2022 44th Annual International Conference of
the IEEE Engineering in Medicine & Biology Society (EMBC)
Scottish Event Campus, Glasgow, UK, July 11-15, 2022

2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC) | 978-1-7281-2782-8/22/$31.00 ©2022 IEEE | DOI: 10.1109/EMBC48229.2022.9870879

Axon Tracing and Centerline Detection using Topologically-Aware 3D U-Nets

Dylan Pollack[1], Lars A. Gjesteby[1], Michael Snyder[1], David Chavez[1], Lee Kamentsky[2],
Kwanghun Chung[2], Laura J. Brattain[1]

Abstract— **As advances in microscopy imaging provide an ever clearer window into the human brain, accurate reconstruction of neural connectivity can yield valuable insight into the relationship between brain structure and function. However, human manual tracing is a slow and laborious task, and requires domain expertise. Automated methods are thus needed to enable rapid and accurate analysis at scale. In this paper, we explored deep neural networks for dense axon tracing and incorporated axon topological information into the loss function with a goal to improve the performance on both voxel-based segmentation and axon centerline detection. We evaluated three approaches using a modified 3D U-Net architecture trained on a mouse brain dataset imaged with light sheet microscopy and achieved a 10% increase in axon tracing accuracy over previous methods. Furthermore, the addition of centerline awareness in the loss function outperformed the baseline approach across all metrics, including a boost in Rand Index by 8%.**

I. INTRODUCTION

As high-resolution brain imaging techniques continue to improve, one of the biggest challenges in generating connectivity maps of the human brain is the automatic identification of cellular structures from raw imagery [1]. Manual human annotation is thought to be accurate, but will not scale to reconstructing projection profiles of the billions of neurons required to fully map a single human brain. Therefore, it is necessary to develop methods of tracing neurons that rival human performance but with minimal human intervention.

One approach that has yielded human-like performance on biomedical image segmentation tasks is supervised learning using deep convolutional neural networks (CNNs). In particular, the U-Net architecture has achieved state-of-the-art results, exceeding the threshold of human performance in the SNEMI3D Connectomics Challenge in 2017 [2], [3]. To obtain a connectivity map from a segmented subcellular-resolution brain image, the segmentation mask can be processed using downstream morphological thinning and graph extraction algorithms [4]. However, artifacts uncovered during these post-processing steps, such as fragments where gaps are mistakenly introduced or incorrect merging of distinct fibers, are not visible to the segmentation model, limiting its ability to learn more accurate representations.

In recent years, there have been several deep CNN architectures proposed to merge the steps of semantic segmentation, morphological thinning, and path identification commonly done separately in a broader class of "tubular segmentation" problems. It has been observed that across varied computer vision tasks including neurite tracing, blood vessel segmentation, and aerial road mapping, a common objective is the recovery of curvilinear spatial trajectories, rather than a full, pixel-wise segmentation [5]. While neurons and roads have completely different morphologies, the techniques needed to accurately map network connectivity may be similar.

This paper makes the following contributions:

- We adapted two recent deep learning-based segmentation techniques making use of ground truth centerline information to the task of axon tracing in light sheet microscopy data:
  1) CasNet: A cascading segmentation and centerline detection network previously used for road detection [6]
  2) soft-clDice: An alternative loss function that incorporates parameter-free skeletonization directly into the training procedure [7]
- We demonstrated an increase in both segmentation and centerline detection performance over a baseline of voxel-wise classification followed by morphological thinning as a post-processing step

