparing untreated HeLa cells and HeLa cells treated with oligomycin or antimycin for a duration of 3 h.

(A) Visual comparison of untreated and treated mitochondrial morphology. Raw images are at the top and MitoSegNet model segmentations at the bottom. The scale bar is 2.5 μm.

(B) Average area, eccentricity, perimeter, and branch length of mitochondria were measured in segmented images of treated and untreated mitochondria. $*p < 0.05$, $***p < 0.001$ using an independent two-sample t test. N = 8.

in terms of mitochondrial branch morphology (Figure 4C). While the average mitochondrial branch length in *catp-6*(lf) mutants is larger than in wild-type (p = 0.01, independent two-sample t test), the average number of mitochondrial branches was found to be significantly smaller in *catp-6*(lf) mutants (p = 0.009, Mann-Whitney U test). Altogether these results indicate that *catp-6(ok3473)* causes a previously unreported mitochondrial morphology phenotype that cannot be detected by simple visual inspection.

MitoSegNet Segmentation of Mitochondria in HeLa Cells

To test the generalizability of our model even further, we applied the MitoSegNet to segment 8 confocal microscopy images depicting mitochondria in HeLa cells (Figure 5A). The fragmentation of mitochondria in HeLa cells treated with oligomycin and antimycin for 3 hr was captured in the segmentation both visually and quantitatively. As expected, the average mitochondrial area is significantly larger in untreated cells compared to treated HeLa cells (p = 0.0068, independent two-sample t test) (Figure 5B). The average eccentricity is lower for the fragmented mitochondria compared to the untreated mitochondria, indicating a more circular shape ($p = 1.32*10^{-8}$, independent two-sample t test) (Figure 5B). The average perimeter distribution reflects a similar pattern as found for the area, showing the fragmented mitochondria to have a smaller perimeter on average (p = 0.00037, independent two-sample t test) (Figure 5B). The average branch length is also significantly smaller in the treated mitochondria compared to the untreated mitochondria ($p = 1.30*10^{-5}$, independent two-sample t test) (Figure 5B).

The MitoS Segmentation and MitoA Analysis Tool

To enable non-experts, we implemented the MitoSegNet in an easy-to-use tool, the MitoS segmentation tool, a Python-based, standalone executable. MitoS can be executed in a basic mode, which utilizes the pretrained MitoSegNet for segmentation of mitochondria and allows us to easily apply the model without prior deep learning experience (Figure 1). We applied the MitoS image segmentation using the pretrained

MitoSegNet on two different systems. For all cases 10 images of each 1300 × 1030 px size (8-bit) were segmented. The MitoS GPU segmentation was run using an NVIDIA GeForce GTX 960M and NVIDIA TITAN X and segmentation took 65 and 15 s, respectively. Segmentation using the MitoS CPU version was performed on an Intel(R) Core(TM) i7-6700HQ CPU and a system using four Intel(R) Xeon(R) CPU E5-2680 v4 processors and lasted 7.5 min and 65 s, respectively. The basic mode also includes a fine-tuning module, which allows us to optimize the pretrained on new images (Figure S1A). The MitoS advanced mode can be used if other structures besides mitochondria should be segmented or if the user wishes to build a self-configured deep learning segmentation model (Figure S1B). The MitoA analyzer is a separate Python-based, standalone executable tool that can be run after successful segmentation for quantification and visualization of potential morphological differences (Figure S3). It measures ten different morphological and three intensity-based features for each object and summary statistics for all object features per image are generated. The tables of two or multiple samples containing these summary statistics can then be subjected to hypothesis testing, visualization, and correlation analysis. The MitoS and MitoA tools require no installation and no prerequisite installations (such as frameworks), and they are available for both Windows and Linux.

DISCUSSION

We present MitoSegNet, a segmentation model that exploits the power of deep learning to address the challenging problem of accurate mitochondria segmentation. We show that the MitoSegNet outperforms feature-based, non-deep learning-based algorithms and that it is generalizable to unseen images from *C. elegans* and mammalian cells.

MitoSegNet Model Segmentation Performance

While the superior visual and quantitative performance of the MitoSegNet model segmentation might not come as a surprise to researchers acquainted with the capabilities of deep learning-based segmentation, we believe these results to be interesting to researchers who commonly use feature-based segmentation methods. For accuracy evaluation, we did not rely only on pixel-based accuracy, as we found this to be an insufficient measure of morphological accuracy but extended our analysis. Our single object shape comparison as well as the calculation of energy distances for five different feature descriptors per image demonstrate that the MitoSegNet model segmentation yields the best morphological accuracy compared with commonly applied segmentation methods.

Comparison of Mitochondrial Morphology between *Catp-6*^{ATP13A2} Mutants and Wild-type

The visual comparison of mitochondrial morphology in wild type and *catp-6*(lf) mutants did not reveal any obvious differences in morphology. In both cases, mitochondria appeared to be largely tubular. However, the quantitative analysis using the MitoSegNet revealed that average area, perimeter, and minor axis length of mitochondria in *catp-6*(lf) mutants are smaller than in wild-type. Since the average number of mitochondria per cell is the same as in wild-type, we concluded that *catp-6*(lf) causes a reduction in overall mitochondrial mass. This is consistent with observations of Anand et al., who found that oxygen consumption is lower in *C. elegans catp-6*(lf) mutants than in wild type (Anand et al., 2020). In mammalian cells, mutation of *ATP13A2* has been found to cause impairment of mitochondrial function and induce mitochondrial fragmentation (Gusdon et al., 2012; Ramonet et al., 2012). However, this is associated with increased mitochondrial mass due to inefficient autophagic clearance (Grünewald et al., 2012). The differences between the mutant phenotypes observed in *C. elegans* and mammalian cells is likely to reflect the multiple roles of ATP13A2 and CATP-6 transport substrates (polyamines) in maintaining mitochondrial function (van Veen et al., 2020). Further research on properties of *catp-6* deficient mitochondria, such as membrane potential or levels of reactive oxygen species, might uncover the cause for the observed differences in mitochondrial morphology.

Application of the MitoSegNet to Mitochondria in HeLa Cells

Although the pretrained model was generated with standard fluorescence microscopy images, depicting mitoGFP-labeled mitochondria in *C. elegans*, the same pretrained model was able to generate visually accurate segmentations of mitoRFP labeled mitochondria in HeLa cells using a laser scanning confocal microscope. This demonstrates the high robustness and generalization capabilities of our pretrained MitoSegNet model and that it can be used for segmentation of mitochondria in organisms other than *C. elegans*. Furthermore, our MitoSegNet Analysis tool quantitatively confirmed the morphological differences of mitochondria between untreated HeLa cells and HeLa cells treated with oligomycin or antimycin.

MitoSegNet Model Architecture

Although the MitoSegNet architecture is largely based on the U-Net, through testing various changes in the original architecture, we found that the validation dice coefficient as well as the validation loss improved upon removing the dropout layers and instead placing a batch normalization layer (Ioffe and Szegedy, 2015) after every convolution layer in the contracting pathway. Interestingly, a recent study found that the combined usage of batch normalization followed by dropout (forming an independent component layer) stabilized the training process, increased convergence speed, and improved the convergence limit (Chen et al., 2019). It would require further testing to find out if the usage of an independent component layer would improve the current MitoSegNet performance even further.

