OPEN ACCESS

3D surface reconstruction of cellular cryo-soft X-ray microscopy tomograms using semisupervised deep learning

Michael C. A. Dyhr[a,1] (ID), Mohsen Sadeghi[b,1,2] (ID), Ralitsa Moynova[a] (ID), Carolin Knappe[a] (ID), Burcu Kepsutlu Çakmak[a], Stephan Werner[c] (ID), Gerd Schneider[c,d], James McNally[c], Frank Noé[b,2], and Helge Ewers[a,2] (ID)

Cryo-soft X-ray tomography (cryo-SXT) is a powerful method to investigate the ultrastructure of cells, offering resolution in the tens of nanometer range and strong contrast for membranous structures without requiring labeling or chemical fixation. The short acquisition time and the relatively large field of view leads to fast acquisition of large amounts of tomographic image data. Segmentation of these data into accessible features is a necessary step in gaining biologically relevant information from cryo-soft X-ray tomograms. However, manual image segmentation still requires several orders of magnitude more time than data acquisition. To address this challenge, we have here developed an end-to-end automated 3D segmentation pipeline based on semisupervised deep learning. Our approach is suitable for high-throughput analysis of large amounts of tomographic data, while being robust when faced with limited manual annotations and variations in the tomographic conditions. We validate our approach by extracting three-dimensional information on cellular ultrastructure and by quantifying nanoscopic morphological parameters of filopodia in mammalian cells.

cryo-soft X-ray microscopy | deep learning | automated segmentation

All cells are filled with a dense and complex mixture of organelles and macromolecular structures that rely on nanoscale interactions to perform vital functions. To understand the organization and interaction of cellular organelles, gathering information at the resolution of native cellular ultrastructure is essential. Electron microscopy (EM) is considered the gold standard for ultrastructural analysis. However, while a significant portion of our current understanding of cellular processes is based on EM methods, several limitations in sample preparation, volume coverage, acquisition speed, and throughput remain. Cryo-soft X-ray tomography (cryo-SXT) offers several advantages over EM approaches, but with the trade-off of somewhat lower resolution.

Cryo-SXT takes advantage of the intrinsic different absorption contrast of elements. In the spectral region called the 'water window', carbon atoms absorb very strongly compared to oxygen atoms. This is of great relevance for biological specimens, where the natural contrast between carbon-dense material such as proteins and lipids and the aqueous environment can be exploited to visualize membranous organelles and small particles [e.g., nanoparticles (1) or viruses (2)] at a resolution of 30 to 40 nm (3). Moreover, specific organelles or structures can be distinguished based on their linear absorption coefficient, which is directly related to their chemical composition (4). Unlike many EM methods, cryo-SXT samples require neither chemical fixation nor embedding or labeling. Instead, cells can be grown on gold transmission electron microscopy (TEM) grids, plunge frozen and directly imaged in the transmission X-ray microscope. Furthermore, the penetration depth of soft X-rays is with 15 μm significantly deeper than what would be accessible by TEM techniques. This allows transmission microscopy and tomography of the entirety of the cellular volume in minutes without the need for physical sectioning (5, 6).

Cryo-SXT yields projection images acquired at incremental tilt angles, which are then aligned and processed by a tomographic reconstruction algorithm to produce a tomogram. With tissue culture cells on a flat support (e.g., EM grids), a tilt range of ±65° is typically used for tomography. In this paper, we use the term "tomogram" instead of "tomographic reconstruction" in order to avoid confusion with the reconstruction procedure performed by our proposed deep network. One such obtained tomogram is composed of hundreds of image slices, from which the 3D ultrastructure of the cell must subsequently be extracted using segmentation techniques. Segmentation of the tomogram yields more quantifiable information, as for instance the size, organization, and distribution of organelles in the cytosol (3). Cryo-SXT covers a large volume in a short period of time and does not suffer from preparation artifacts such as section compression or shrinkage (7, 8). Fast-paced development of 3D-EM techniques such as focused ion beam scanning electron microscopy (FIB SEM) toward larger volumes and imaging under cryogenic and noncryogenic

Significance

Cryo-soft X-ray tomography (cryo-SXT) is a powerful microscopy technique, which resolves the three-dimensional ultrastructure of cells in their near-native state at high resolution. The depth of ultrastructural detail presented in these datasets is highly valuable for cell biological research, but also poses the problem of extracting all that information by segmentation for more quantifiable analysis. The increasing implementation of artificial intelligence in bioimage analysis has produced a plethora of tools for semiautomated or fully automated segmentation for electron and fluorescence microscopy; however, for cryo-SXT, segmentation still relies on extensive manual input. In this work, we developed a convolutional neural network, which is capable of automatically segmenting cryo-SXT datasets at high accuracy within a few minutes.

[1]M.C.A.D. and M.S. contributed equally to this work.

[2]To whom correspondence may be addressed. Email: mohsen.sadeghi@fu-berlin.de, frank.noe@fu-berlin.de, or helge.ewers@fu-berlin.de.

conditions has led to remarkable results in the ultrastructure of biological specimens (9–12). However, these techniques are still limited by extensive sample preparation procedures in comparison to cryo-SXT (13). Using cryo-SXT, it is possible to acquire a tomogram covering ~1000 μm^3 within less than 30 min. This allows for the study of changes in, for example, the organization of cellular organelles with statistically relevant amounts of data in 3D volumes as in the yeast cell cycle (14), nanoparticle uptake (1), and parasite infection (15). Moreover, the compatibility of cryo-SXT with cryo-fluorescence microscopy or cryo-hard X-ray fluorescence microscopy facilitates the identification and visualization of specific cellular structures and processes (16, 17).

While volumetric data can be generated rather quickly in cryo-SXT, the data analysis is still time-consuming. Specifically, in the cryo-SXT workflow, manual segmentation remains the key bottleneck (18). The complete manual segmentation of a dataset may take several days or even weeks, discouraging the effort to segment the data beyond the structures of interest. As a result, subtle but potentially important changes in cellular ultrastructure may go unnoticed, even if that information was present in the original data. Automated or semiautomated segmentation using machine learning algorithms, especially artificial neural networks, hold the promise of overcoming this barrier. Deep learning (DL) has already made a great impact in microscopy, with promising applications in resolution enhancement (19–22) and denoising (23), missing data reconstruction (24) and rapid phenotyping (25). More specifically, significant contributions have already been made toward automated segmentation in various microscopy applications, including volume-EM (9, 10, 26–29). Open-source solutions offering basic functionality are also available (30–33), and the design of dedicated fine-tuned networks will likely lead to further improvements. The design of methods suited for automated segmentation of cryo-SXT tomograms is an ongoing effort (34–38) and faces inherent challenges such as limited depth of focus and the missing wedge problem (6). The missing wedge refers to larger tilt angles that are missing in the 3D tomogram, for example in cryo-SXT, +/–65° instead of +/– 90°. Reconstruction of data with a missing wedge introduces artifacts, such as blurring and apparent elongation in the z axis (39, 40). Addressing these challenges using conventional DL image processing requires a large pool of consistent, high-quality training data, which in turn requires extensive access to suitable cryo-soft X-ray microscopes and significant manual effort in building suitable training datasets.

Focusing on the automatic image segmentation task, supervised learning with deep convolutional networks (ConvNets) remains the most popular approach (41–43). While other approaches, from using the general-purpose vision transformers (44, 45) to more specific methods such as the weighted random walks (46) have also been suggested, ConvNets that are being actively fine-tuned and optimized to compete for the top performance in image processing tasks (47) still provide a reliable and versatile solution. Nevertheless, relying exclusively on supervised learning with ConvNets requires large amounts of manually annotated data. Moreover, in supervised training the network can overfit on the instrument(s) used in acquiring the training data and fail to generalize to a different imaging source.

To address these challenges, while benefitting from the performance and ease of use of ConvNets, we introduce a method for semisupervised learning of cryo-SXT image segmentation. We propose a deep convolutional model that combines the well-established U-Net architecture (26, 27) with an image reconstruction path, and uses a mixture of manually annotated and unannotated images for training. While the network learns to classify features of interest, such as membrane structures, from manual annotations,

significantly larger amounts of unannotated data enable it to learn a representation of various possible image features and imaging conditions. This approach significantly enhances the network's understanding of the variability of input, while reducing the number of annotated images required to make reliable predictions. Once trained on both the annotated and unannotated data, zero-shot segmentation and surface reconstruction of a variety of cryo-SXT tomograms from different instruments can be achieved without the need for parameter tweaking and retraining. While the surface reconstruction enables rapid visual inspection of the cellular structures captured in the dataset, the automatic segmentation allows faster quantitative evaluation of the data. The entire application pipeline requires minutes for whole-cell tomograms, offering a versatile tool to fully harness the structural information in cryo-soft X-ray tomographic datasets. We test our model on unseen data acquired at three different synchrotrons to demonstrate transferability and quantify a significant amount of subresolution 3D data to demonstrate the throughput of our technique.