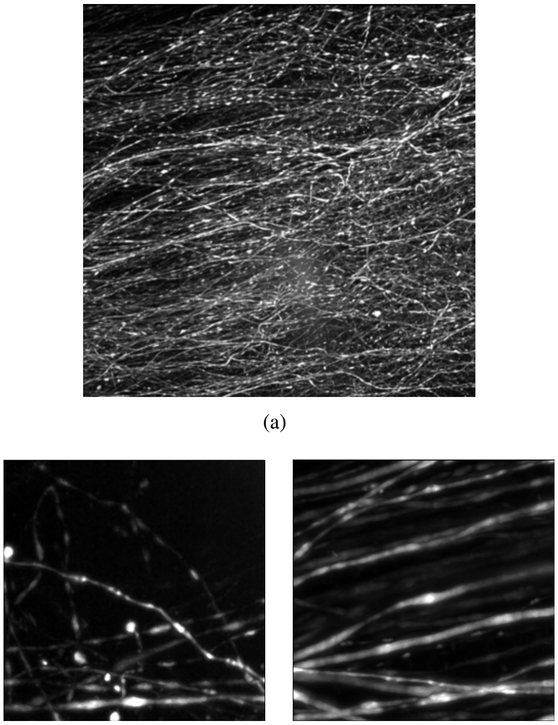
II. METHODS AND MATERIALS

A. Dataset

For algorithm development, we used a light sheet microscopy dataset imaged from a piece of mouse brain tissue prepared under 3× expansion, with a stain targeting Parvalbumin positive neurons from the globus pallidus externus (PVGPe). These methods stabilize the tissue with clear hydrogels that preserve biomolecules and enable removal of lipids, rendering unstained portions of the sample optically transparent [8]. The full PVGPe volume is 2048×2048×1271 voxels, with a voxel resolution of 0.6×0.6×2 μm, but only a 256×256×206 voxel (148×148×412 μm) subvolume was annotated manually (Fig. 1). Two labelers traced axon fibers in the annotation subvolume using ImageJ [9]. Following previous experiments on this dataset [10], we preprocessed the imagery by clipping the highest and lowest 0.01% of values, applying a median filter, and scaling between 0 and 1. The dataset is subdivided into contiguous training (50%), validation (25%), and testing (25%) regions, which were fed into the model as 128×128×64 voxel samples.

[1] Lincoln Laboratory, Massachusetts Institute of Technology, Lexington, MA, USA Lars.Gjesteby@ll.mit.edu

[2] Institute for Medical Engineering and Science, Massachusetts Institute of Technology, Cambridge, MA, USA khchung@mit.edu



(a)

(b) (c)

Fig. 1. Maximum intensity projections along the Z-axis of the PVGPe dataset: a) Full volume; b) Labeled subvolume (Region 1); c) Unlabeled subvolume (Region 2).

B. Architectures

1) 3D U-Net: The U-Net architecture consists of convolutional layers interspersed with downsampling to form sequentially lower-resolution modules in an encoder path, followed by upsampling to restore the original resolution in a decoder path, with lateral skip connections between same-resolution encoder and decoder modules [2]. We used 3×3×3 convolutional layers followed by group normalization and exponential linear units. Summation joining was chosen for skip connections with an additional skip connection between the first and last convolutional layer in each module. We also used strided 2×2×2 max-pooling for downsampling, and strided transpose convolutions with max-pooling for upsampling [3], [11] . In our baseline approach, we trained a residual 3D U-Net with 4 resolution blocks to perform voxel-wise classification using binary cross entropy loss between the output and training images (Fig. 2(a)). The resulting segmentation was then binarized using a threshold of 0.5, and skeletonized using morphological thinning to obtain a single voxel wide centerline [12].

2) CasNet: Originally proposed for the task of road mapping, a cascaded CNN can be used to perform simultaneous segmentation and centerline prediction [6]. In this approach, the output of a CNN trained to perform semantic segmentation provides the input to a second, simpler CNN that directly predicts centerline pixels. During back propagation, the segmentation network accumulates gradients of the loss functions for both tasks, forcing it learn a representation retaining information that is useful for centerline detection as well. We adapted a CasNet architecture consisting of a depth-4 (i.e. 4 resolution blocks) 3D U-Net for segmentation, cascading into a depth-3 3D U-Net for centerline prediction (Fig. 2(b)). The loss function is the sum of the segmentation loss, computed between the output of the upstream network and the training image, and the centerline detection loss, computed between the output of the downstream network and a thin binary skeleton extracted from the training image. Cheng et al. [6] used cross entropy loss for both terms; however, due to the extreme class imbalance present in the 3D centerline dataset, we chose modified Dice loss instead [13].