MitoS and MitoA Tools

Most deep learning applications in the field of biological image segmentation were created for the purpose of 2D cell segmentation (Chen et al., 2017; Al-Kofahi et al., 2018; Falk et al., 2019; Kusumoto and Yuasa, 2019), while organelle-specific deep learning applications are scarce. Although most tools allow the user to retrain available 2D cell segmentation models to segment other biological structures of interest, this often requires computer science-related skills, such as familiarity with programming languages, shell interaction or knowledge on how to install various deep learning frameworks. One of the main motivations behind MitoSegNet and the MitoS and MitoA toolbox was to make deep learning segmentation accessible to researchers that do not have an extensive background in computer science or deep learning. The MitoS tool can be run without installation. The simple graphical user interface allows users to quickly navigate the MitoS and MitoA tools. The MitoS basic mode also comes with a fine-tuning module that allows researchers that would like to segment other organelles or images taken under different conditions than those used for training the MitoSegNet model. Since the subsequent step after segmentation is usually the analysis, we included the MitoA tool to save researchers the time to look up appropriate analysis tools and instead be able to quickly obtain potentially interesting insights.

Conclusion

The MitoSegNet has been shown to outperform both conventional feature-based and machine-learning-based segmentation of mitochondria. The pretrained model can be easily applied to new 2D microscopy images of mitochondria through the usage of the MitoS tool, which is available for both standard and high-end Windows and Linux systems. Successfully segmented images of mitochondria can be subjected to quantification, statistical testing, and visualization with the MitoA tool.

Limitations of the Study

The MitoSegNet model used for segmentation of images depicting fluorescently labeled mitochondria in *C. elegans* and HeLa cells was trained with 12 pairs of raw images and the appendant hand-labeled ground truth images. Although both the visual and quantitative segmentation accuracy was shown to be high in this study, there remains a bias which is based on the two annotators who generated 6 ground truth images each. To reduce the ground truth bias and increase the generalizability of the MitoSegNet model, more images labeled by different annotators can be added. Furthermore, image sections in which single mitochondria were not clearly distinguishable due to optical constraints of the microscope used to generate the image, made it difficult for the annotator to create labeled images that accurately represent the ground truth. This uncertainty introduced to the MitoSegNet model can be decreased by adding images recorded with higher resolution, thus reducing the sections in which such visually indistinguishable mitochondria exist.

Resource Availability

Lead Contact

Further information and requests for resources should be directed to and will be fulfilled by the Lead Contact, Carsten Marr (carsten.marr@helmholtz-muenchen.de).

Materials Availability

Images used for training and testing the model are available upon request.

Data and Code Availability

The software documentation for the MitoS and MitoA tool can be found at https://github.com/mitosegnet. The MitoSegNet segmentation model, the MitoA analysis and MitoS segmentation tool (GPU/CPU) for

Linux and Windows are available at https://zenodo.org/search?page=1&size=20&q=mitosegnet. The Python code used for generating the figures displayed in the manuscript is available at https://github.com/MitoSegNet/MitoSegNet_AccuracyTesting_Manuscript.

METHODS

All methods can be found in the accompanying Transparent Methods supplemental file.

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.isci.2020.101601.

ACKNOWLEDGMENTS

We thank E. Lambie, H. Hartz and members of the Conradt and Marr lab for comments on the project development, the manuscript, and testing the MitoS and MitoA tools. We also thank Iva Dzhilyanova for testing the MitoS tool on multiple systems. This work was supported by the Deutsche Forschungsgemeinschaft (Center for Integrated Protein Science Munich [CIPSM; EXC 114], CO204/6-1 and CO204/9-1). C.M. has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (Grant agreement No. 866411). Some strains were provided by the CGC, which is funded by NIH Office of Research Infrastructure Programs (P40 OD010440).

AUTHOR CONTRIBUTIONS

C.A.F. designed and conducted the computational experiments, performed the data analysis, and wrote the paper. B.C. designed the biological experiments and wrote the paper. C.M. designed the computational experiments and wrote the paper. S.G.R., S.H., L.B.C., M.D. and K.S. designed and conducted the biological experiments.

DECLARATION OF INTERESTS

None.

Received: February 7, 2020
Revised: August 18, 2020
Accepted: September 17, 2020
Published: October 23, 2020

REFERENCES

Abdolhoseini, M., Kluge, M.G., Walker, F.R., and Johnson, S.J. (2019). Segmentation, tracing, and quantification of microglial cells from 3D image stacks. Sci. Rep. *9*, 8557.

Akram, F., Garcia, M.A., and Puig, D. (2017). Active contours driven by difference of Gaussians. Sci. Rep. *7*, 14984.

Al-Kofahi, Y., Zaltsman, A., Graves, R., Marshall, W., and Rusu, M. (2018). A deep learning-based algorithm for 2-D cell segmentation in microscopy images. BMC Bioinformatics *19*, 365.

Anand, N., Holcom, A., Broussalian, M., Schmidt, M., Chinta, S.J., Lithgow, G.J., Andersen, J.K., and Chamoli, M. (2020). Dysregulated iron metabolism in C. elegans catp-6/ATP13A2 mutant impairs mitochondrial function. Neurobiol. Dis. *139*, 104786.

Berg, S., Kutra, D., Kroeger, T., Straehle, C.N., Kausler, B.X., Haubold, C., Schiegg, M., Ales, J., Beier, T., Rudy, M., et al. (2019). Ilastik: interactive machine learning for (bio)image analysis. Nat. Methods *16*, 1226–1232.

van der Bliek, A.M., Sedensky, M.M., and Morgan, P.G. (2017). Cell biology of the mitochondrion. Genetics *207*, 843–871.

de Boer, R., Smith, R.L., De Vos, W.H., Manders, E.M., Brul, S., and van der Spek, H. (2015). Caenorhabditis elegans as a model system for studying drug induced mitochondrial toxicity. PLoS One *10*, e0126220.

Breckenridge, D.G., Kang, B.-H., Kokel, D., Mitani, S., Staehelin, L.A., and Xue, D. (2008). Caenorhabditis elegans drp-1 and fis-2 regulate distinct cell-death execution pathways downstream of ced-3 and independent of ced-9. Mol. Cell *31*, 586–597.

Chan, D.C. (2020). Mitochondrial dynamics and its involvement in disease. Annu. Rev. Pathol. Mech. Dis. *15*, 235–259.

Chen, M., Dai, W., Sun, S.Y., Jonasch, D., He, C.Y., Schmid, M.F., Chiu, W., and Ludtke, S.J. (2017). Convolutional neural networks for automated annotation of cellular cryo-electron tomograms. Nat. Methods *14*, 983–985.

Chen, G., Chen, P., Shi, Y., Hsieh, C.Y., Liao, B., and Zhang, B. (2019). Rethinking the usage of batch normalization and dropout in the training of deep neural networks. arXiv, arxiv:1905.05928.

Chlebus, G., Schenk, A., Moltz, J.H., van Ginneken, B., Hahn, H.K., and Meine, H. (2018). Automatic liver tumor segmentation in CT with fully convolutional neural networks and object-based postprocessing. Sci. Rep. *8*, 15497.

Falk, T., Mai, D., Bensch, R., Cicek, O., Abdulkadir, A., Marrakchi, Y., Bohm, A., Deubner, J., Jackel, Z., Seiwald, K., et al. (2019). U-Net: deep learning for cell counting, detection, and morphometry. Nat. Methods *16*, 67–70.

