

# AUTOMATIC MULTI CLASS ORGANELLE SEGMENTATION FOR CELLULAR FIB-SEM IMAGES

Cyril Meyer <sup>a\*</sup>, Véronique Mallouh <sup>b</sup>, Danièle Spehner <sup>b</sup>,  
Étienne Baudrier <sup>a</sup>, Patrick Schultz <sup>b</sup>, Benoît Naegel <sup>a</sup>

<sup>a</sup> ICube, UMR 7357 CNRS, Université de Strasbourg

<sup>b</sup> IGBMC, UMR 7104 CNRS, U1258 INSERM, Université de Strasbourg

## ABSTRACT

Focused Ion Beam milling combined with Scanning Electron Microscopy (FIB-SEM) technique is an electron microscopy imaging method that offers the possibility of acquiring 3D isotropic images of biological structures at the nanometric scale. Automated image segmentation is required for morphological analysis of huge image stacks and to save time consuming manual intervention. Current methods are either specific to data and organelles or lack accuracy. We propose a robust multi-class semantic segmentation method for FIB-SEM images, based on deep neural networks. We evaluate and compare our proposed method on two FIB-SEM images, for the segmentation of mitochondria, cell membrane and endoplasmic reticulum. We achieve results close to inter-expert variability.

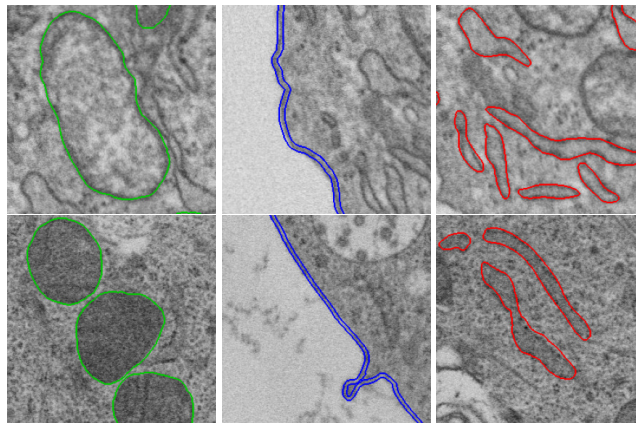
**Index Terms**— Deep Learning, Segmentation, Convolutional Neural Network, Connected Operators, Mathematical Morphology, Electron Microscopy, FIB-SEM

## 1. INTRODUCTION

Focused Ion Beam milling combined with Scanning Electron Microscopy (FIB-SEM) [1, 2] offers the possibility of imaging  $40 \times 40 \mu\text{m}$  areas with isotropic spatial resolution down to 3 nm (typically 5 nm [3]). The Focused Ion Beam (FIB) mills away a thin layer at the surface of the sample, while a Scanning Electron Microscope (SEM) images the newly revealed block face. The iteration of these two processes records a 3D stack of 2D slices separated by as little as 5 nm.

The FIB-SEM technology can provide whole cell images of very large size (of the order of  $2000 \times 1500 \times 500$ , i.e. 1.5 billion pixels), with high spatial resolution and containing many complex biological objects. The annotation of cell compartments and organelles is crucial to extract quantitative information such as size, distribution and morphology. This information could be useful for medical analysis at the cell scale. A human expert can annotate these cellular features

(see Figure 1), but this task is time consuming and not efficient in a context of massive data.



**Fig. 1.** Example of FIB-SEM image patches of size  $256 \times 256$  pixels for the considered classes. Left, mitochondria in green, middle, cell membrane in blue, right, endoplasmic reticulum in red. First line, FS-1 image, second line, FS-2 image.

Automation of this step is required but remains a major challenge for several reasons. First, there are multiple complex objects to segment with different characteristics. Second, the images have an extremely noisy nature (low signal-to-noise ratio, low contrast). Third, resolution, staining and fixation methods vary, resulting in objects with different contrast, texture and shape. Last, the drift during image acquisition may result in shift on the  $z$  axis and isotropy defects, and the contrast may vary from slice to slice.