Results

Semisupervised Training of Deep Convolutional Network for Cryo-SXT Segmentation.

We aimed to develop an end-to-end deep-learning network suited for the analysis of cryo-SXT datasets. After testing a large number of different network architectures, we determined that the significant variety of image features and the relatively low contrast in cryo-SXT images demand a very deep network with a large representation power (*Methods*).

Consequently, we constructed a specific-purpose convolutional feed forward network (48), as illustrated in Fig. 1 and explained in detail in *Methods*. This network takes 2D slices of cryo-SXT tomograms, and transforms them in parallel to two outputs by using "image" and "annotation" decoders (48). From a general perspective, our model augments a U-Net, that is used for the main segmentation tasks, with an unsupervised autoencoder, with them sharing the same encoder (Fig. 1). The U-Net architecture and its variants have proven very successful in supervised segmentation tasks (26, 27, 38, 41, 49). We will demonstrate how the addition of the parallel autoencoder improves the segmentation performance, and brings it up to the task of segmenting cryo-SXT tomograms based on a rather small training set.

As is usual with ConvNets, besides the width and height dimensions, a 2D image processed by the network is assigned an extra "channel" dimension. The input grayscale images have thus only one channel. The channel dimension contains parallel sets of pixel-wise information about an image, that are being processed simultaneously at each layer of the network. Since using residual learning (i.e., making skip connections spanning one or more intermediate layers in a feed-forward architecture) facilitates training of significantly deeper networks (50), our proposed model relies heavily on residual blocks. These blocks are made up of convolutional layers in which residuals (i.e., information carried across the skip connections) are concatenated along the channel dimension, accumulating extended pixel-wise information. Similar designs have successfully been used in previous segmentation applications (51, 52).

Overall, our proposed network contains 31 layers in the segmentation and 23 layers in the image reconstruction path (Fig. 1*A*) with a total of more than 76 million trainable parameters (network weights). To improve the training procedure of the network, we have employed batch normalization in the majority of blocks, which leads to faster convergence by regularizing the mean and variance of pixel intensity values across batches of image data (53) (Fig. 1). The global bottleneck of the network introduces a
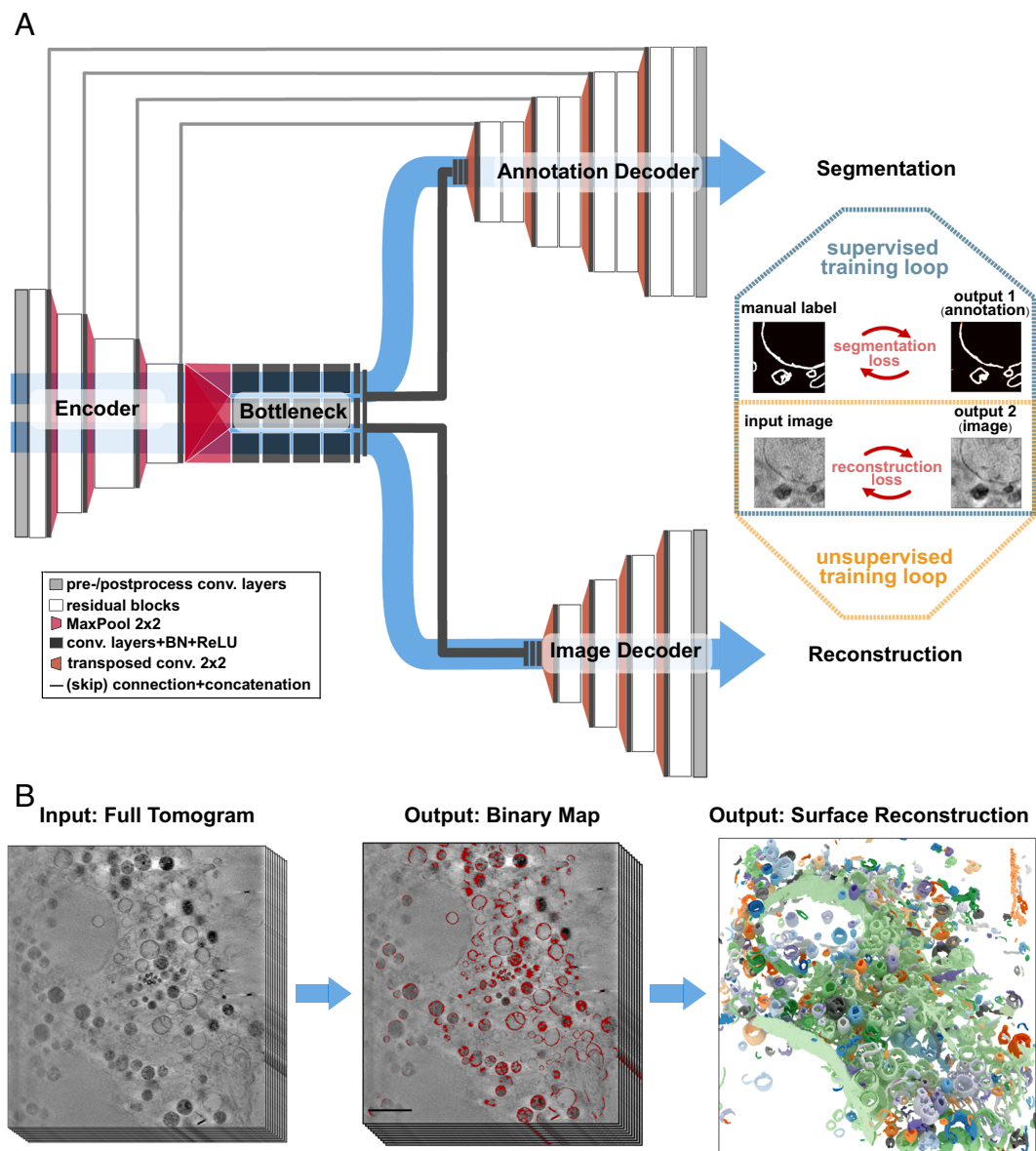
Fig. 1. Processing workflow from raw tilt data to 3D volume rendering of cryo-SXT datasets. (*A*) Schematic of the two-headed convolutional neural network, illustrating the image segmentation and image reconstruction paths. On the right, the training procedure is shown. (*B*) Whole cryo-SXT tomograms can be used as input for the CNN. Shown is a 0° tilt image of a representative cryo-soft X-ray tomographic dataset, showing a part of the cytosol of a plunge-frozen CRFK cell. The output is a binary map of annotated labels, shown here in red and overlaid with the raw data. The Network can also produce a 3D volume rendering of all network output labels from the entire tomogram in random colors. (Scale bar, 2 μm.)

cardinality of 3, i.e., three parallel convolutional paths with different kernel sizes (Fig. 1*A*). This allows for higher representation power with less network depth and prevents vanishing gradients during the training procedure (54). The outputs of the bottleneck paths are concatenated and passed to the two decoders. The image decoder (Fig. 1*A*) is tasked with reproducing the input image, while the annotation decoder performs the segmentation. Global encoder-to-decoder skip connections, which are characteristic of the U-Net architecture (26, 27), are only introduced between the encoder and the annotation decoder (Fig. 1*A*).

For the image segmentation branch, we have made hyperparameter choices similar to the original U-Net design in the number of down/up-sampling stages, kernel sizes, and the number of convolutional filters. For other hyperparameters, such as the number of residual blocks in each stage, the cardinality of the bottleneck region, final layers in the decoders, and the learning rate, we have relied on extensive experimentation and manual selection of the

combinations that resulted in best results, both qualitatively as judged by the final segmentation, and quantitatively as measured via pixel-wise accuracy and precision (details in *Methods*). Despite the large number of trainable parameters, with several key decisions in our network design (details in *Methods*) it is memory-efficient enough to feed full-sized slices from our dataset in one go.