De Fauw, J., Ledsam, J.R., Romera-Paredes, B., Nikolov, S., Tomasev, N., Blackwell, S., Askham, H., Glorot, X., O'Donoghue, B., Visentin, D., et al. (2018). Clinically applicable deep learning for diagnosis and referral in retinal disease. Nat. Med. *24*, 1342–1350.

Di Fonzo, A., Chien, H.F., Socal, M., Giraudo, S., Tassorelli, C., Iliceto, G., Fabbrini, G., Marconi, R., Fincati, E., Abbruzzese, G., et al. (2007). ATP13A2 missense mutations in juvenile parkinsonism and young onset Parkinson disease. Neurology *68*, 1557–1562.

Grünewald, A., Arns, B., Seibler, P., Rakovic, A., Münchau, A., Ramirez, A., Sue, C.M., and Klein, C. (2012). ATP13A2 mutations impair mitochondrial function in fibroblasts from patients with Kufor-Rakeb syndrome. Neurobiol. Aging *33*, 1843.e1–1843.e7.

Gusdon, A.M., Zhu, J., Van Houten, B., and Chu, C.T. (2012). ATP13A2 regulates mitochondrial bioenergetics through macroautophagy. Neurobiol. Dis. *45*, 962–972.

Ichishita, R., Tanaka, K., Sugiura, Y., Sayano, T., Mihara, K., and Oka, T. (2008). An RNAi screen for mitochondrial proteins required to maintain the morphology of the organelle in Caenorhabditis elegans. J. Biochem. *143*, 449–454.

Ioffe, S., and Szegedy, C. (2015). Batch normalization: accelerating deep network training by reducing internal covariate shift. In Proceedings of the 32nd International Conference on International Conference on Machine Learning - Volume 37Proceedings of the 32nd International Conference on International Conference on Machine Learning - Volume 37, pp. 448–456.

Kanazawa, T., Zappaterra, M.D., Hasegawa, A., Wright, A.P., Newman-Smith, E.D., Buttle, K.F., McDonald, K., Mannella, C.A., and van der Bliek, A.M. (2008). The C. elegans Opa1 homologue EAT-3 is essential for resistance to free radicals. PLoS Genet. *4*, e1000022.

Kreshuk, A., and Zhang, C. (2019). Machine learning: advanced image segmentation using ilastik. Methods Mol. Biol. *2040*, 449–463.

Krizhevsky, A., Sutskever, I., and Hinton, G.E. (2012). ImageNet classification with deep convolutional neural networks. In Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1 (Curran Associates Inc.), pp. 1097–1105.

Kusumoto, D., and Yuasa, S. (2019). The application of convolutional neural network to stem cell biology. Inflamm. Regen. *39*, 14.

Labrousse, A.M., Zappaterra, M.D., Rube, D.A., and van der Bliek, A.M. (1999). C. elegans dynamin-related protein DRP-1 controls severing of the mitochondrial outer membrane. Mol. Cell *4*, 815–826.

Lambie, E.J., Tieu, P.J., Lebedeva, N., Church, D.L., and Conradt, B. (2013). CATP-6, a C. elegans ortholog of ATP13A2 PARK9, positively regulates GEM-1, an SLC16A transporter. PLoS One *8*, e77202.

LeCun, Y., Boser, B., Denker, J.S., Henderson, D., Howard, R.E., Hubbard, W., and Jackel, L.D. (1989). Backpropagation applied to handwritten zip code recognition. Neural Comput. *1*, 541–551.

Li, Y., Gong, H., Wu, W., Liu, G., and Chen, G. (2015). An automated method using hessian matrix and random walks for retinal blood vessel segmentation. 2015 8th International Congress on Image and Signal Processing (CISP).

Lindeberg, T. (1994). Scale-space theory: a basic tool for analysing structures at different scales. J. Appl. Stat. *21*, 225–270.

Long, J., Shelhamer, E., and Darrell, T. (2014). Fully convolutional networks for semantic segmentation. arXiv, arxiv:1411.4038.

Marr, D., and Hildreth, E. (1980). Theory of edge detection. Proc. R. Soc. Lond. B Biol. Sci. *207*, 187–217.

McQuin, C., Goodman, A., Chernyshev, V., Kamentsky, L., Cimini, B.A., Karhohs, K.W., Doan, M., Ding, L., Rafelski, S.M., Thirstrup, D., et al. (2018). CellProfiler 3.0: next-generation image processing for biology. PLoS Biol. *16*, e2005970.

Meijering, E. (2012). Cell segmentation: 50 Years down the road [life sciences]. IEEE Signal Process. Mag. *29*, 140–145.

Moller, J.V., Juul, B., and le Maire, M. (1996). Structural organization, ion transport, and energy transduction of P-type ATPases. Biochim. Biophys. Acta *1286*, 1–51.

Orozco-Fuentes, S., Neganova, I., Wadkin, L.E., Baggaley, A.W., Barrio, R.A., Lako, M., Shukurov, A., and Parker, N.G. (2019). Quantification of the morphological characteristics of hESC colonies. Sci. Rep. *9*, 17569.

Pernas, L., and Scorrano, L. (2016). Mito-morphosis: mitochondrial fusion, fission, and cristae remodeling as key mediators of cellular function. Annu. Rev. Physiol. *78*, 505–531.

Ramirez, A., Heimbach, A., Grundemann, J., Stiller, B., Hampshire, D., Cid, L.P., Goebel, I., Mubaidin, A.F., Wriekat, A.L., Roeper, J., et al. (2006). Hereditary parkinsonism with dementia is caused by mutations in ATP13A2, encoding a lysosomal type 5 P-type ATPase. Nat. Genet. *38*, 1184–1191.

Ramonet, D., Podhajska, A., Stafa, K., Sonnay, S., Trancikova, A., Tsika, E., Pletnikova, O., Troncoso, J.C., Glauser, L., and Moore, D.J. (2012). PARK9-associated ATP13A2 localizes to intracellular acidic vesicles and regulates cation homeostasis and neuronal integrity. Hum. Mol. Genet. *21*, 1725–1743.

Regmi, S.G., and Rolland, S.G. (2017). New imaging tools to analyze mitochondrial morphology in Caenorhabditis elegans. Methods Mol. Biol. *1567*, 255–272.

Rolland, S.G., Lu, Y., David, C.N., and Conradt, B. (2009). The BCL-2-like protein CED-9 of C. elegans promotes FZO-1/Mfn1,2- and EAT-3/Opa1-dependent mitochondrial fusion. J. Cell Biol. *186*, 525–540.

Rolland, S.G., Motori, E., Memar, N., Hench, J., Frank, S., Winklhofer, K.F., and Conradt, B. (2013). Impaired complex IV activity in response to loss of LRPPRC function can be compensated by mitochondrial hyperfusion. Proc. Natl. Acad. Sci. U S A *110*, E2967–E2976.

Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: Convolutional Networks for Biomedical Image Segmentation, Cham (Springer International Publishing).

Sato, Y., Nakajima, S., Shiraga, N., Atsumi, H., Yoshida, S., Koller, T., Gerig, G., and Kikinis, R. (1998). Three-dimensional multi-scale line filter for segmentation and visualization of curvilinear structures in medical images. Med. Image Anal. *2*, 143–168.

Scherer, D., Müller, A., and Behnke, S. (2010). Evaluation of Pooling Operations in Convolutional Architectures for Object Recognition (Springer Berlin Heidelberg).