Few methods have addressed the problem of organelle segmentation in electron microscopy. Multi-class methods designed to segment multiple objects simultaneously exist [4], but lack precision and are mostly focused on mitochondria and membranes. Some methods addressed the segmentation of endoplasmic reticulum in FIB-SEM images, but they are either non-automatic, requiring many user interactions [5], or use class-specific methods [6] which cannot be generalized to other organelles. Currently, there is no generic and accurate multi-class segmentation method for FIB-SEM

\* IdEx Doctoral contract, Université de Strasbourg.

organelles.

In the context of complex biomedical images, state of the art methods are currently based on convolutional neural networks [7, 8, 9]. These methods have also been used with success in electron microscopy [6, 7, 10, 11, 12].

In this paper, we propose a semantic segmentation method based on deep neural networks and dedicated to cellular FIB-SEM images. Our method is evaluated quantitatively by comparison with two expert annotations on a 3-class segmentation task.

## 2. METHOD

We base our method on convolutional networks for semantic segmentation, experimenting two different model architectures and using or not grain filter dynamic preprocessing.

### 2.1. Segmentation models

We choose two different U-shaped model architectures:

- First, a UNet [7] like architecture, which is a reference for biomedical image segmentation, adding batch normalization after each convolution and 50% dropout after the last convolution of the deepest block to avoid over fitting. We use padded convolution to maintain the spatial dimension of the output, and the softmax activation function to provide a categorical output.

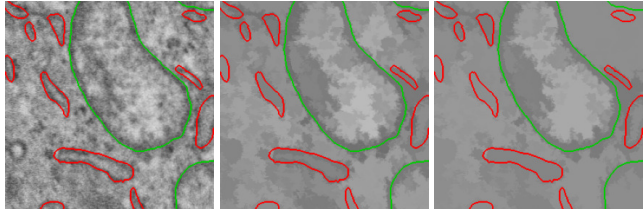
- Second, an EfficientUNet architecture, based on EfficientNet [13] and similar to Eff-UNet [14]. We use an EfficientNet for the contracting path, U-Net like concatenation and single convolution followed with a batch normalization for decoding path. We choose to use an EfficientNet-B4 encoder, resulting in two models with similar number of parameters (31 million for our UNet and 33 million for our EfficientUNet).

We train our models using a Dice loss based on weighted classes:  $\text{Loss} = 1 - \frac{1}{c} \sum_{i=1}^c \alpha_i \times \text{Dice}(X_i, Y_i)$  with  $c$  the number of classes,  $\alpha_i$  the weight for class  $i$ ,  $X_i$  the label for class  $i$  and  $Y_i$  the output for class  $i$ .

### 2.2. Grain filter preprocessing

The grain filter [15] is a non-linear filter based on mathematical morphology and hierarchical representation. This filter can simplify an image without altering its contours (i.e. a contour cannot be created nor shifted), by removing small connected components (see Figure 2). A threshold value  $\lambda$  defines the minimum size of components to keep. Qualitatively, a grain filter can help humans to distinguish objects in noisy images.

We propose to feed the network with preprocessed images using various  $\lambda$  values and concatenate the results in the color (spectral) channels. The selection of  $\lambda$  values is determined empirically (more details in 3.2).



**Fig. 2.**  $256 \times 256$  patch (left) and filtered patches using grain filter with threshold value  $\lambda$  of 650 (middle) and 3200 (right) on the image FS-1, with all classes annotated.

### 2.3. Patch extraction and prediction

We use a patch approach with the smallest patch size keeping good classification results, so as to have the best results in the fastest way. The training data in terms of patches is much larger than the number of training images. Even if we only have 40 training slices, we can extract millions of unique patches. Moreover, smaller inputs give us the possibility to use our method on less powerful hardware or larger models.

For the training set, we use a patch extraction policy which oversamples patches containing annotation: 80% of this set is composed of patches containing at least one class, the remaining 20% is composed of randomly extracted patches. We use random  $90^\circ$  rotation, horizontal and vertical flips for data augmentation. We choose not to use deformations based augmentation due to the nature our data, where shape is very important in characterizing objects.

We predict a slice using a sliding window with a stride of one third of the patch size, and keep the center as the result.

## 3. EXPERIMENTS

### 3.1. Data

We performed our experimentation on two different FIB-SEM images. Each image is a stack of 2D slices. Both images represent HeLa cells, embedded in Lowicryl, and stained using osmium tetroxide and uranyl acetate.