The network is trained in a semisupervised fashion by following a two-step protocol (Fig. 1*A* and *SI Appendix*, Fig. S1*A*): i) in the supervised training step, we use manually segmented image-label pairs to train the network. These are tomogram slices for which segmentation labels in the form of binary masks are prepared. In the supervised training step, the loss function is a weighted sum between image reconstruction and label prediction errors. Subsequently, weights in both branches of the network are updated by the optimizer. The aim of this step is to best reproduce the image and its accompanying label. ii) in the unsupervised training step, we use a large number of slices from different tomograms to train only the image reconstruction path of the network. For these

images no manual segmentations are available. Thus, when the images are used in the training loop, the segmentation decoder has to be frozen and its weights are not updated (*SI Appendix*, Fig. S1*A*). This step amounts to training an autoencoder for reproducing cryo-SXT images and relies on a loss function between the output and input images only (*Methods*). What this step achieves is to force the encoder toward learning a compressed representation of cryo-SXT images that is transferable between different cell and tomography conditions. The overall training protocol entails a loop that alternates between the two steps, and iteratively updates the network weights. Satisfactory convergence of the training protocol is achieved when the network is capable of simultaneously restoring the input image as well as producing the segmentation labels where provided (Fig. 1*B*). We used early stopping (decided based on validation accuracy) to prevent the network from overfitting, as well as reducing the chance for false positives.

To quantify the performance of the trained network, we have used pixel-wise scores that measure the percentage of pixels correctly reproduced, both in the reconstructed image and in the predicted label. The predicted segmentation labels have been first transformed into binary data by applying a hard threshold of 0.9 (only pixels are accepted as positively segmented in the predicted label that have intensities greater than 0.9). This results in an unequivocal way of obtaining true and false predictions of positive (belonging to a segmented area) and negative (belonging to the background) pixels. We have thus measured accuracy, precision, recall, and F1-score of the data using this information (Fig. 2*A*, details in *Methods*). To estimate the reconstruction accuracy of images, pixels with values within one SD of intensity for reconstructed images are considered positive predictions

We have run the training protocol for 700 epochs on 256 × 256 pixel images, which have been prepared on-the-fly by one sweep of our data augmentation protocol over all the full-sized slices in the training set (*Methods*). The receptive field of the combined convolutional and pooling layers in the encoder and bottleneck of our network is 206 pixels (55). This justifies the chosen size of the images used for training the network, and should also serve as an estimate for the scale of features learnable by the network.

After training, the image reconstruction path achieves a pixel-wise accuracy of > 80%. It is noteworthy that reconstructing images also includes reproducing the background noise. Because the network lacks any stochastic input for noise sampling, the background noise is implicitly approximated (Fig. 1 and *SI Appendix*, Fig. S2). The segmentation path, on the other hand, reproduces manual labels with pixel-wise mean accuracy of 97% and precision of 74% over the test dataset (Fig. 2*A* and *SI Appendix*, Fig. S2). We emphasize that all the reported scores have been measured on hold-out test datasets that the network has not been exposed to during the training.

To demonstrate that our modification of the U-Net architecture and the addition of the image-to-image autoencoder indeed improves the segmentation task, we have separately trained two networks, one a pure U-Net trained in a supervised manner on image-label training data, and the other, our model, with the same U-Net segmentation branch (Fig. 1*A*), and trained with the proposed semisupervised protocol. Our model with the semisupervised training shows a significantly higher segmentation precision across the test dataset (Fig. 2*A*). The median segmentation precision has been improved by at least 14% by the semisupervised approach (Fig. 2*A*). Additionally, as will be shown, semisupervised learning not only improves the segmentation quality on one dataset, but also contributes to the transferability of the model to different imaging instruments.

Evaluation of Segmentation Speed, Quality, and Robustness.
Using a GeForce RTX 3090 graphics card with 24 GB of graphics memory, and using the already trained network, automated segmentation took less than 5 min for a tomogram with 350 slices of 1324 × 1284 pixel images or 8 min for a tomogram of 500 similarly sized slices. The fully convolutional nature of the network makes it possible to apply it, without adjustment, to input images of almost arbitrary size. The only limiting factor is the available graphics memory for simultaneous storage of the bitmap images, in all stages of processing by the network, including the auxiliary channels added in between. With our hardware, we found out that slices with the maximum size of 1664 × 1664 pixels can undergo automatic segmentation in one shot. This resulted in a processing time of 2 s per slice. To allow for processing of larger images on a similar hardware, or similar images on a graphics card with less memory, we added a preprocessing step, run on the CPU, during which the image is sliced into smaller overlapping square chunks. The set of all chunks are processed by the GPU, and in the postprocessing step, reassembled into a fully segmented slice, averaged with constant weights in the overlapping regions. The option also remedies the fact that the current network model with 4 stages of pooling requires input sizes to be multiples of 16.

To evaluate the segmentation quality of the network on the tomogram-level, a human expert user classified the automatically generated labels and filled in missing features into a representative set of 59 slices from three different tomograms. By comparing the pixels in each category, we found that out of all labeled pixels per slice, an average of 69.8% (s.d. = 10.4) were *true positives*, i.e. correctly labeled by the network and 27.7% (s.d. = 10.8) were *false negatives*, i.e., structures additionally labeled by the human user. 1.3 % (s.d. = 2.2) and 1.2 % (s.d. = 2.1) were *false positives* or labeled *reconstruction artifacts*, respectively (Fig. 2 *B–E*). Additionally, we performed a side-by side comparison of 100 slices of a high-quality tomogram of a CRFK cell that had not be used for training after analysis by an expert user (we allowed 20 h for analysis) and the trained network and found good agreement (*SI Appendix*, Fig. S4).

Since SXT requires synchrotron radiation, it can at this time only be performed at a few institutions worldwide. Thus, to be widely applicable, our method should generalize to data acquired from other synchrotrons. To test this, we applied our trained deep-learning network to X-ray tomography datasets obtained from two other synchrotrons. We used a dataset of A549 lung cancer cells incubated with nanoparticles for endocytosis that was imaged on the MISTRAL beamline at the ALBA synchrotron (Barcelona, Spain; *SI Appendix*, Fig. S5) (1), and a dataset of mock-infected U2OS cells (EMPIAR-10416) imaged on the B24 beamline at Diamond Light Source (Didcot, UK) (16) (*SI Appendix*, Fig. S6). The data provided by the Diamond Light Source were acquired with a different X-ray microscope objective with somewhat lower resolution but higher depth of field compared to the objectives used for all the other data (40-nm vs. 25-nm zone plates). Despite the different experimental conditions, these two additional datasets were also readily segmented in minutes by our trained network, which identified the vast majority of membranous organelles inside the cytosol (*SI Appendix*, Figs. S6 and S7). These results demonstrate the utility of our deep network for data generated at cryo-soft X-ray microscopes from other synchrotron facilities and under different imaging and cell-biological conditions.