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al. (2012). Fiji: an open-source platform for biological-image analysis. Nat. Methods *9*, 676–682.

Stember, J.N., Chang, P., Stember, D.M., Liu, M., Grinband, J., Filippi, C.G., Meyers, P., and Jambawalikar, S. (2018). Convolutional neural networks for the detection and measurement of cerebral aneurysms on magnetic resonance angiography. J. Digit. Imaging *32*, 808–815.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S.E., Anguelov, D., Erhan, D., Vanhoucke, V., and Rabinovich, A. (2014). Going deeper with convolutions. arXiv, arxiv:1409.4842.

Taha, A.A., and Hanbury, A. (2015). Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool. BMC Med. Imaging *15*, 29.

Tan, F.J., Husain, M., Manlandro, C.M., Koppenol, M., Fire, A.Z., and Hill, R.B. (2008). CED-9 and mitochondrial homeostasis in C. elegans muscle. J. Cell Sci. *121*, 3373–3382.

Tilokani, L., Nagashima, S., Paupe, V., and Prudent, J. (2018). Mitochondrial dynamics: overview of molecular mechanisms. Essays Biochem. *62*, 341–360.

Tondera, D., Grandemange, S., Jourdain, A., Karbowski, M., Mattenberger, Y., Herzig, S., Da Cruz, S., Clerc, P., Raschke, I., Merkwirth, C., et al. (2009). SLP-2 is required for stress-induced mitochondrial hyperfusion. EMBO J. *28*, 1589–1600.

Torborg, C.L., and Feller, M.B. (2004). Unbiased analysis of bulk axonal segregation patterns. J. Neurosci. Methods *135*, 17–26.

van Veen, S., Martin, S., Van den Haute, C., Benoy, V., Lyons, J., Vanhoutte, R., Kahler, J.P., Decuypere, J.-P., Gelders, G., Lambie, E., et al. (2020). ATP13A2 deficiency disrupts lysosomal polyamine export. Nature *578*, 419–424.

Wai, T., and Langer, T. (2016). Mitochondrial dynamics and metabolic regulation. Trends Endocrinol. Metab. *27*, 105–117.

Wiesel, T.N., and Hubel, D.H. (1963). Single-cell responses in striate cortex of kittens deprived of vision in one eye. J. Neurophysiol. *26*, 1003–1017.

iScience, Volume *23*

Supplemental Information

MitoSegNet: Easy-to-use Deep Learning Segmentation for Analyzing Mitochondrial Morphology

Christian A. Fischer, Laura Besora-Casals, Stéphane G. Rolland, Simon Haeussler, Kritarth Singh, Michael Duchen, Barbara Conradt, and Carsten Marr

Figure S1

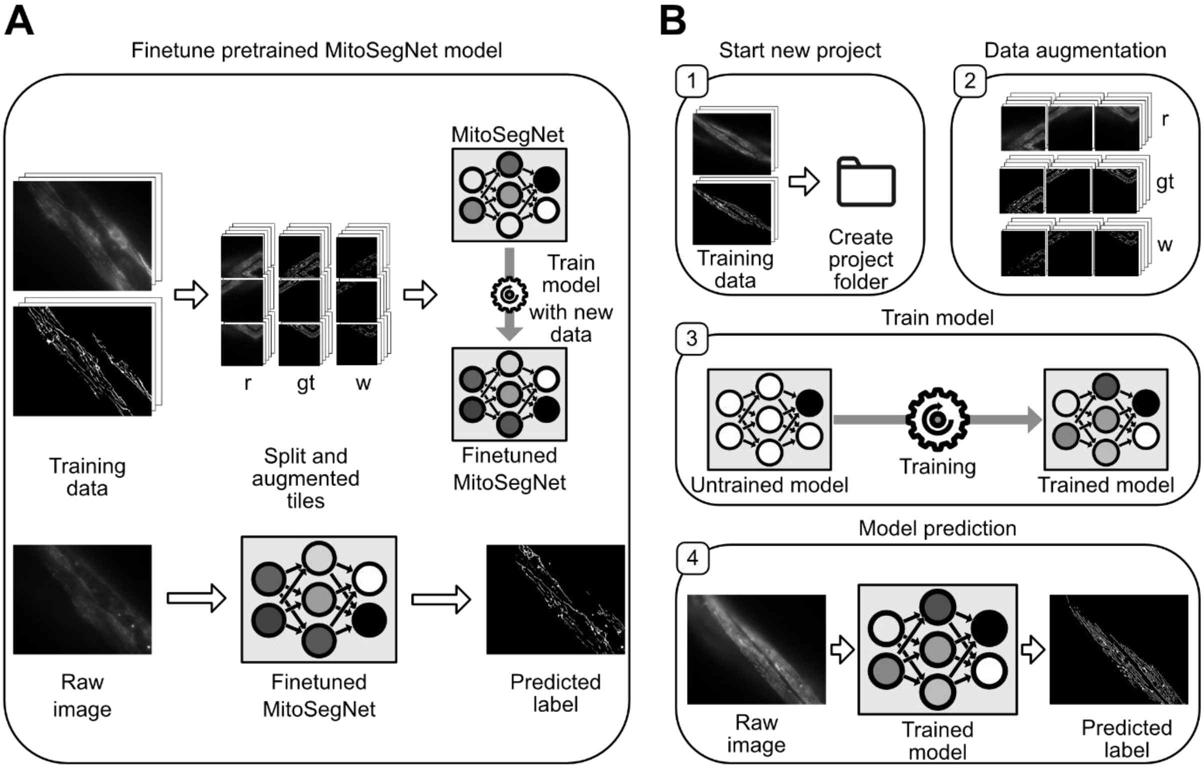


Figure S1: MitoS segmentation tool finetuning module and advanced mode workflow, Related to Figure 1 **(A)** The finetuning module enables the user to add new training data to finetune the existing MitoSegNet. This function automatically splits the new training data into tiles, performs augmentation (r: raw images, gt: ground truth images, w: weight map images) and the augmented data is then used to train the pretrained MitoSegNet. After completion of training, the finetuned MitoSegNet can be used for segmentation of new images using the prediction function. **(B)** The advanced mode allows the user to create a new deep learning segmentation model by adjusting network parameters. [1] Generation of a new project folder, in which all subsequent advanced functions will be carried out. [2] Data augmentation parameters can be specified and users can decide whether weight maps should be generated. [3] Model training allows the user to set the learning rate, batch size, class balance weight factor and if weight maps should be used. [4] Once the model has been trained the prediction function of the MitoS tool can be used to predict the segmentation on previously unseen images.

Figure S2

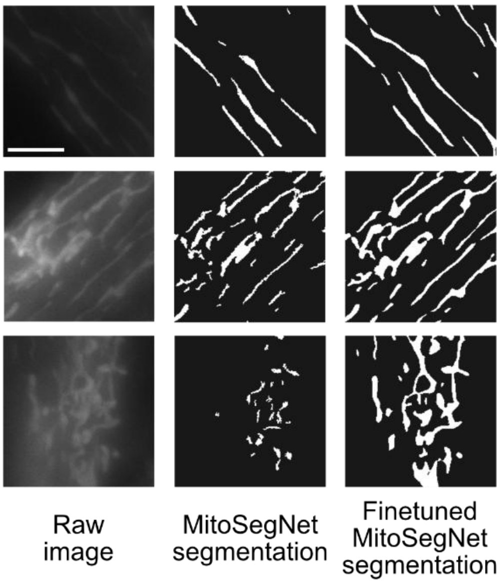


Figure S2: Finetuning MitoSegNet improves visual segmentation results, Related to Figure 2. Due to the usage of a non-integrated $P_{myo\text{-}3}$::mitoGFP reporter, the intensity of fluorescently labelled mitochondria appeared weaker and hence the MitoSegNet model failed to accurately segment these images. The input images are shown on the left, in the middle are the binary masks before finetuning and to the right the same masks after finetuning. By segmenting one image by hand and finetuning the pretrained segmentation model for 10 epochs, the visual segmentation results could be largely improved. The scale bar is 5µm.

