doi:10.1017/S1431927620023983
Microsc. Microanal. 26 (Suppl 2), 2020

Task Based Semantic Segmentation of Soft X-ray CT Images Using 3D Convolutional Neural Networks

Axel Ekman[1], Jian-Hua Chen[1], Gerry Mc Dermott[1], Mark A. Le Gros[2] and Carolyn Larabell[2]

[1]Lawrence Berkeley National Laboratory and UCSF, Berkeley, California, United States, [2]Molecular Biophysics and Integrated Bioimaging Division, Lawrence Berkeley National Laboratory, Berkeley, California, United States

Semantic segmentation refers to the process of linking each pixel in an image to a class label, for example, cell phenotype, membrane, nucleus, or mitochondria. In practice, semantic segmentation simplifies a tomographic reconstruction and enables quantifiable analysis of the organelles in the cell, such as density differences, spatial distance metrics such as distances between organelles, and morphometrics. In Soft X-ray Tomography (SXT), segmentation is based on the measured Linear Absorption Coefficient (LAC) and guided and confirmed by structural cues, sub-cellular location, together with data from other imaging modalities. Currently, segmentation is a time-consuming, mostly manual process that often depends on specialist knowledge of the specimen to identify features in the reconstruction. To address this bottleneck, and make segmentation less dependent on user effort and expertise, we aim to incorporate automated, machine learning segmentation algorithms into our existing data processing and analysis pipeline.

To date, NCXT staff and users have segmented thousands of individual cells. Since each segmentation represents many hours or even days of effort, the accumulated information and data is an enormously valuable resource for training machine learning algorithms. We have begun taking advantage of this resource by developing a data-driven machine learning segmentation pipeline Fig. 1. Our aim is to refine our segmentation pipeline to improve accuracy and set up a task-based framework, where users can define the required semantic information for their problem, and get a tailor-made algorithm that draws on all the available reference data for that specific task.

In their paper on handwritten digit recognition, et al. [1] state that "Classical work in visual pattern recognition has demonstrated the advantage of extracting local features and combining them to form higher-order features." Convolution Neural Networks (CNNs) provide a general functional space, where both the extracted features and their combinations are embedded into a single, trainable minimization function. CNNs are essentially neural networks that replace --- to some degree --- fully connected layers with convolutional operators [2]. That is, the network consists of a nonlinear combination of parameterized convolution kernels, which are "learned" through training the network for a specific task. Due to the shared-weights architecture and translation invariance of the convolutional operator, CNNs reduce the number of parameters in the model, making the training, i.e., optimization of the function, more robust [2].

In our work, we chose to use a fully 3D U-Net-type [3] CNN. To circumvent the problem of preserving both high-level features and high resolution, we use cascading networks (as e.g. in [4]). Here we train two (or more) models that we use sequentially for the final result. A highly downsampled image is used for a coarse map of the segmented feature. By downsampling the data we can increase the relative receptive field and use deeper network architectures, as the memory requirement is much lower. The result of the first network can then be used as *a priori* information for a second network, that is trained by using sub-

patches of the whole image. This network input is then a multi-channel image with e.g. the original gray value image and the *n* probability fields from the output of the first CNN. An example of a result of a cascading CNN is shown in Fig. 2.

The methods we will investigate/develop in this aim could greatly increase the complexity and depth of analysis that can be accomplished in a given period of time. Automated segmentation will allow users to carry out SXT analyses that are not currently feasible due to the time and effort required to analyze large numbers of SXT reconstructions.

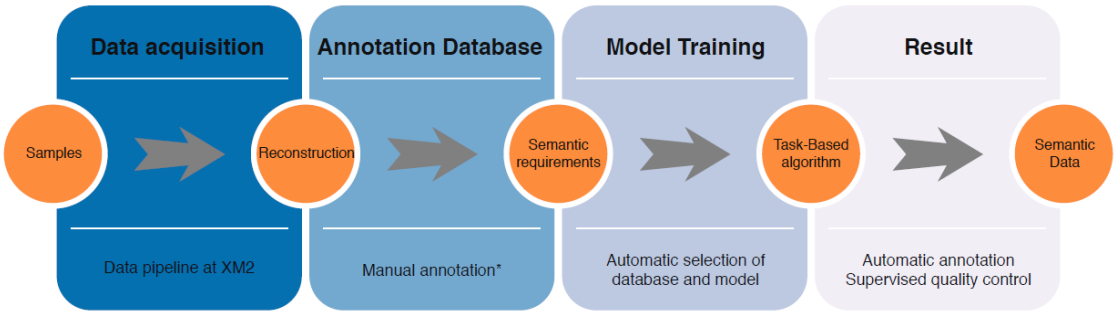
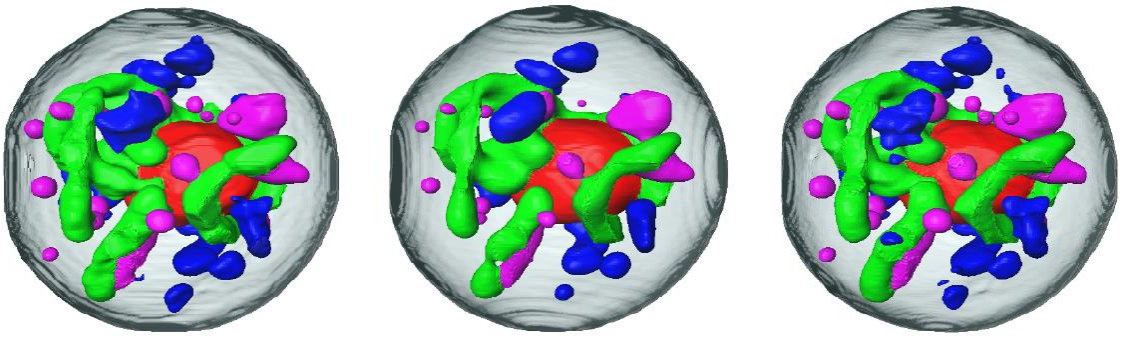


Figure 1. Each research question comes with a different set of semantic needs. The aim is to establish an easy-to-use pipeline that can collect the necessary data and train a task-specific algorithm that is tailor-made for the specific questions being asked.*For data that differ substantially from the samples already in the database, some manual input to update the existing database.



(a) Manual segmentation (b) Low-Resolution CNN (c) Cascading CNN

Figure 2. Example of a multi-label segmentation of the green alga *Chromochloris zofingiensis* showing the nucleus (red), lipid bodies (magenta), mitochondria (green) and starch granules (blue). In cases, where a full-resolution CNN model image is too large to fit in a single GPU full detail can be preserved by subsequently refining the model with a cascading second model.

References

[1] Le Cun et al. Handwritten digit recognition: Applications of neural network chips and automatic learning. IEEE Communications Magazine, 27(11):41–46, 1989.
[2] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. Deep Learning. MIT Press, 2016. http://www.deeplearningbook.org.
[3] Özgün Çiçek et al. C¸ ic¸ek. 3d u-net: Learning dense volumetric segmentation fromsparse annotation. InMedical

https://doi.org/10.1017/S1431927620023983 Published online by Cambridge University Press

Image Computing and Computer-Assisted Intervention – MICCAI 2016, pages 424–432, Cham, 2016. Springer International Publishing.
[4] Fabian Isensee et al. nnu-net: Self-adapting framework for u-net-based medical image segmentation, 2018.

https://doi.org/10.1017/S1431927620023983 Published online by Cambridge University Press