Using DL to Quantify Filopodia Morphology. Next, we asked whether our automated segmentation would allow for medium-throughput, high-resolution image analysis. As a ubiquitous example,
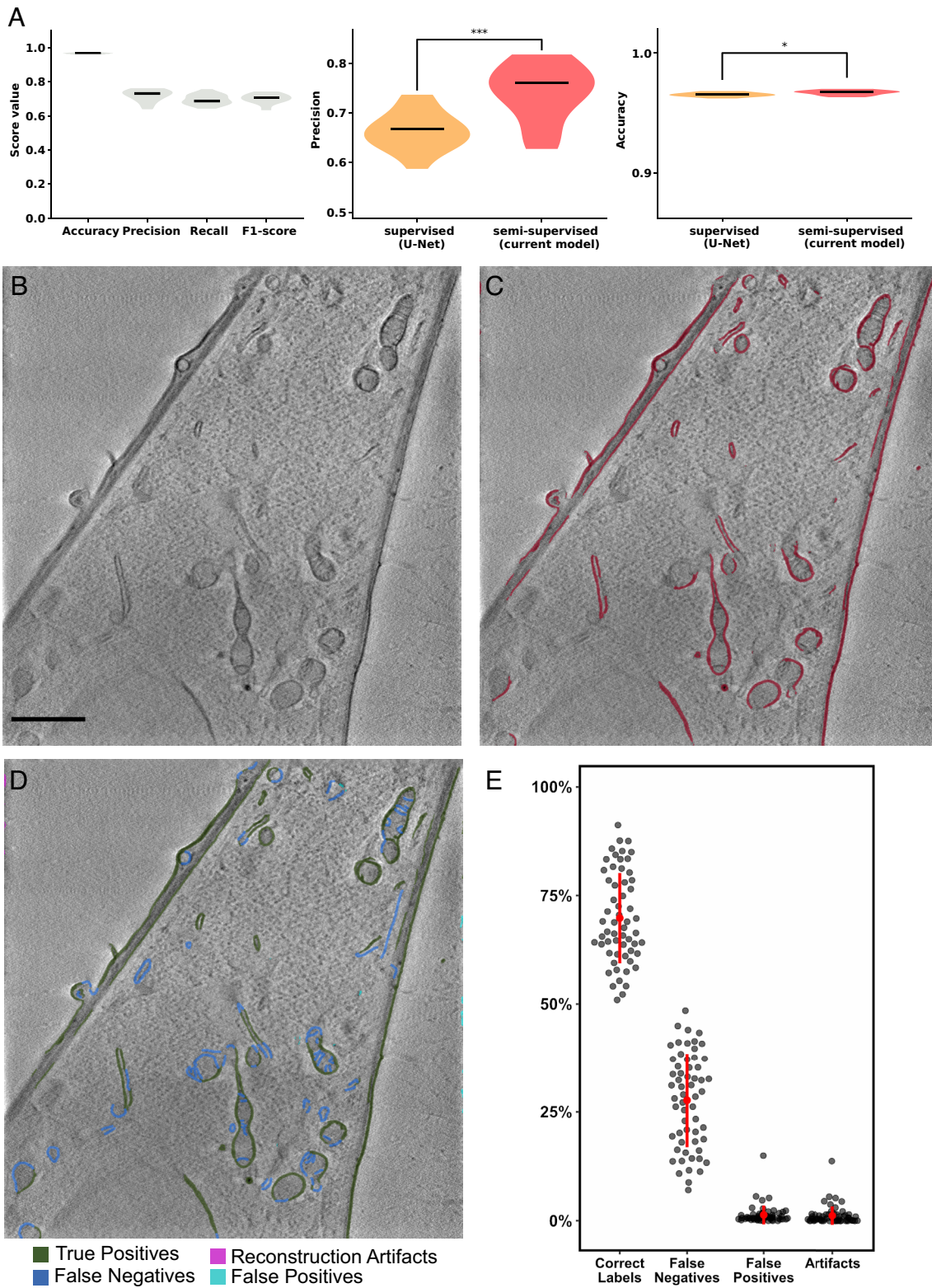
Fig. 2. Quantification of image segmentation robustness and accuracy. (*A*, *Left*) pixel-wise metrics (accuracy, precision, recall, and F1-score) of the segmentation task, measured for the trained deep network when applied to the test dataset. (Middle and right) comparison between the performance of two distinct networks, trained, respectively, in the supervised and semisupervised manner, in terms of precision and accuracy of predictions on the test dataset. (*B*) A representative 2D slice from the 3D tomographic image stack that was analyzed. (*C*) Overlay of the example slice in *A* with the labels automatically produced by the deep network. (*D*) The labels produced by the network were manually filled in on a series of randomly selected single slices and classified into true positives (green), reconstruction artifacts (pink), false positives (cyan). Manually added labels (false negatives) are shown in blue. (*E*) The network performance was evaluated based on the percentage of surface pixels per class for each analyzed slice. The red dots indicate the mean percentage and red bars the SD (Scale bar, 2 μm.)

we investigated cellular filopodia, which are abundant, narrow plasma membrane protrusions whose spatial dimensions are inaccessible by visible-light imaging. These elongated, submicron-thick, actin-filled structures play an important role in cellular migration and spreading of cells (56). We obtained three tomograms of CRFK cells, containing a total of 59 filipodia. These three datasets were then

rapidly segmented, reconstructed and analyzed (Fig. 3*A*). Manual corrections were performed where necessary to close any open loops and separate the filopodia objects from the plasma membrane, from which they originate (*Methods*).

As expected, we found that the spatial orientation of the filopodia affected their analysis in 2D tomogram slices. Filopodia crossing a lateral tomographic slice at a low angle would be observed as a very elongated ellipse (i.e., having a long first axis length, Fig. 3*A*). Consistent with this observation, we noticed a high variation of the first axis length of the filopodia cross-section (302 ± 189 nm). The second axis length, more robustly reflecting filopodial diameter, was with 132 ± 28 nm generally more homogeneous. The diameter of filopodia is thus essentially constant but the 3D orientation prevents accurate measurements of volume, surface area, and perimeter from 2D cross-sections. It is thus essential to analyze entire filopodia from 3D reconstructions.

Our DL–based reconstruction provided the basis for swift 3D analysis as well. Using a pixel size of 9.8 nm in all dimensions, we found a major axis length of 966 ± 394 nm by fitting an ellipsoid to the segmented data. The 3D second axis length also had a higher error with 320 ± 141 nm (Fig. 3*B*). We found that the average volume and surface area of filopodia was highly variable, with $2.2 \cdot 10^{7} \pm 1.7 \cdot 10^{7}$ nm^3 and $6.1 \cdot 10^{5} \pm 3.6 \cdot 10^{5}$ nm^2, respectively. However, the surface area to volume ratio was consistent with 30.2 ± 5.3 µm^{-1} (or 0.3021 ± 0.0534 nm^{-1}, respectively) for the entire population, indicative of a consistent shape of all filopodia. Taken together, we concluded that the combination of SXT with DL image analysis allowed for fast reconstruction and analysis of small subresolution 3D structures.

Discussion

In this work, we have developed a deep neural network that allows for rapid and robust segmentation and surface reconstruction of three-dimensional cryo-SXT data. On average, it recognizes 70% of the feature pixels in a cryo-soft X-ray tomogram within 10 min or less, regardless of the complexity of the dataset. Hence, highly complex datasets such as mammalian cells can be segmented in a very short period of time. Even a semimanual segmentation requires significantly more hands-on time by the user. Using early stopping as well as maintaining sparsity in the annotated features in the training data, we intentionally guided the training of the network such that the risk of false positives was minimal, reasoning that false-positive labels would further increase the connectivity of the objects and thus require significant additional manual user input for separating objects and correcting the segmentation. We are convinced that future iterations with additional training data may recognize significantly more features, while maintaining a low false-positive rate.