Figure S3

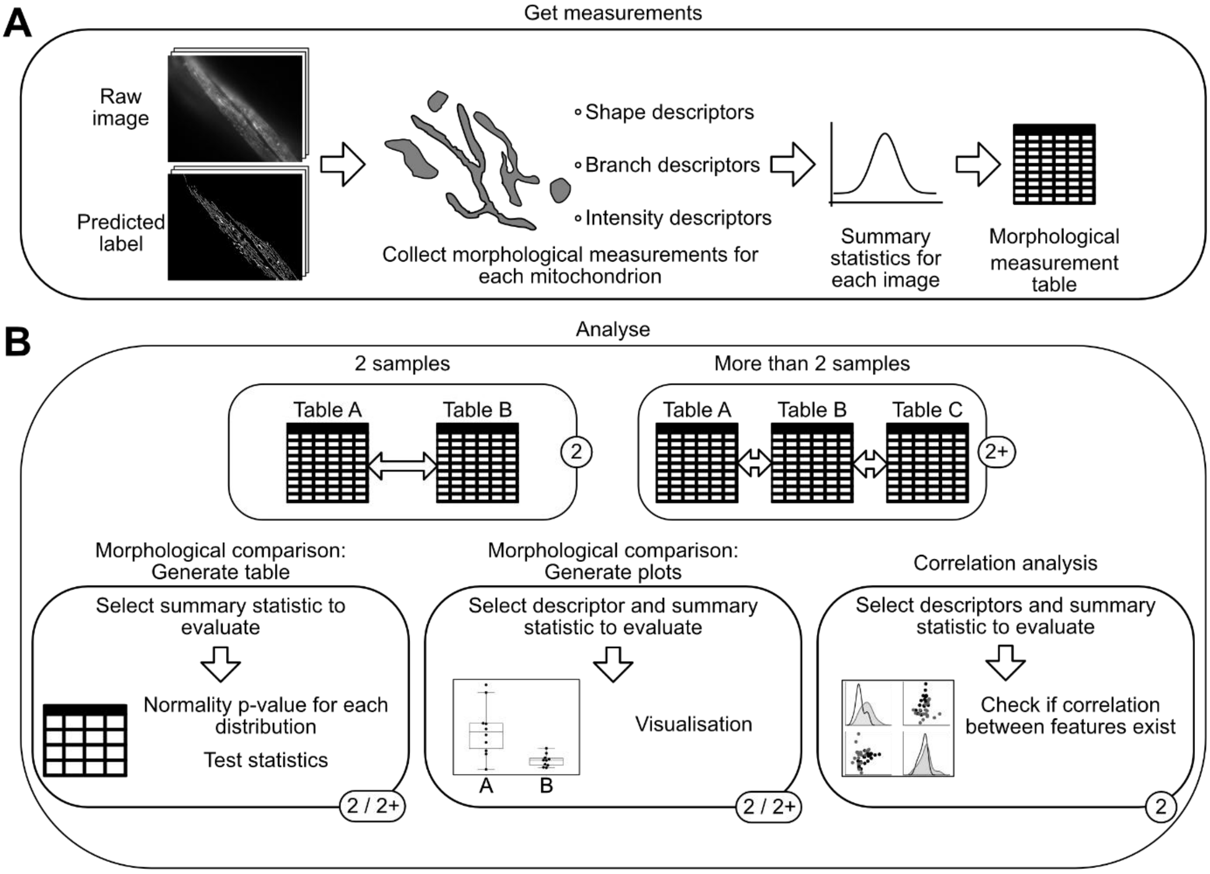


Figure S3: MitoA analyser tool workflow, Related to Figure 4 and 5. The two main functions of the MitoA tool are to get measurements or to analyse. To analyse images of interest, one has to first get the relevant measurements. **(A)** The “Get measurements” function uses images and their predicted labels to measure 6 different shape descriptors, 4 branch descriptors and the mean, maximum and minimum fluorescence intensity for each segmented object. The summary statistics are then saved to a measurements table. **(B)** Analysis can either be performed on only two samples (2) or more than two samples (2+). Subsequent statistical analysis and visualisation can be used in both cases but the correlation analysis is currently only implemented for the two sample comparison. The generate table function allows the user to select a summary statistic to evaluate (such as average, median or standard deviation) after which a table is generated in which sample distribution normality is tested and an appropriate statistical test is selected. The generate plots function lets the user select a feature descriptor and summary statistic to display as a boxplot. With the correlation analysis, up to 4 different feature descriptors from two samples can be visualised as scatterplots, including the display of the correlation coefficients.

Figure S4

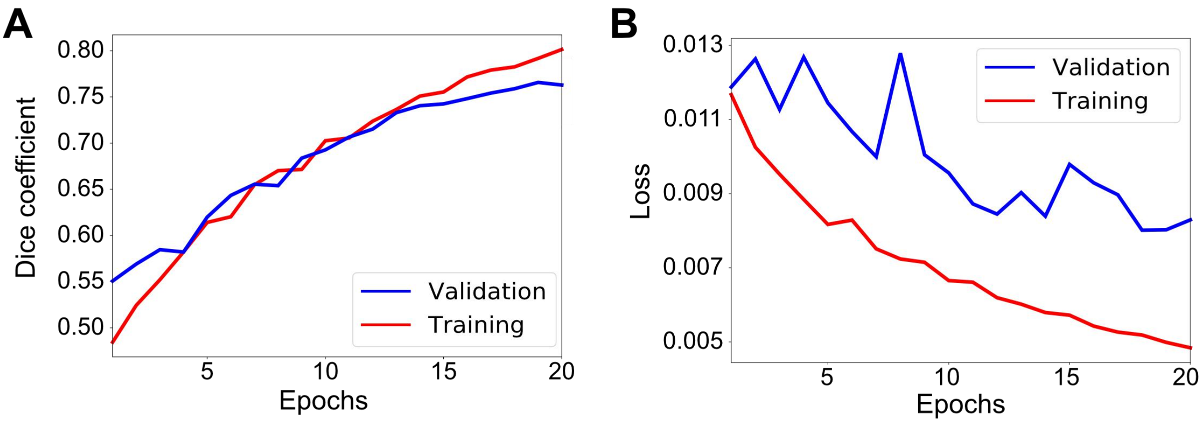


Figure S4: Training performance of final MitoSegNet trained on 12 images, related to Figure 1. (A) Training (red) and validation (blue) dice coefficient increase much like during cross validation and after epoch 15 discrepancy between training and validation dice coefficient begins to increase. **(B)** Training and validation loss decrease and reaches its minimum of 0.008 at 18 epochs.