The FS-1 image (FIB-SEM-1) depicts a cell chemically fixed at room temperature and was recorded with an  $x, y, z$  resolution of  $5 \text{ nm} \times 5 \text{ nm} \times 20 \text{ nm}$ , respectively. The manually annotated data is composed of 79 slices. FS-2 has a  $7.5 \text{ nm} \times 7.5 \text{ nm} \times 15 \text{ nm}$  resolution and represents a high pressure frozen and freeze-substituted cell. The annotated data is composed of 80 slices, coming from a subsection containing 120 slices, where we have kept the first and last 40 slices of the subsection.

Each slice has been annotated by an expert using 3 different classes: mitochondrion, cell membrane and endoplasmic reticulum. A default background class is affected to each non-assigned pixel. We divided each image in 3 sets, training (first 40 slices), validation (next 20 slices) and test (last slices). The training set is used to train the network, the validation set to

choose hyper-parameters and select models, and the test set to evaluate our method. One slice from the test set has also been annotated by another expert, to assess inter-expert variation.

### 3.2. Methodology

We performed four experiences in the same way on both images, extracting  $256 \times 256$  patches and using the two different models presented in 2.1 with or without grain filtered inputs.

We trained our model using Adam [16] optimizer with the following parameters:  $\alpha = 0.001$ ,  $\beta_1 = 0.9$ ,  $\beta_2 = 0.999$ ,  $\epsilon = 1e-07$ . We used the following weights for our loss, background 0.1, mitochondria 0.7, cell membrane 1.0 and endoplasmic reticulum 0.9. We trained our models during 100 epochs, composed of 256 mini-batches, with mini-batches of 4 patches. After each epoch, we used 64 mini-batches from the validation set to evaluate our model and save the best model using validation loss.

For each experiment, we used either a single input model, or a 3 input models with 2 filtered versions of the patch as additional input. We chose filter parameters  $\lambda$  equal to 650 and 3200, which is approximately equal to 1% and 5% of the patch surface. These parameters have been determined empirically, keeping the best results after trying several values between 0.1% and 20% of the patch surface.

For each configuration, we performed 15 independent trainings. We then evaluated them by calculating F1 score of each slice for each class and averaging the score for each class. Based on these results, we selected the 10 best models using the sum of the 3 scores on the validation set.

### 3.3. Implementation

All the project is running with Python 3 using TensorFlow [17], Keras and Higras [18]. Our EfficientUNet model is based on the EfficientNet implementation from TensorFlow, to which we add a decoder path and concatenations. Our implementation and all our codes can be found here: <https://github.com/Cyril-Meyer/NeNIST>.

### 3.4. Results

For evaluation, we performed a prediction on the test set. We compared our results with Ilastik 1.3.3 [19] using the pixel classification workflow and using all features. The Ilastik workflow used is based on a random forest classifier. We did not use the 40 training slices but only 20, keeping one in two because of performance issues, as Ilastik is not designed to work with labels as dense as ours. We also compared the annotations between the first and second expert.

We computed the F1 score for each class. Results are shown in Table 1. We compared all the results using a second metric, which counts detected connected components from the labels. A connected component is well detected

if  $recall > 0.80$ , undetected if  $recall < 0.10$ , and under-detected if  $0.10 \leq recall \leq 0.80$ . These results are shown in Table 2. The cell membrane class is not included in this table because the metric is not suitable for few thin linear structures.

Visual results are shown in Figure 3. For Table 2 and Figure 3, we choose a single prediction from the best method in Table 1.