Based on our segmentation evaluation, only about 1.8% or 1.6% of the automatic label pixels were either false positives or reconstruction artifacts on average per slice. These erroneous labels could at most times easily be distinguished from cellular features, since they were either localized in extracellular space and/or on the direct boundaries of the reconstructed volume. The appearance of these errors can be further reduced prior to classification without loss of information by cropping the dataset by a few pixels on the boundaries to avoid most reconstruction artifacts and by using a rough mask to differentiate the cellular volume from extracellular
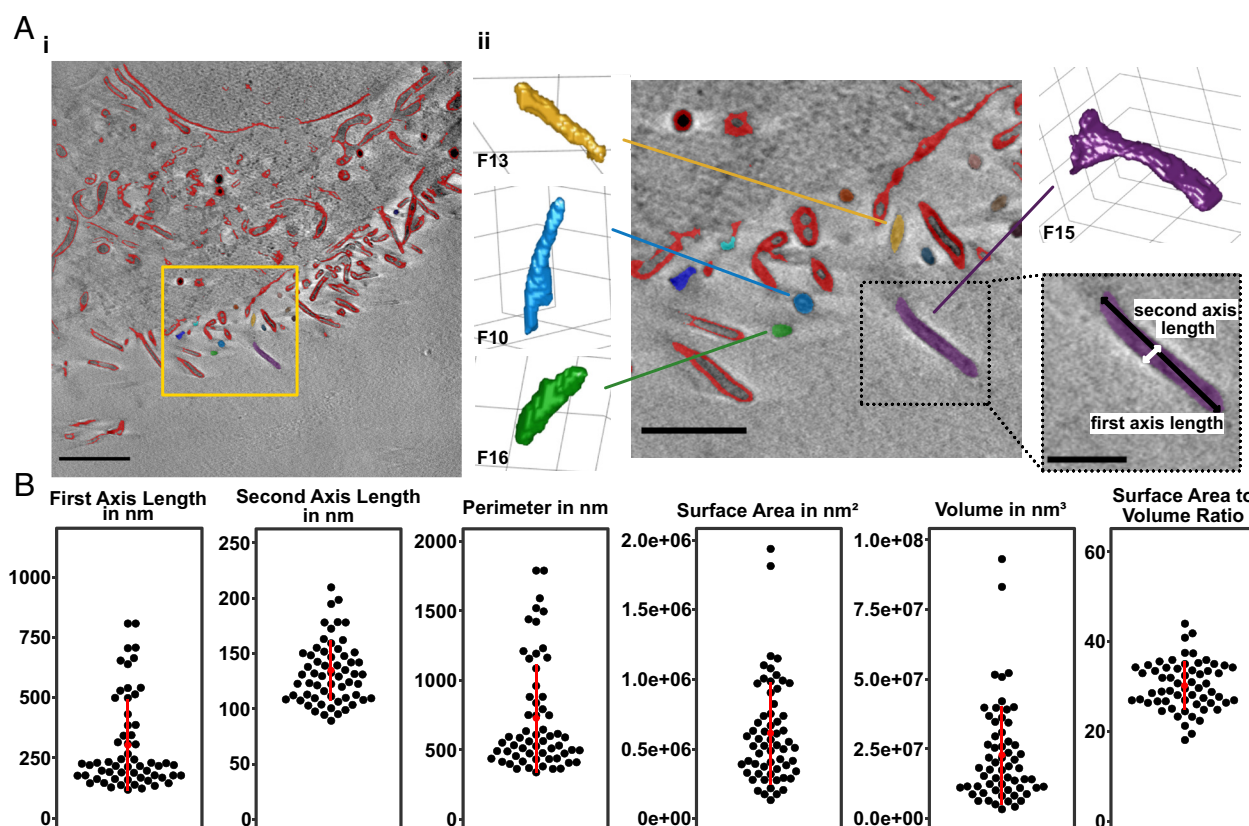
Fig. 3. Quantification of filopodial 3D structure. (*A*) Representative tomographic slice of an analyzed dataset, overlaid with the automatically generated labels in red (*i*). The yellow box in *i* denotes the area shown up close in *ii*. A selection of individual filopodia used for analysis are shown in isosurface renderings, illustrating the general rod-like morphology of the filopodia. The box with dashed outline illustrates the orientation of first and second axis lengths on a representative filopodium. (*B*) Dot-plots illustrating the mean first and second axis lengths and perimeter of the 2D cross-sections of the filopodia, as well as the 3D surface area, volume and surface area to volume ratio of all of the filopodia analyzed in this study. The red dot indicates the mean, and the bar denotes the SD of the respective parameter across the population (n = 59). (Scale bars, (*i*) 2 µm, (*ii*) 1 µm, and (*ii*) *Inset* 0.5 µm.)

space to reduce the presence of false positives. *SI Appendix*, Fig. S3 illustrates examples of the falsely labeled reconstruction artifacts and false positives that were the most common in our datasets. However, we must also acknowledge the potential introduction of bias by having a human user interpret and classify automatically generated labels.

It is possible to perform the automatic segmentation with additional filtering to remove small objects, or to perform several iterations of erosion, leading to increasingly thinner contours and better object separation, as well as increasing the rate of false negatives. Depending on the intended use of the segmentation, i.e., rapid inspection vs. accurate analysis, as well as the size of the smallest objects of interest, tweaking of filtering and erosion settings may be useful for optimal results.

Filopodia Analysis. To illustrate the utility of our deep network, we quantified several spatial parameters of a high number of filopodia. We found that 2D analysis is not sufficient for the accurate determination of critical parameters, but 3D analysis after fast reconstruction through our algorithm allows for the swift determination of filopodial volume and surface area. These values have a high SD, since the measured population represents filopodia in various states of extension and retraction of these highly dynamic and often transient structures (57). The measurement of filopodial dimensions is of great interest, due to their important role in cell migration and cancer (56, 58–61). Our method will allow for a fast and convenient readout at the nanoscale and opens broad avenues of research on filopodial growth.

Advantages of our DL Approach. Our proposed deep network provides several key advantages for segmenting cryo-SXT data:

Using this network, a segmentation can be produced in less than 10 min on a modern GPU. This dramatically simplifies and accelerates the data analysis step to the point where data can already be analyzed at the synchrotron during the acquisition of the tomogram. This saves not only a significant amount of time, but also enables researchers to learn from their data faster and potentially adjust critical sample parameters before acquiring the next datasets, thereby increasing the quality of the data they obtain. This is especially important since cryo-SXT is predominantly available at synchrotrons, which only allows for limited amounts of beamtime.

The manual input required for obtaining a 3D segmentation of the data is significantly reduced, saving time and providing more objective segmentation results. The time required for a full segmentation of a dataset is considerable, so that users may focus their segmentation efforts on only a part of their data, looking for an expected result. An unbiased, automated segmentation can help in revealing more subtle or unexpected changes in the cells under the experimental conditions. At the same time, it becomes possible to generate statistically relevant amounts of data in a reasonable timeframe. This allows to take advantage of the full potential of cryo-SXT in indiscriminately visualizing all carbon-dense cellular structures, such as membranous and cytoskeletal structures.

Our network can be retrained with new training data. This should allow to improve the network's performance under specific tomographic reconstruction conditions, or when different cell types other than adherent mammalian cells are being investigated. Thanks to the high processing speed and accuracy, it is much faster to generate a high number of training data based on preliminary segmentations, instead of manually segmenting a comparable amount of data. This can also be helpful for experimenting with other DL tools.

The output data of the network can be exported as common file formats, including .TIFF format and thus easily imported into a variety of programs commonly used by biologists, such as FIJI (62) or Microscopy Image Browser (MIB) (63). This ensures an easier integration into the user's individual image processing and analysis workflow.

Remaining Challenges. A few issues remain to be addressed in the future. While cryo-SXT as a method allows for rapid acquisition of a large portion of a cell, this imaging technique is still predominantly available at only a few synchrotron facilities worldwide. Laboratory-based cryo-soft X-ray microscopes are under development, but are currently not capable of achieving results of comparable quality. The limited accessibility of this technique, together with the diversity of applications and generated datasets results in a scarcity of comparable training data for effective training, which was one of the central challenges of this work.

The interior ultrastructure of mammalian cells is often highly complex, with many organelles in close proximity or even in direct contact with each other. This high degree of connectivity among the organelles poses a problem of object separation. This network was designed to rapidly and objectively segment all membranous structures of the cell. A user interested in separately quantifying one organelle from another must still invest manual effort in classifying and separating the objects (i.e., the organelles) from each other. This can be observed in *SI Appendix*, Figs. S6 and S7, where several organelles were joined together during the 3D segmentation procedure. The high speed and accuracy of our automated segmentation will now help to generate sufficient training data for such a classification network, which will learn to distinguish the major organelle types present in mammalian cells and assist in separating objects.

Finally, our automated procedure can be further improved to address the complications introduced by the *missing wedge problem.* This problem arises from incomplete rotation of a tomographic specimen, and cannot be avoided when imaging specimens grown on a flat substrate. Reconstruction of data with a missing wedge leads to blurring, elongation, and ray artifacts, which may obscure actual cellular features. As a consequence, cellular features distorted by missing-wedge artifacts become very difficult to discern, even for the human eye. Such cases formed only a small part of our training data space, and so our network is not yet sufficiently well trained to compensate for these artifacts. This will be overcome as the network is exposed to more and more data which suffer from such artifacts, and so with time the network will learn to more accurately segment features obscured by missing-wedge artifacts.

Conclusion

In conclusion, we present here a convolutional neural network that dramatically reduces the manual effort and time required for segmentation of cryo-soft X-ray tomographic datasets. We show that the network correctly recognizes about 70% of the features present in different datasets correctly and thus allows for an instant assessment of the dataset. This will allow soft X-ray microscope users to analyze their data more rapidly, efficiently, and objectively, and thereby optimize their microscope time, which is constrained by the heavy demand for these instruments at synchrotron beamlines. Furthermore, our work opens the door to the generation of statistically relevant amounts of data on large cytosolic volumes from many cells. This will be of great importance in the quantitative

investigation of the organization and reorganization of the cellular organelle complement.