Figure S5

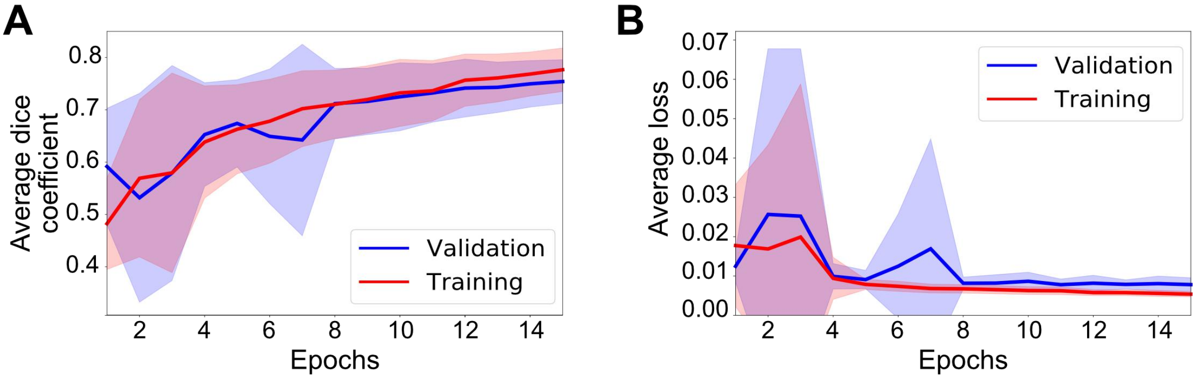


Figure S5: Average training performance of 12 different MitoSegNet models trained on 11 images for cross validation, Related to Figure 1. (A) Average validation (blue) and training (red) dice coefficient steadily increases over 15 epochs of training. **(B)** Average training loss decreases as expected, while validation loss enters a plateau after 8 epochs. Blue and red areas indicate the standard deviation added (upper border) or subtracted (lower border) from the average of the validation and training metric respectively.

Transparent Methods

General C. elegans methods and strains

C. elegans strains were cultured on Nematode Growth Medium (NGM) in petri plates (Brenner 1974). All experiments were carried out at 20°C and all strains were maintained at 20°C. *C. elegans* mitochondria in body wall muscle cells were labelled using an integrated transgene (*bcI78*) expressing mitochondrial matrix-targeted GFP under the control of the body wall muscle-specific promoter *myo-3* (P_{myo3}*::mitoGFP*) (Rolland, Motori et al. 2013). The following alleles were used LGII: *fzo-1(tm1133)* (National BioResource Project); *eat-3(ad426)* (Kanazawa, Zappaterra et al. 2008); LGIV: *drp-1(tm1108)* (National BioResource Project) and *catp-6(ok3473)* (Lambie, Tieu et al. 2013).

Generating the MitoSegNet model

The MitoSegNet model was trained for 20 epochs with 12 fluorescence microscopy images (1300 x 1030 px), depicting adult *C. elegans* mitochondria (tubular, elongated, fragmented and mixed) in body wall muscle cells labelled with mitoGFP (Fig. 1A). The images were recorded on a fluorescence microscope equipped with a 63x 1.4 NA oil lens (Axioskop 2; Carl Zeiss Inc.) and a charge-coupled device camera (1300; Micromax). Two expert annotators manually segmented a set of 6 microscopy images. The image annotations were generated by drawing separate regions of interest over all mitochondria within one image and then converting the selections into a binary mask. Raw 3D microscopy images were converted to 2D images through maximum intensity projection. Based on the expert annotation of four mitochondrial phenotypes (elongated, fragmented, mixed and tubular), 3 images of each phenotype were used for training. Prior to augmentation each image was divided into four overlapping 656 x 656 px images due to GPU memory constraints. The input tile size was chosen to allow for the 2x2 max-pooling operations during training. Each image (and the corresponding ground truth) underwent random augmentation using the Keras library, producing 80 differently augmented images per input image, which involved random shearing (in a range of 30%), rotations (in a range of 180°), change of zoom (in a range of 30%), brightness change (in a range of 20%), horizontal and vertical flipping, x- and y-shifts (in a range of 20%) and mirroring. This increased the total number of images used for training from 48 to 3,840. We used a batch size of 1 and included the weight map as implemented in the U-Net publication and set w_{bal} to 0.042 (1 dived by the number of background pixels per object pixel to achieve a 1:1 ratio between object and background pixels). The MitoSegNet model was trained at a learning rate of $7*10^{-5}$ for 20 epochs and reached its minimum validation loss at epoch 18 (Fig. S4B).

MitoSegNet model prediction

To evaluate the segmentation performance of the MitoSegNet model on all 12 images, we performed a cross validation and generated 12 separately trained models (in the same manner as described above) (Fig. S5). Each model was trained with 11 images for 15 epochs and applied each trained MitoSegNet model to the test image that was excluded from training (Fig. 1B). Training and prediction were performed using our Python-based MitoS tool on a Devuan GNU/Linux server with an Nvidia TITAN X (12 GB GDDR5 RAM), using CUDA 10.0 and cuDNN 7.6.1.

The MitoS segmentation and MitoA analysis tool

The MitoS tool is a Python-based, standalone executable application that can be used for deep learning based segmentation of 2-dimensional microscopy images. The tool can be executed in two modes: the “basic mode” is designed for researchers with no prior deep learning experience (Fig. 1), the “advanced mode” is intended for people with such experience and allows customizing training parameters (Fig. S1). For those users that do not have a CUDA-capable GPU, a CPU-only version of the MitoS segmentation tool is available.

The MitoS basic mode

uses the pretrained mitochondria-specific MitoSegNet model and any images depicting mitochondria images similar to those used for training can be segmented with the MitoS tool (Fig. 1). If the MitoSegNet model segmentation does not yield satisfactory results, self-generated training data can be used to finetune the existing segmentation model (Fig. S1A). Because the MitoSegNet model was generated using 656x656 px image tiles, any images intended for segmentation are fitted to this pre-set tile size. Images larger than the pre-set tile size are split into overlapping tiles with a mirrored border to avoid prediction in border regions. Smaller images are not split but instead a mirrored border is added to increase the size to the pre-set tile size. After prediction the mirrored parts of the tiles are removed and the tiles are stitched back together. A final threshold is applied to the fully stitched image to convert it to an 8-bit binary mask ($p_i < 128$ to 0 and $p_i >= 128$ to 255). Furthermore, any objects below 10 px in size are considered to be noise and are thus removed from the final segmentation image. The finetuning module automatically generates weight maps and augments the novel training data using the following augmentation operations: shearing (in a range of 30%), rotation (in range of 180°), zoom (in a range of 30%) and brightness change (in a range of 20%), horizontal and vertical flip, width and height shift (in a range of 20%). Once augmentation is completed, model training begins with a pre-set learning rate of 0.0001, a batch size of 1 and a class balance factor of 1/(foreground to background pixel ratio). We tested the MitoS finetuning module on images depicting adult *C. elegans* mitochondria visualised with a non-integrated $P_{myo\text{-}3}$*::mitoGFP* reporter (Rolland, Motori et al. 2013). The intensity of the fluorescently labelled mitochondria appeared weaker and hence the MitoSegNet model failed to accurately segment these images (Fig. S2). One image of weak fluorescent intensity mitochondria was therefore manually segmented and used to finetune the pretrained model for 10 epochs, generating a finetuned MitoSegNet model that generated satisfactory segmentation results (Fig. S2). Alternatively, it is possible to create a custom segmentation model using the advanced mode of the MitoS tool (Fig. S1B). The advanced mode consists of four modules: Module 1 automatically generates a folder structure that is later used for model training. Module 2 enables data augmentation, which allows the user to specify the range of the above listed augmentation operations. Module 3 carries out the model training for which the parameters can be freely chosen. Module 4 predicts new segmentations using the fully trained model (Fig. S2).