3) 3D U-Net + clDice: Recently, soft centerline-Dice (clDice) was proposed as a loss function for segmentation problems emphasizing preservation of topology [7]. clDice between a predicted segmentation (V_P) and a ground truth segmentation (V_L) and their extracted skeletons (S_P and S_L, respectively) is defined as the harmonic mean between topological precision (T_{prec}) and topological sensitivity (T_{sens}):

$$T_{prec} = \frac{|S_P \cap V_L|}{|S_P|} \tag{1}$$

$$T_{sens} = \frac{|S_L \cap V_P|}{|S_L|} \tag{2}$$

$$clDice = 2 \times \frac{T_{prec} \times T_{sens}}{T_{prec} + T_{sens}} \tag{3}$$

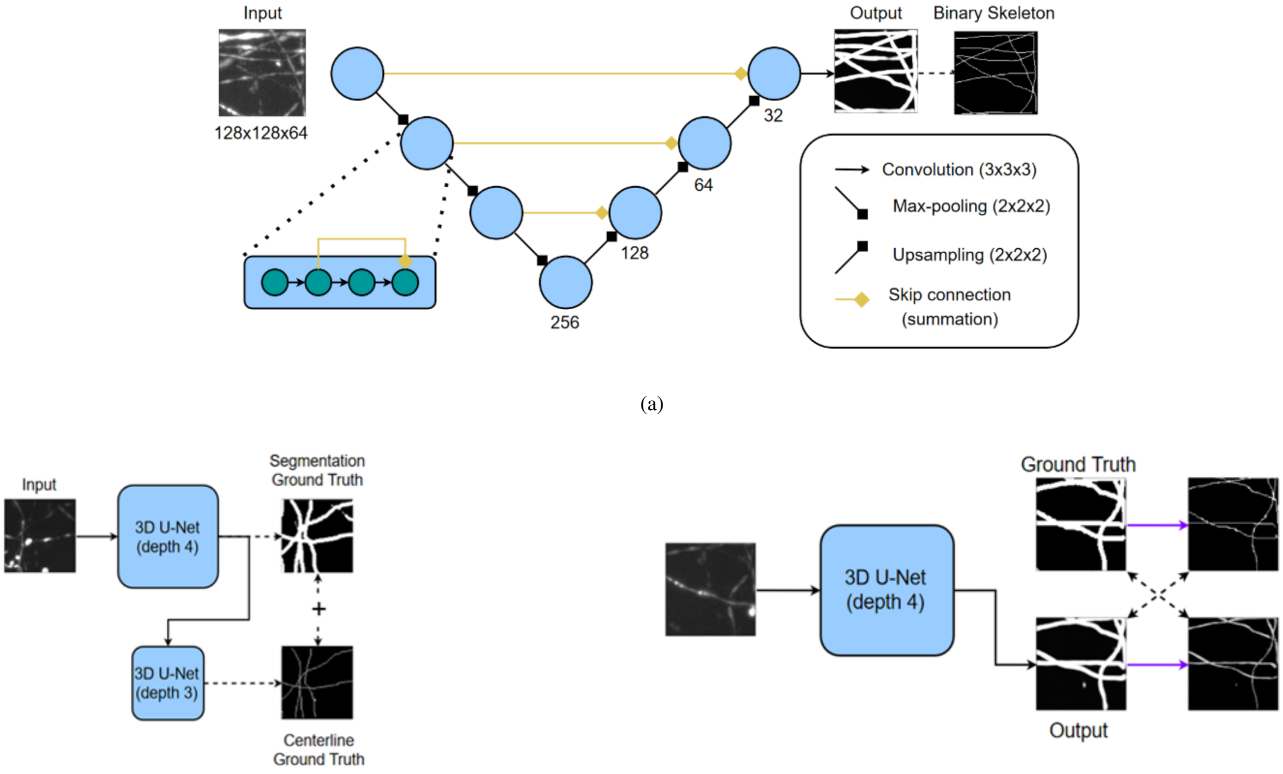
Specifically, T_{prec} is the proportion of the predicted skeleton that lies within the ground truth segmentation, and T_{sens} is the proportion of the ground truth skeleton recovered by the predicted segmentation. Since the morphological dilation and erosion operations traditionally used to extract pixel-wide skeletons are not differentiable, a "soft-skeletonization" operation was also introduced by using iterative min- and max-pooling to achieve a similar effect. Finally, a loss function combining clDice and Dice terms was proposed:

$$L = \alpha(1 - clDice) + (1 - \alpha)(1 - Dice) \tag{4}$$

To test this method for axon tracing using our own data, we used the baseline 3D U-Net architecture, but replaced binary cross entropy with soft-clDice loss, using $\alpha = 0.5$ (Fig. 2(c)).