Methods

Sample Preparation. Crandell-Rees feline kidney (CRFK) cells were cultivated in DMEM with 4,5 g/l D-glucose, supplemented with 2 mM glutamine and 10% FCS inside a humidified incubator at 37 °C. For preparing specimens, Quantifoil R2/2 AuG200F1 finder or Au-HZB-2 grids (Quantifoil Micro Tools GmbH, Germany) were first cleaned for 30 s at maximum power in a Diener Zepto plasma cleaner (Diener Electronic, Germany), then sterilized in 70% ethanol and placed inside empty wells of a 24-well plate. After letting the remaining ethanol evaporate, fresh full medium was added and subsequently CRFK cells were seeded, such that they had reached a confluency of 70 to 80 % the next day, when the cells were either fixed for two hours at room temperature in freshly prepared 2% glutaraldehyde or directly plunge-frozen in liquid ethane. The frozen specimens were stored in liquid nitrogen until imaging.

Cryo-Soft X-Ray Tomography and Tomographic Reconstruction. All acquisitions were performed at the U41 beamline at the BESSY II synchrotron facility, unless specifically mentioned otherwise. Tilt images ranging from ± 60° (round Quantifoil grids) or ± 65° (HZB-2 grids) with an increment of 1° were acquired at 510 eV with a 25 nm Fresnel Zone Plate (FZP). Tracking, alignment, and tomographic reconstruction were performed using 30 iterations of the simultaneous iterative reconstruction technique (64) in IMOD (65), using gold nanoparticles as fiducial markers.

Deep Network for Image Segmentation. The model shown in Fig. 1 is an end-to-end convolutional feed-forward network, taking 2D slices of cryo-SXT tomograms, and processing them to two concurrent outputs, respectively, from "image" and "annotation" decoders (48, 50). Convolutional layers of the network apply a convolution with a learnable kernel plus an additive bias, followed by the non-linear activation function g. Thus, for image X being transformed by the network, $X_{n+1} = g(W_n * X_n + b_n)$, with W_n and b_n being the trainable kernel and the bias, respectively. We have used rectified linear unit (ReLU) as the activation function g throughout the network. Each individual convolution step in the network is composed of mini "bottle-neck" convolutional blocks, as described by He *et al.* (50). In these blocks, the number of image channels is reduced by a convolution step with a kernel of size 1 × 1, followed by the main convolution with a 3 × 3 kernel, and a mapping back to the desired output channels by another 1 × 1 kernel. This setup reduces the number of trainable parameters in the network, while enhancing the abstraction learned by each convolution step.

Data Pipeline. The manually annotated data consist of 79 pairs of raw 1324 × 1284 pixel slices from a single tomogram and their corresponding manually fully annotated binary labels. For semisupervised training, these data are complemented with 513 unannotated slices from 17 different tomograms, for which manual labels are not produced. These unannotated data are used for training the image-reconstruction branch of the network.

The network is trained on 256 × 256 pixel images, or image/label pairs, depending on which dataset the input data are taken from. We designed an image augmentation procedure, which randomly crops the input slices into these smaller images, while additionally applying random 90° rotations and horizontal/vertical reflections (*SI Appendix*, Fig. S1B). We have implemented the entire data preparation and augmentation pipeline using the TensorFlow (66) Dataset API, which benefits from automated GPU processing, hence offering uninterrupted training of the network while augmented data are prepared on-the-fly. The stochastic manner in which the training data are prepared and fed to the network results in a more robust prediction as well as less chance of overfitting. The original data are split 90 to 10% between training and hold-out test sets before the augmentation, with the quality of image reconstruction and the accuracy and precision of label prediction verified on the test set.

Training and Optimization. For the decoder performing image segmentation, we have used the pointwise Huber loss between manual (Y) and predicted (X) annotations (67),

$$L_H = \begin{cases} \frac{1}{2}(X-Y)^2, & |X-Y| < \delta \\ -\frac{1}{2}\delta^2 + \delta \cdot |X-Y|, & \text{otherwise} \end{cases}$$

where L_H is the Huber loss, and δ is a scalar parameter, which we set to 0.1 in training our model. This function provides the robustness of the l_1-norm with constant gradients, with well-behaving near-zero gradients similar to the l_2-norm. We found Huber loss to produce the sharpest segmentations with minimal artifacts. For the image reconstruction decoder, we have used as the loss function a combination of the Huber function and the structural similarity index (SSIM), proposed by Wang and Sheikh (68),

$$L_S = 1 - \text{SSIM}(X, Y),$$

$$\text{SSIM}(X, Y) = \frac{2\mu_X\mu_Y + C_1}{\mu_X^2 + \mu_Y^2 + C_1} \cdot \frac{2\sigma_{XY} + C_2}{\sigma_X^2 + \sigma_Y^2 + C_2},$$

where L_S is the image structural similarity loss, μ_X and μ_Y are the mean pixel intensities of the respective images, σ_X and σ_Y are the corresponding SDs, and σ_{XY} is the covariance between the two images. C_1 and C_2 are parameters of SSIM, which we, respectively, chose as 0.01 and 0.03. In SSIM, mean and (co)variances are measured locally using a Gaussian kernel. We chose the size of the Gaussian kernel to be 11 pixels, with a SD of 1.5 pixels. This choice has been backed by the systematic study of Zhao *et al.* on the loss functions for image reconstruction (69).

During the forward pass and the subsequent backpropagation, gradients of the loss function with respect to trainable parameters are calculated and used by the optimizer algorithm to update these parameters (22). We monitor the training procedure by comparing the loss between training and 10% validation datasets. We stop the training loop when the validation loss plateaus, which implies the network might begin to overfit the training data.

The performance of the trained network is evaluated based on the following scores:

$$accuracy = \frac{true\ pos. + true\ neg.}{all\ pos. + all\ neg.}, \quad precision = \frac{true\ pos.}{all\ pos.},$$

$$recall = \frac{true\ pos.}{true\ pos. + false\ neg.}, \quad F\text{-}Score = \frac{2}{precision^{-1} + recall^{-1}}.$$

Hyperparameter Optimization. We considered three classes of hyperparameters to be manually optimized based on the test data. i) the decision between the presence or the absence of the image decoder resulting in the supervised-only versus semisupervised learning. We showed the significance improvement resulting from this choice (Fig. 2A), resulting in the decision to move forward with the semi-supervised approach. ii) the model complexity pertaining to the number of convolutional kernels considered in each layer of the network. We considered four choices for this hyperparameter, controlled by a tunable scaling parameter that determines the number of kernels when setting up the network layers. Our choices result in the network being set up with the range of 1 million to 76 million trainable parameters (*SI Appendix*, Fig. S7A). Both the accuracy and the precision of the segmentation task improve with the model complexity in the studied range. Neither of the two scores show hints of overfitting (*SI Appendix*, Fig. S7A). The improvement however has an asymptotic trend, seemingly saturating with models more complex than what we used here. iii) the learning rate used for training the network. We experimented with values in the range 10^{-5} to 0.001, and found the best result in terms of the accuracy and the precision to be obtained for the value of 10^{-4}. With the learning rate of 0.001 and higher, the network essentially fails to train (*SI Appendix*, Fig. S7B).

3D Volumetric Reconstruction. The two-dimensional outputs of the segmentation network provide a voxelized representation of the membranes in the tomogram. We have applied the well-established marching-cube algorithm to obtain surface reconstructions from the voxels (70). The implementation from the SciPy package has been used for this purpose (71). The PyVista python package has been used for processing of the 3D data (72). Final renderings are achieved via the POV-Ray package (73).

Output Export and Import into Software for Downstream Applications. The output labels generated by the deep network were exported as stacks

of binary mask images in .TIFF format. These stacks were then imported into MIB (63). By thresholding, the binary labels could be quickly selected and used as a mask to be overlaid on the original input image. This result was saved as a .MODEL file and could be used for inspection of segmentation quality, label classification, manual complementation of labels, and quantification of objects.

Evaluating Segmentation Quality. In order to determine the quality of the automatic segmentation results, an expert user manually inspected the predicted labels of 59 randomly selected slices of three representative tomographic datasets, covering the entire range of the datasets by setting appropriate intervals between the slices. The labels generated by the network were overlaid with the corresponding, original slice and then classified into correct labels (a feature was being recognized correctly), false positives (a label was placed where no feature was present) and reconstruction artifacts (a label was placed where the cryo SXT tomographic reconstruction algorithm produced an obvious image artifact). Cellular features that were missing in the network's segmentation were manually labeled and classified as false negatives. The evaluation was performed by scoring the number of pixels of each category for each slice analyzed versus the total amount of labeled pixels. The inspection, classification, and evaluation of slices were all performed in MIB (63). Due to the missing wedge problem that arises in tomographic techniques with incomplete sample rotation and the limited depth of focus of the 25 nm FZP, some apparent features were present but too blurred for a human user to confidently trace the feature contour with a four pixel brush tool. These features were excluded from the analysis, since training a network to recognize such features would lead to a significant increase in false positives.