The MitoA analyser tool is a separate Python-based application that can be run after successful segmentation for quantification and visualisation of potential morphological differences (Fig. S3). It measures 13 different features using both raw and segmented image for each object: area, minor and major axis length, eccentricity, perimeter, solidity, mean, max and min intensity, number of branches, branch length, total branch length and curvature index. The minor or major axis lengths are defined as the lengths of the line segment connecting the two co-vertices or vertices of an ellipse fitted around an object. The eccentricity of a conic section is a non-negative real number that characterizes its shape and the eccentricity for a circle, ellipse and parabola are 0, $0 > x > 1$ or 1 respectively. The solidity is defined as the object area divided by the convex object area. The curvature index indicates if the branches of the mitochondria are straight or curved and is defined as

$$c_i = \frac{branch\ length - euclidean\ distance}{euclidean\ distance}$$

The Euclidean distance is the shorted path between the start and end point of one branch. The single object features are summarised for each image as average, median, standard deviation, standard error, minimum and maximum and saved in an Excel table. One table is generated for a group of images, which can subsequently be compared. Both the two- and multi-comparison of samples allows to generate a table containing statistically relevant information and a visualisation using boxplots. Each statistical table contains the descriptors p-value of the D'Agostino's K-squared test to determine if the data is normally distributed (D'Agostino 1971, D'Agostino and Pearson 1973). Based on this result and the Levene's test p-value to test for equality of variances an appropriate hypothesis test is chosen and the name of the test, including its p-value are displayed in the table (Levene 1960). For comparison of two normally distributed datasets with equal variance a Student's t-test is used . If either of the two criteria are not met, the Mann–Whitney U test is used. For comparison of multiple samples with equal variance and normal distributions, a one-way ANOVA is used or alternatively, the Kruskal-Wallis-Test is applied (Kruskal and Wallis 1952). For a two-sample comparison the effect size is also calculated, and the boxplot visualisation indicates statistically significant differences. Furthermore, the two-sample comparison also enables the user to perform a correlation analysis. The correlation analysis visualises correlation by letting the user select up to four different descriptors to correlate against each other in both samples and then have them displayed as scatterplots. The distributions are subjected to the D'Agostino's K-squared test to determine the usage of the Pearson or Spearman rank-order correlation. Correlation coefficients and p-values are displayed in the MitoA terminal or can be saved to an Excel file. The MitoSegNet Analyser was used for statistical testing and visualisation of morphological differences for Figure 4 and 5.

Both tools have been tested internally on Windows and Linux platforms and will be actively maintained on GitHub.

The MitoSegNet architecture

The MitoSegNet architecture is based on the previously published U-Net (Ronneberger, Fischer et al. 2015) and was implemented in Keras (Python 3.7.3). It consists of a contracting path, which follows the standard architecture of convolutional neural networks, consisting of repeated application of 3x3 convolutions (padded convolutions), each followed by a batch normalization layer, a rectified linear unit (ReLU) and a 2x2 max pooling operation with a stride of 2. Unlike the original U-Net, the MitoSegNet does not utilize any drop-out layers at the end of the contracting path as we have found the batch normalization layer to reduce the training time. After 1024 feature channels have been generated, the expanding pathway uses 2x2 up-convolutions to halve the number of channels, followed by a concatenation with the corresponding feature map from the contracting path and subsequent 3x3 convolutions, followed by a ReLU. The final convolutional layer (1x1) is followed by a sigmoid function. In total, the MitoSegNet consists of 24 convolutional layers. The optimization algorithm chosen for the training process is Adam (adaptive moment estimation), which is an extension to stochastic gradient descent (Diederik P. Kingma 2014). During training, 20% of the data is excluded from the process and instead used for model validation after each training epoch. The energy function is computed by a pixel-wise sigmoid function over the final feature map combined with a binary cross entropy loss function

$$p_k(x) = \frac{1}{1 + e^{-a_k(x)}}$$

where $a_k(x)$ denotes the activation in feature channel k at the pixel position x and $p_k(x)$ is the approximated maximum-function for that feature channel. The binary cross entropy is then used to calculate the loss at each pixel

$$E = -\sum_{i=1}^{2} w \cdot y_i' \cdot \log(y_i)$$

where y_i is the predicted probability of the class i, y_i' is the true probability for that class and w is the weight map. The separation weight map w_{sep} is implemented as described in (Ronneberger, Fischer et al. 2015) and prevents touching objects to be segmented as one object by increasing the weights in border regions. The MitoSegNet also includes a class balancing weight map, which decreases the weight of background pixels by a factor of w_{bal} so the final weight map is defined as

$$w = w_{bal} + w_{sep}$$

Just as in the original U-Net, initial weights are drawn from a Gaussian distribution with a standard deviation of $\sqrt{2/N}$, where N is the number of incoming nodes of one neuron.

Other segmentation methods

All segmentation methods include the removal of uninformative image slices from the stack and the remaining stack is reduced to a single image via maximum intensity projection. Prior to all Fiji (ImageJ) segmentation methods a background subtraction (rolling ball algorithm with a radius of 15 pixels) is applied (Schneider, Rasband et al. 2012). After applying the three different feature enhancement approaches with Fiji, they are all followed by a final filter step in which all particles smaller than 10 px in size are removed from the final mask.

The method denoted as **Laplacian** is based on the de Boer lab workflow on quantitative analysis of mitochondrial morphology in *C. elegans* (de Boer, Smith et al. 2015). It is based on the application of a local contrast enhancement (CLAHE, in Fiji with a block size = 15, histogram = 256, maximum = 3) and subsequent object enhancement using a multi-scale Laplacian operator (FeatureJ Laplacian in Fiji with a compute smoothing setting of 1.0) (De Vos, Van Neste et al. 2010). The images are then binarized according to a Yen autothresholding procedure (Yen, Chang et al. 1995).

The **Hessian** method calculates eigenvalues of a Hessian matrix using the Tubeness plugin in Fiji with a sigma value of 1.0 (Sato, Nakajima et al. 1998). To generate a binary image the IsoData autothresholding is used (Huang and Wang 1995).

The **Gaussian** method calculates the difference of Gaussians by generating two gaussian blurred versions of the original image and subtracting the blurred image with the higher sigma (σ= 4) from the image with lower sigma (σ= 2). The binary mask is generated using default autothresholding.

To also include a machine learning segmentation approach we used the open-source software **Ilastik**. This software learns from labels provided by the user, using a random forest classifier in the learning step, in which each pixel's neighbourhood is characterized by a set of nonlinear features (Kreshuk and Zhang 2019). To train the classifier, 12 images were used for training. Each appendant label was created with ilastik by partially marking objects of interest and background.

Segmentation performance on test set

We evaluated the segmentation performance of the six approaches with different measures, each focusing on particular aspects of the segmentation result with specific limitations. The Dice coefficient (dc, also known as F1 score) (Dice 1945, Sørensen 1948) is a statistic value used for comparing the similarity of two binary datasets.

$$dc = \frac{2 \cdot |ground\,truth \, \cap \, prediction|}{|ground\,truth| + |prediction|}$$

On each image, the dc is evaluated on all pixels. The distributions of dice coefficients from all images were tested for normality using the D'Agostino's K-squared test and statistical difference was determined using the Kruskal-Wallis test followed by Dunn's multiple comparisons test. Upon comparing the dc's with the visual segmentation results, we realised that the dc values do not fully reflect the morphological segmentation accuracy (see Fig. 4B).