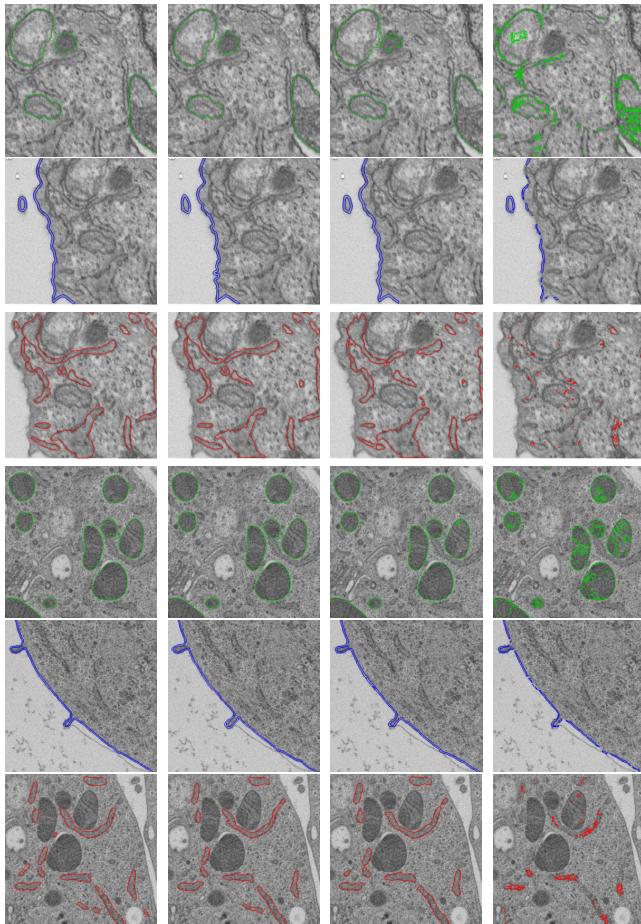
Method	F1 Class 1	F1 Class 2	F1 Class 3
FS-1			
UNet	.937 (.009)	<b>.806</b> (.012)	.724 (.010)
UNet+F	.939 (.013)	<b>.800</b> (.019)	.731 (.012)
EUNet	<b>.951</b> (.008)	<b>.764</b> (.036)	.733 (.015)
EUNet+F	<b>.947</b> (.011)	<b>.757</b> (.040)	<b>.739</b> (.011)
Ilastik	.185	.672	.112
2 <sup>nd</sup> expert	.941	0.751	.740
FS-2			
UNet	<b>.955</b> (.002)	<b>.754</b> (.005)	.727 (.003)
UNet+F	<b>.950</b> (.002)	<b>.742</b> (.009)	.716 (.005)
EUNet	<b>.957</b> (.001)	<b>.752</b> (.009)	<b>.739</b> (.007)
EUNet+F	<b>.955</b> (.002)	<b>.748</b> (.009)	.729 (.008)
Ilastik	.872	.520	.216
2 <sup>nd</sup> expert	.949	.593	.783

**Table 1.** F1 score on test set, mean (standard deviation when necessary for comparison). Bold: best score or higher than second expert. EUNet: EfficientUNet. +F: grain filtered inputs

Method	TP	FN	UD
FS-1 Class 1			
EUNet+F (/193)	92.7% (179)	4.7% (9)	2.6% (5)
Ilastik (/193)	0.0% (0)	49.2% (95)	50.8% (98)
2 <sup>nd</sup> expert (/10)	100.0% (10)	0.0% (0)	0.0% (0)
FS-1 Class 3			
EUNet+F (/1034)	46.9% (485)	19.1% (198)	33.9% (351)
Ilastik (/1034)	0.0% (0)	82.2% (850)	17.8% (184)
2 <sup>nd</sup> expert (/55)	23.6% (13)	9.1% (5)	67.3% (37)
FS-2 Class 1			
EUNet (/363)	89.0% (323)	6.1% (22)	5.0% (18)
Ilastik (/363)	62.8% (228)	4.7% (17)	32.5% (118)
2 <sup>nd</sup> expert (/21)	100.0% (21)	0.0% (0)	0.0% (0)
FS-2 Class 3			
EUNet (/1468)	56.1% (824)	21.5% (316)	22.3% (328)
Ilastik (/1468)	0.0% (0)	60.6% (889)	39.4% (579)
2 <sup>nd</sup> expert (/68)	26.5% (18)	19.1% (13)	54.4% (37)

**Table 2.** Results on test set, in percentage of component. TP: true positive, FN: false negative, UD: under detected.

All the variations of our method outperform the results obtained using Ilastik method. The EfficientUNet models give



**Fig. 3.** Comparison on  $500 \times 500$  subsections of image. From left to right: ground truth (1<sup>st</sup> expert), 2<sup>nd</sup> expert, our method and Ilastik. The 3 first rows are derived from FS-1 and 3 last rows from FS-2. Notice that in some cases (third row), the model and the 2<sup>nd</sup> expert agree on the disagreement with the first expert.

better results than the UNet ones, especially for the endoplasmic reticulum class. The filtered input improves slightly the results on FS-1 endoplasmic reticulum class, but not on the FS-2 image.