Filopodia Analysis. A total of sixty automatically labeled filopodia from three different datasets were selected for analysis. Filopodia had to be covered entirely by the reconstructed volume, i.e., filopodia protruding beyond the x, y, or z boundaries of the field of view were excluded, since 3D parameters cannot be conclusively determined from incomplete objects. Similarly, filopodia partially interrupted by e.g., reconstruction artifacts from gold fiducials were also excluded, since these artifacts prevent unambiguous labeling of underlying structures in tomographic datasets.

The automatically generated labels were inspected individually for each filopodium, ensuring that the entire structure from its base at the plasma membrane to the tip of its protrusion was correctly labeled by the network. While doing so, the object was manually separated from any adjacent labels, which arose most frequently most frequently at the filopodial juncture with the plasma membrane. Any open loops of the labels were closed and the contour's shape filled, such that the volume and surface area values could be determined accurately. This process took approximately 2 h for each of the three tomograms, with most of the time required for separating the filopodia labels from the plasma membrane labels. The isolated and filled objects were then quantified both as a series of 2D objects and as a fitted ellipsoid around the 3D object using the built-in quantification toolbox in MIB (16). For 2D quantification of each individual filopodium, the mean and SD of all values in all slices of its component 2D objects were determined. The final quantification of all filopodia is based on the mean and SD of all 59 individual means (2D) or 3D parameters, respectively. For calculating the nanometer-scale values, an isotropic voxel size of 9.8 nm was used.

Author affiliations: [a]Institute of Chemistry and Biochemistry, Department of Biology, Chemistry and Pharmacy, Free University of Berlin, 14195 Berlin, Germany; [b]Artificial Intelligence of the Sciences Group, Department of Mathematics and Informatics, Free University of Berlin, 14195 Berlin, Germany; [c]Helmholtz Zentrum Berlin für Materialien und Energie GmbH, 12489 Berlin, Germany; and [d]Institute of Physics, Humboldt Universität zu Berlin, 12489 Berlin, Germany

1. B. Kepsutlu *et al.*, Cells undergo major changes in the quantity of cytoplasmic organelles after uptake of gold nanoparticles with biologically relevant surface coatings. *ACS Nano* **14**, 2248–2264 (2020).
2. F. J. Chichón *et al.*, Cryo X-ray nano-tomography of vaccinia virus infected cells. *J. Struct. Biol.* **177**, 202–211 (2012).
3. G. Schneider *et al.*, Three-dimensional cellular ultrastructure resolved by X-ray microscopy. *Nat. Methods* **7**, 985–987 (2010).
4. D. Y. Parkinson, G. McDermott, L. D. Etkin, M. A. Le Gros, C. A. Larabell, Quantitative 3-D imaging of eukaryotic cells using soft X-ray tomography. *J. Struct. Biol.* **162**, 380–386 (2008).
5. R. Carzaniga, M.-C. Domart, L. M. Collinson, E. Duke, Cryo-soft X-ray tomography: A journey into the world of the native-state cell. *Protoplasma* **251**, 449–458 (2014).
6. J. Groen, J. J. Conesa, R. Valcárcel, E. Pereiro, The cellular landscape by cryo soft X-ray tomography. *Biophys. Rev.* **11**, 611–619 (2019).
7. N. Korogod, C. C. Petersen, G. W. Knott, Ultrastructural analysis of adult mouse neocortex comparing aldehyde perfusion with cryo fixation. *eLife* **4**, e05793 (2015).
8. A. Perktold, B. Zechmann, G. Daum, G. Zellnig, Organelle association visualized by three dimensional ultrastructural imaging of the yeast cell. *FEMS Yeast Res.* **7**, 629–638 (2007).
9. L. Heinrich *et al.*, Whole-cell organelle segmentation in volume electron microscopy. *Nature* **599**, 141–146 (2021), 10.1038/s41586-021-03977-3.
10. M. Chen *et al.*, Convolutional neural networks for automated annotation of cellular cryo-electron tomograms. *Nat. Methods* **14**, 983–985 (2017), 10.1038/nmeth.4405.
11. M. Cortese *et al.*, Integrative imaging reveals SARS-CoV-2-induced reshaping of subcellular morphologies. *Cell Host. Microbe.* **28**, 853–866.e5 (2020).
12. M. Schaffer *et al.*, A cryo-FIB lift-out technique enables molecular-resolution cryo-ET within native caenorhabditis elegans tissue. *Nat. Methods* **16**, 757–762 (2019).
13. V. Baena *et al.*, FIB-SEM as a volume electron microscopy approach to study cellular architectures in SARS-CoV-2 and other viral infections: A practical primer for a virologist. *Viruses* **13**, 611 (2021).
14. M. Uchida *et al.*, Quantitative analysis of yeast internal architecture using soft x-ray tomography. *Yeast* **28**, 227–236 (2011).
15. S. Kapishnikov *et al.*, Unraveling heme detoxification in the malaria parasite by in situ correlative X-ray fluorescence microscopy and soft X-ray tomography. *Sci. Rep.* **7**, 7610 (2017).
16. I. Kounatidis *et al.*, 3D correlative cryo-structured illumination fluorescence and soft x-ray microscopy elucidates reovirus intracellular release pathway. *Cell* **182**, 515–530.e17 (2020).
17. J. J. Conesa *et al.*, Unambiguous intracellular localization and quantification of a potent iridium anticancer compound by correlative 3D cryo x-ray imaging. *Angew. Chemie Int. Ed.* **59**, 1270–1278 (2020).
18. V. Loconte *et al.*, Using soft X-ray tomography for rapid whole-cell quantitative imaging of SARS-CoV-2-infected cells. *Cell Rep. Methods* **1**, 100117 (2021).
19. Y. Rivenson *et al.*, Deep learning microscopy. *Optica* **4**, 1437–1443 (2017).
20. E. Nehme, L. E. Weiss, T. Michaeli, Y. Shechtman, Deep-STORM: Super-resolution single-molecule microscopy by deep learning. *Optica* **5**, 458–464 (2018).
21. H. Wang *et al.*, Deep learning enables cross-modality super-resolution in fluorescence microscopy. *Nat. Methods* **16**, 103–110 (2019).
22. Z. Wang, J. Chen, S. C. H. Hoi, Deep learning for image super-resolution: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **43**, 3365–3387 (2021).
23. L. Fang *et al.*, Deep learning-based point-scanning super-resolution imaging. *Nat. Methods* **18**, 406–416 (2021).
24. X. Chai *et al.*, Deep learning for irregularly and regularly missing data reconstruction. *Sci. Rep.* **10**, 3302 (2020).
25. C. Sommer, R. Hoefler, M. Samwer, D. W. Gerlich, A deep learning and novelty detection framework for rapid phenotyping in high-content screening. *MBoC* **28**, 3428–3436 (2017).
26. O. Ronneberger, P. Fischer, T. Brox, "U-Net: Convolutional networks for biomedical image segmentation" in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, N. Navab, J. Hornegger, W. M. Wells, A. Frangi, Eds. (Springer International Publishing, 2015), pp. 234–241, 10.1007/978-3-319-24574-4_28.
27. T. Falk *et al.*, U-Net: Deep learning for cell counting, detection, and morphometry. *Nat. Methods* **16**, 67–70 (2019).
28. R. Conrad, K. Narayan, CEM500K, a large-scale heterogeneous unlabeled cellular electron microscopy image dataset for deep learning. *eLife* **10**, e65894 (2021).