C. Training

Each candidate architecture was trained via stochastic gradient descent using the ADAM optimizer [14] with an initial learning rate of 1×10^{-4} and weight decay of 1×10^{-3}. For training, mini batches of 16 samples were randomly cropped from the input dataset, and augmented using random grayscale perturbations, random 90° X-Y rotations, and random flipping. For inference, samples were cropped using a sliding window with 50% overlap and blended using a 3D Hann window to minimize segmentation errors along the edge of the receptive field. We also incorporated test-time augmentation, by averaging predictions across 16



(a)

(b)

(c)

Fig. 2. a) Baseline approach uses a residual 3D U-Net architecture to segment the image, followed by binary thresholding and morphological thinning to obtain a voxel-wide skeleton. b) Cascading 3D U-Net (CasNet) passes its segmentation output into a smaller 3D U-Net that predicts centerline voxels. c) Soft-clDice thins the 3D U-Net segmentation output using soft-skeletonization (purple arrows) so that centerline information can be incorporated into the loss function.

transformations for each input sample (unique combinations of X-Y 90° rotations and flipping along each axis). Test-time augmentation has been shown to improve segmentation performance at the expense of a linear increase in inference time [2]. For each model, training proceeded until convergence or until no improvement in validation loss was observed over a period of 20 epochs. To account for noise due to random initialization and the relatively small test set size, each experiment was repeated 10 times.

III. RESULTS

A. Evaluation Metrics

Segmentation performance was evaluated using three metrics: 1) Dice coefficient, which measures voxel-wise similarity between the ground truth and predicted binary segmentation masks, 2) clDice, which emphasizes homotopy-equivalence between ground truth and predicted segments, and 3) the adjusted-for-chance Rand Index (ARI) [15], which measures agreement between ground truth and predicted clusterings. Clusters were obtained by finding the connected components of foreground structures.

Centerline detection was evaluated using a variation of Dice score in which any prediction falling within ρ voxels of the ground truth is considered a true positive. The motivation behind this metric, which we will refer to as ρ-Dice, is to measure centerline accuracy in a way that does not excessively penalize minor deviations from the ground truth [6]. It is formulated as the inverse of clDice, defining ρ-precision and ρ-sensitivity as:

$$\rho_{prec} = \frac{|S_P \cap S_{L,\rho}|}{|S_P|} \tag{5}$$

$$\rho_{sens} = \frac{|S_L \cap S_{P,\rho}|}{|S_L|} \tag{6}$$

$$\rho\text{-}Dice = 2 \times \frac{\rho_{prec} \times \rho_{sens}}{\rho_{prec} + \rho_{sens}} \tag{7}$$

where $S_{L,\rho}$ and $S_{P,\rho}$ denote the ground truth and predicted skeletons, respectively, following binary dilation by ρ voxels. Centerline predictions are always single-voxel-wide skeletons obtained via morphological thinning of a model's output.

B. Experimental Results

The mean and standard deviation for each metric over 10 trials of model training and testing (Table I) were reported. For CasNet, which makes two predictions, we computed ρ-Dice using its downstream centerline detection output, and Dice and clDice using its segmentation output. 3D U-Net +

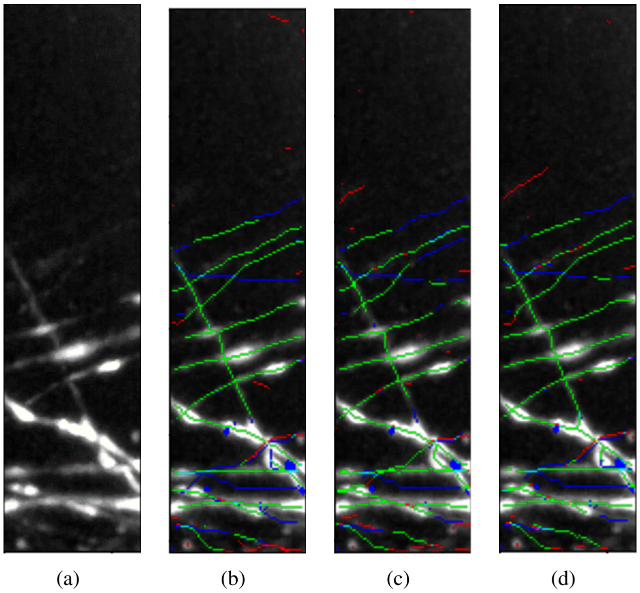

(a) (b) (c) (d)