These results show that our method is robust and efficient. The UNet models are closer to the expert for the cell membrane class but this class is annotated as a thin linear structure, and a small variation in the way of annotating results in a large F1-score drop. Comparing with the 2<sup>nd</sup> expert, and viewing the standard deviation for this case, these results are closer to an over-fitting to the 1<sup>st</sup> expert than actual better results. We can notice that, although bringing a visual improvement for the human expert, the grain filter does not provide significant improvement in the classification results. This is an interesting result and can be explained by the capacity of the deep neural network to extract relevant features, even in a noisy environment. Ilastik performs well on classes with more con-

trast (mitochondria in FS-2 and cell membrane) but not on the others.

The mitochondria class is well segmented, and less than 5% of the objects are missed. For the endoplasmic reticulum class, more objects are missed. Using both metrics, we can conclude from the high F1-score that the missed connected components are mainly small entities which represent a small proportion of pixels in the images. This issue probably comes from the loss function used, which penalized less small missing components than large components. The variability between the results of the images shows that our method is robust and works with more or less contrasted images.

#### 4. CONCLUSION

In this paper, we focused on the multi-class organelle semantic segmentation problem for biomedical FIB-SEM images. We proposed a deep learning based method, experimented a grain filter preprocessing and two different model architectures. Our method achieves results close to inter-expert variability, and we provided a comparison with another method. Our method has the advantage of being generic to multi-class organelles segmentation, allowing the addition of new classes easily.

False negative detection could be improved by using morphological aware loss or 2.5D approaches.

Adding more organelles to the segmentation task, such as the golgi apparatus, endosome, and the nucleus with its heterochromatin and euchromatin or nucleolar compartments or the nuclear envelope will be an interesting challenge. Our method still requires a large amount of annotated data, but we expect that interactive segmentation methods based on our current results may result in faster data annotation process with less human intervention. Experimenting with the use of our current models on different images, and exploring the possibility of transfer learning between models is likely to provide a general model which can segment efficiently several organelles on multiple images.

#### 5. ACKNOWLEDGMENTS

We thank F. Alpy from the "Institut de Génétique et de Biologie Moléculaire et Cellulaire" for providing the FS-2 cellular system. We acknowledge the use of resources of the French Infrastructure for Integrated Structural Biology FRISBI ANR-10-INBS-05 and of Instruct-ERIC. We acknowledge that this work is supported by an IdEx doctoral contract, Université de Strasbourg. Conflict of Interest: The authors declare that they have no conflict of interest.

#### 6. COMPLIANCE WITH ETHICAL STANDARDS

The authors declare that the study does not involve human- or animal-derived samples that require ethical approval. This article does not contain any studies involving human participants performed by any of the authors.