29. M. D. Guay et al., Dense cellular segmentation for EM using 2D–3D neural network ensembles. *Sci. Rep.* **11**, 2561 (2021).
30. E. Gómez-de-Mariscal et al., DeepImageJ: A user-friendly environment to run deep learning models in ImageJ. *Nat. Methods* **18**, 1192–1195 (2021).
31. L. von Chamier et al., Democratising deep learning for microscopy with ZeroCostDL4Mic. *Nat. Commun.* **12**, 2276 (2021).
32. I. Belevich, E. Jokitalo, DeepMIB: User-friendly and open-source software for training of deep learning network for biological image segmentation. *PLoS Comput. Biol.* **17**, e1008374 (2021).
33. S. Berg et al., ilastik: Interactive machine learning for (bio)image analysis. *Nat. Methods* **16**, 1226–1232 (2019).
34. T. Plautz et al., Progress toward automatic segmentation of soft x-ray tomograms using convolutional neural networks. *Microsc. Microanal.* **23**, 984–985 (2017).
35. A. Ekman, J.-H. Chen, G. M. Dermott, M. A. L. Gros, C. Larabell, Task based semantic segmentation of soft x-ray CT images using 3D convolutional neural networks. *Microsc. Microanal.* **26**, 3152–3154 (2020).
36. I. Luengo et al., SuRVoS: Super-region volume segmentation workbench. *J. Struct. Biol.* **198**, 43–53 (2017).
37. K. L. Nahas, J. F. Fernandes, C. Crump, S. Graham, M. Harkiolaki, Contour, a semi-automated segmentation and quantitation tool for cryo-soft-X-ray tomography. *Biol. Imaging* **17**, e3 (2021), 10.1101/2021.12.03.470962.
38. A. Pennington et al., SuRVoS 2: Accelerating annotation and segmentation for large volumetric bioimage workflows across modalities and scales. *Front. Cell Dev. Biol.* **10**, 842342 (2022).
39. A. J. Koster et al., Perspectives of molecular and cellular electron tomography. *J. Struct. Biol.* **120**, 276–308 (1997).
40. P. Ercius, O. Alaidi, M. J. Rames, G. Ren, Electron tomography: A three-dimensional analytic tool for hard and soft materials research. *Adv. Mater.* **27**, 5638–5663 (2015).
41. Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, O. Ronneberger, 3D U-Net: Learning dense volumetric segmentation from sparse annotation. arXiv [Preprint] (2016). https://doi.org/10.48550/arXiv.1606.06650 (Accessed 5 May 2023).
42. P. D. Lösel et al., Introducing biomedisa as an open-source online platform for biomedical image segmentation. *Nat. Commun.* **11**, 5577 (2020).
43. D. M. Pelt, J. A. Sethian, A mixed-scale dense convolutional neural network for image analysis. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 254–259 (2018).
44. A. Dosovitskiy et al., An image is worth 16x16 Words: Transformers for image recognition at scale (2021). https://doi.org/10.48550/.2010.11929.
45. A. Kirillov et al., Segment anything (2023). https://doi.org/10.48550/2304.02643. (Accessed 5 May 2023).
46. P. Lösel, V. Heuveline, "Enhancing a diffusion algorithm for 4D image segmentation using local information" in *Medical Imaging 2016: Image Processing* (SPIE, 2016), vol. 9784, pp. 707–717.
47. Z. Liu et al., A ConvNet for the 2020s (2022). https://doi.org/10.48550/2201.03545. (Accessed 5 May 2023).
48. I. Goodfellow, Y. Bengio, A. Courville, *Deep Learning* (MIT Press, 2016).
49. F. Milletari, N. Navab, S.-A. Ahmadi, "V-Net: Fully convolutional neural networks for volumetric medical image segmentation" in *2016 Fourth International Conference on 3D Vision (3DV)* (2016), pp. 565–571, 10.1109/3DV.2016.79.
50. K. He, X. Zhang, S. Ren, J. Sun, "Deep residual learning for image recognition" in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (2016), pp. 770–778, 10.1109/CVPR.2016.90.
51. K. Lee, J. Zung, P. Li, V. Jain, H. S. Seung, Superhuman accuracy on the SNEMI3D connectomics challenge. arXiv [Preprint] (2017). https://doi.org/10.48550/arXiv.1706.00120 (Accessed 5 May 2023).
52. L. Yu, X. Yang, H. Chen, J. Qin, P. A. Heng, "Volumetric ConvNets with mixed residual connections for automated prostate segmentation from 3D MR images" in *Proceedings of the AAAI Conference on Artificial Intelligence* (2017), vol. 31.
53. S. Ioffe, C. Szegedy, Batch normalization: Accelerating deep network training by reducing internal covariate shift. arXiv [Preprint] (2015). https://doi.org/10.48550/arXiv.1502.03167 (Accessed 5 May 2023).
54. S. Xie, R. Girshick, P. Dollár, Z. Tu, K. He", Aggregated residual transformations for deep neural networks" in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2017), pp. 5987–5995, 10.1109/CVPR.2017.634.
55. A. Araujo, W. Norris, J. Sim, Computing receptive fields of convolutional neural networks. *Distill* **4**, e21 (2019).
56. G. Jacquemet, H. Hamidi, J. Ivaska, Filopodia in cell adhesion, 3D migration and cancer cell invasion. *Curr. Opin. Cell Biol.* **36**, 23–31 (2015).
57. T. Bornschlögl, How filopodia pull: What we know about the mechanics and dynamics of filopodia. *Cytoskeleton* **70**, 590–603 (2013).
58. Q. Zhang et al., PTENε suppresses tumor metastasis through regulation of filopodia formation. *EMBO J.* **40**, e105806 (2021).
59. S. L. Gupton, F. B. Gertler, Filopodia: The fingers that do the walking. *Sciences STKE* 2007, re5 (2007).
60. J. L. Gallop, Filopodia and their links with membrane traffic and cell adhesion. *Semin. Cell Dev. Biol.* **102**, 81–89 (2020).
61. G. Jacquemet et al., FiloQuant reveals increased filopodia density during breast cancer progression. *J. Cell Biol.* **216**, 3387–3403 (2017).
62. J. Schindelin et al., Fiji: An open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012).
63. I. Belevich, M. Joensuu, D. Kumar, H. Vihinen, E. Jokitalo, Microscopy image browser: A platform for segmentation and analysis of multidimensional datasets. *PLOS Biol.* **14**, e1002340 (2016).
64. P. Gilbert, Iterative methods for the three-dimensional reconstruction of an object from projections. *J. Theoret. Biol.* **36**, 105–117 (1972).
65. J. R. Kremer, D. N. Mastronarde, J. R. McIntosh, Computer visualization of three-dimensional image data using IMOD. *J. Struct. Biol.* **116**, 71–76 (1996).
66. TensorFlow, Developers zenodo (2022). 10.5281/zenodo.6574269. (Accessed 5 May 2023).
67. P. J. Huber, "Robust estimation of a location parameter" in *Breakthroughs in Statistics: Methodology and Distribution*, S. Kotz, N. Johnson, Eds. (Springer, 1992), pp. 492–518. 10.1007/978-1-4612-4380-9_35.
68. Z. Wang, A. C. Bovik, H. R. Sheikh, E. P. Simoncelli, Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Image Process.* **13**, 600–612 (2004).
69. H. Zhao, O. Gallo, I. Frosio, J. Kautz, Loss functions for neural networks for image processing. arXiv [Preprint] (2018). https://doi.org/10.48550/arXiv.1511.08861 (Accessed 5 May 2023).
70. W. E. Lorensen, H. E. Cline, "Marching cubes: A high resolution 3D surface construction algorithm" in *Proceedings of the 14th annual conference on Computer graphics and interactive techniques* (Association for Computing Machinery, 1987), pp. 163–169, 10.1145/37401.37422.
71. P. Virtanen et al., SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nat Methods* **17**, 261–272 (2020).
72. C. B. Sullivan, A. A. Kaszynski, PyVista: 3D plotting and mesh analysis through a streamlined interface for the Visualization Toolkit (VTK). *J. Open Source Software* **4**, 1450 (2019).
73. POV-Ray, The persistence of vision raytracer. https://www.povray.org/. (Accessed 5 May 2023).
74. M. C. A. Dyhr et al., deep_sxt. *Github*. https://github.com/noegroup/deep_sxt. Deposited 19 April 2023.
75. M. C. A. Dyhr et al., 3D-surface reconstruction of cellular cryo-soft X-ray microscopy tomograms using semi-supervised deep learning. EMPIAR. https://www.ebi.ac.uk/empiar/EMPIAR-11392/. (Accessed 26 January 2023).
76. M. C. A. Dyhr et al., Datasets for *training*, validation and hyperparameter *optimization* of the deep *network*. Refubium. https://refubium.fu-berlin.de/handle/fub188/37508. (Accessed 7 February 2023).