To obtain information about morphological segmentation accuracy, we compared single object shapes. The single object shape comparison uses the following shape descriptors for comparison: area, eccentricity, aspect ratio (dividing the major axis length by the minor axis length), perimeter and solidity. The difference between predicted and ground truth shape descriptor is calculated as fold deviation for each object and defined as

$$sd_{dev} = \frac{|sd_p - sd_{gt}|}{sd_p}$$

where sd_p is the predicted and sd_{gt} the ground truth shape descriptor. Object correspondence was assumed if at least one identical pixel coordinate was found in both the ground truth and the predicted object. Only objects that were predicted to correspond with a single object in the ground truth or vice versa were included in this analysis. For single correspondence, a fold deviation (the predicted value multiplied or divided by that fold deviation would yield the ground truth value) for all five morphological descriptors from the ground truth is then calculated for each object (Fig. 4C). The single object fold deviation values were tested for normality using the D'Agostino's K-squared test and statistical differences between methods were determined using the Kruskal-Wallis test followed by Dunn's multiple comparisons test. Because the single object shape comparison neglects false positive predictions, we compared the predicted shape distribution of all objects per image against the ground truth. The shape descriptors are the same as used for the single object shape comparison. For segmented objects, distributions of these five descriptors were obtained and compared to the distributions of the ground truth images by calculating the energy distance (Fig. 4E). The energy distance between two distributions d1 and d2 is defined as

$$D(d1, d2) = (2E\,|X - Y| - E|X - X'| - E|Y - Y'|)^{1/2}$$

where X and X' (resp. Y and Y') are independent random variables with a probability distribution of *d1* (resp. *d2*) (Szekely 2002). Energy distance values were normalised prior to being tested for normality using the D'Agostino's K-squared test, followed by applying a Kruskal-Wallis test and a subsequent Dunn's multiple comparisons test.

All methods to test the performance of each segmentation approach were implemented in Python 3.7.3 using the following libraries: NumPy, OpenCV, scikit-image, scikit-learn, scikit-posthocs, Matplotlib, Seaborn, Pandas and SciPy.

Comparison of mitochondrial morphology between *catp-6* mutant and wild-type

Images segmented with the pretrained segmentation model were recorded with a fluorescent microscope, using a 63x 1.4 NA oil lens (Axioskop 2; Carl Zeiss Inc.) and a charge-coupled device camera (1300; Micromax). The *C. elegans* strain carrying the *bcIs78* transgene and the *catp-6(ok3473)* mutation was compared to the wild-type strain carrying the bcIs78 transgene. Quantitative analysis was performed using the MitoA tool.

Application of the MitoSegNet model segmentation to mitochondria in HeLa cells

Images segmented with the pretrained segmentation model were recorded with an inverted Zeiss LSM 880 system equipped with a DPSS 561-nm laser, using a Plan-Apochromat 63x / 1.4 oil DIC objective and a GaAsP detector. Images were collected with the ZEN 2 software at 1024 x 1024 pixels resolution. To visualize mitochondria, HeLa cells (ATCC) were transfected with mitoRFP using Lipofectamine 2000. After 24 hours of transfection, the HeLa cells were either left untreated or treated with 2.5 μM oligomycin (Calbiochem) and 1 μM antimycin A (Sigma) in fresh growth medium for 3 hours. Segmented and raw images were subjected to quantitative analysis using the MitoA tool.

References

Brenner, S. (1974). "The genetics of Caenorhabditis elegans." Genetics 77(1): 71-94.
D'Agostino, R. B. (1971). "An omnibus test of normality for moderate and large sample size." Biometrika 58: 341-348.

D'Agostino, R. B. and E. S. Pearson (1973). "Tests for departure from normality." Biometrika 60: 613-622.

de Boer, R., R. L. Smith, W. H. De Vos, E. M. Manders, S. Brul and H. van der Spek (2015). "Caenorhabditis elegans as a Model System for Studying Drug Induced Mitochondrial Toxicity." PLoS One 10(5): e0126220.

De Vos, W. H., L. Van Neste, B. Dieriks, G. H. Joss and P. Van Oostveldt (2010). "High content image cytometry in the context of subnuclear organization." Cytometry Part A 77A(1): 64-75.

Dice, L. R. (1945). "Measures of the Amount of Ecologic Association Between Species." Ecology 26(3): 297-302.

Diederik P. Kingma, J. B. (2014). "Adam: A Method for Stochastic Optimization." CoRR abs/1412.6980.

Huang, L. and M. Wang (1995). "Image thresholding by minimizing the measure of fuzziness." Pattern Recognition 28(1): 41-51.

Kanazawa, T., M. D. Zappaterra, A. Hasegawa, A. P. Wright, E. D. Newman-Smith, K. F. Buttle, K. McDonald, C. A. Mannella and A. M. van der Bliek (2008). "The C. elegans Opa1 Homologue EAT-3 Is Essential for Resistance to Free Radicals." PLOS Genetics 4(2): e1000022.

Kreshuk, A. and C. Zhang (2019). "Machine Learning: Advanced Image Segmentation Using ilastik." Methods Mol Biol 2040: 449-463.

Kruskal, W. H. and W. W. Wallis (1952). "Use of Ranks in One-Criterion Variance Analysis." Journal of the American Statistical Association 47(260): 583-621.

Lambie, E. J., P. J. Tieu, N. Lebedeva, D. L. Church and B. Conradt (2013). "CATP-6, a C. elegans ortholog of ATP13A2 PARK9, positively regulates GEM-1, an SLC16A transporter." PLoS One 8(10): e77202.

Levene, H. (1960). "Robust tests for equality of variances. In Ingram Olkin; Harold Hotelling; et al. (eds.). Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling." Stanford University Press: 278-292.

Rolland, S. G., E. Motori, N. Memar, J. Hench, S. Frank, K. F. Winklhofer and B. Conradt (2013). "Impaired complex IV activity in response to loss of LRPPRC function can be compensated by mitochondrial hyperfusion." Proc Natl Acad Sci U S A 110(32): E2967-2976.

Ronneberger, O., P. Fischer and T. Brox (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation, Cham, Springer International Publishing.

Sato, Y., S. Nakajima, N. Shiraga, H. Atsumi, S. Yoshida, T. Koller, G. Gerig and R. Kikinis (1998). "Three-dimensional multi-scale line filter for segmentation and visualization of curvilinear structures in medical images." Med Image Anal 2(2): 143-168.

Schneider, C. A., W. S. Rasband and K. W. Eliceiri (2012). "NIH Image to ImageJ: 25 years of image analysis." Nat Methods 9(7): 671-675.

Sørensen, T. (1948). "A method of establishing groups of equal amplitude in plant sociology based on similarity of species and its application to analyses of the vegetation on Danish commons." Kongelige Danske Videnskabernes Selskab 5(4): 1-34.

Szekely (2002). "E-statistics: The energy of statistical samples." Technical Report 02-16.
Yen, J., F. Chang and S. Chang (1995). "A New Criterion for Automatic Multilevel Thresholding." IEEE Trans. on Image Processing 4(3): 370-378.