## 7. REFERENCES

- [1] J. A. W. Heymann, M. Hayles, I. Gestmann, L. A. Gianuzzi, B. Lich, and S. Subramaniam, "Site-specific 3D imaging of cells and tissues with a dual beam microscope," *J. Struct. Biol.*, vol. 155, no. 1, pp. 63–73, Jul. 2006.
- [2] G. Knott, H. Marchman, D. Wall, and B. Lich, "Serial Section Scanning Electron Microscopy of Adult Brain Tissue Using Focused Ion Beam Milling," *J. Neurosci.*, vol. 28, no. 12, pp. 2959–2964, Mar. 2008.
- [3] K. L. Briggman and D. D. Bock, "Volume electron microscopy for neuronal circuit reconstruction," *Curr. Opin. Neurobiol.*, vol. 22, no. 1, pp. 154–161, Feb. 2012.
- [4] K. Cetina, J. M. Buenaposada, and L. Baumela, "Multi-class segmentation of neuronal structures in electron microscopy images," *BMC Bioinformatics*, vol. 19, no. 1, p. 298, Aug. 2018.
- [5] S. Machado, V. Mercier, and N. Chiaruttini, "LimeSeg: A coarse-grained lipid membrane simulation for 3D image segmentation," *BMC Bioinformatics*, vol. 20, no. 1, p. 2, Jan. 2019.
- [6] J. Liu, L. Li, Y. Yang, B. Hong, X. Chen, Q. Xie, and H. Han, "Automatic Reconstruction of Mitochondria and Endoplasmic Reticulum in Electron Microscopy Volumes by Deep Learning," *Front. Neurosci.*, vol. 14, 2020.
- [7] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *Med. Image Comput. Comput. Assist. Interv. – MICCAI 2015*, ser. Lecture Notes in Computer Science, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, Eds. Cham: Springer International Publishing, 2015, pp. 234–241.
- [8] E. Gibson, W. Li, C. Sudre, L. Fidon, D. I. Shaker, G. Wang, Z. Eaton-Rosen, R. Gray, T. Doel, Y. Hu, T. Whyntie, P. Nachev, M. Modat, D. C. Barratt, S. Ourselin, M. J. Cardoso, and T. Vercauteren, "NiftyNet: A deep-learning platform for medical imaging," *Comput. Methods Programs Biomed.*, vol. 158, pp. 113–122, May 2018.
- [9] E. Moen, D. Bannon, T. Kudo, W. Graf, M. Covert, and D. Van Valen, "Deep learning for cellular image analysis," *Nat Methods*, vol. 16, no. 12, pp. 1233–1246, Dec. 2019.
- [10] T. Zeng, B. Wu, and S. Ji, "DeepEM3D: Approaching human-level performance on 3D anisotropic EM image segmentation," *Bioinformatics*, vol. 33, no. 16, pp. 2555–2562, Aug. 2017.
- [11] M. G. Haberl, C. Churas, L. Tindall, D. Boassa, S. Phan, E. A. Bushong, M. Madany, R. Akay, T. J. Deerinck, S. T. Peltier, and M. H. Ellisman, "CDeep3M—Plug-and-Play cloud-based deep learning for image segmentation," *Nat Methods*, vol. 15, no. 9, pp. 677–680, Sep. 2018.
- [12] C. Xiao, X. Chen, W. Li, L. Li, L. Wang, Q. Xie, and H. Han, "Automatic Mitochondria Segmentation for EM Data Using a 3D Supervised Convolutional Network," *Front. Neuroanat.*, vol. 12, 2018.
- [13] M. Tan and Q. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," ser. Proc. Mach. Learn. Res., K. Chaudhuri and R. Salakhutdinov, Eds., vol. 97. Long Beach, California, USA: PMLR, Jun. 2019, pp. 6105–6114.
- [14] B. Baheti, S. Innani, S. Gajre, and S. Talbar, "Eff-UNet: A Novel Architecture for Semantic Segmentation in Unstructured Environment," in *2020 IEEE/CVF Conf. on Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2020, pp. 1473–1481.
- [15] V. Caselles and P. Monasse, "Grain Filters," *J. Math. Imaging Vis.*, vol. 17, no. 3, pp. 249–270, Nov. 2002.
- [16] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *3rd Int. Conf. Learn. Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, Y. Bengio and Y. LeCun, Eds., 2015.
- [17] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, M. Kudlur, J. Levenberg, R. Monga, S. Moore, D. G. Murray, B. Steiner, P. Tucker, V. Vasudevan, P. Warden, M. Wicke, Y. Yu, and X. Zheng, "TensorFlow: A system for large-scale machine learning," in *Proceedings of the 12th USENIX Conf. Oper. Syst. Des. Implement.*, ser. OSDI'16. USA: USENIX Association, Nov. 2016, pp. 265–283.
- [18] B. Perret, G. Chierchia, J. Cousty, S. J. F. Guimarães, Y. Kenmochi, and L. Najman, "Higra: Hierarchical Graph Analysis," *SoftwareX*, vol. 10, p. 100335, Jul. 2019.
- [19] S. Berg, D. Kutra, T. Kroeger, C. N. Straehle, B. X. Kausler, C. Haubold, M. Schiegg, J. Ales, T. Beier, M. Rudy, K. Eren, J. I. Cervantes, B. Xu, F. Beuttenmueller, A. Wolny, C. Zhang, U. Koethe, F. A. Hamprecht, and A. Kreshuk, "Ilastik: Interactive machine learning for (bio)image analysis," *Nat Methods*, vol. 16, no. 12, pp. 1226–1232, Dec. 2019.