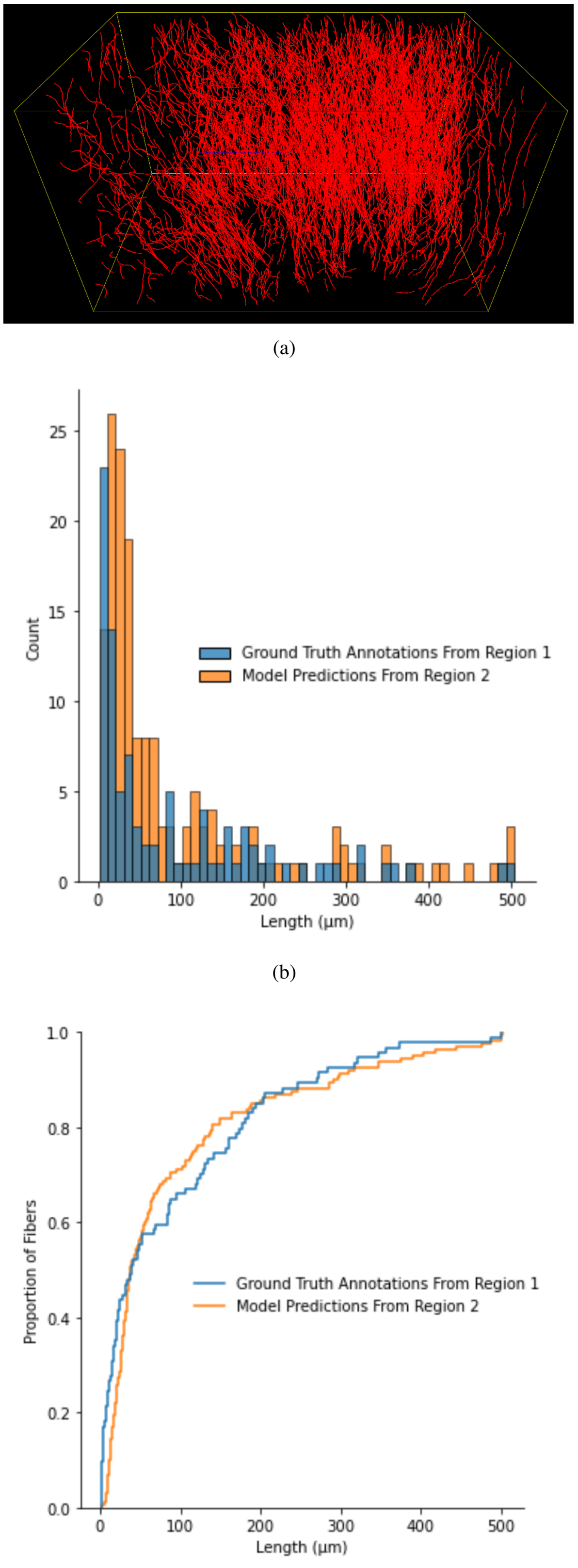
Fig. 3. a) Maximum intensity projection along the Z-axis of test volume. b) 3D U-Net, c) CasNet, and d) 3D U-Net + clDice tracing results overlaid with true positive centerline predictions (green), false negatives (blue) and false positives (red). 3D U-Net + clDice results show fewer fragmenting errors caused by false negative predictions.

clDice outperformed both CasNet and the baseline 3D U-Net across every metric, and the Dice score exceeded the best score previously reported for segmentation on the same dataset [10] by 10%.

Qualitative results for centerline detection are shown in Fig. 3. The fiber tracings from 3D U-Net + clDice (Fig. 3(d)) show fewer splitting errors caused by false negative predictions. In Fig. 4(a), a 3D view of fiber tracings generated by applying the trained 3D U-Net + clDice to the full PVGPe volume are displayed. The shortest 10% of tracing lengths are filtered out to suppress noise and improve visual clarity. Empirical distributions of fiber lengths (Fig. 4(b) and 4(c)) were computed for ground truth tracings in the annotated subvolume (Region 1,, Fig. 1(b)), as well as for automated tracings on a nearby, previously unlabeled subvolume of identical size (Region 2, Fig. 1(c)). It can been seen that the fiber distribution from the automated tracing resemble closely that from the manual tracing.

IV. DISCUSSION

Soft-clDice loss appears to improve both segmentation and centerline detection of dense axon imagery from light sheet microscopy. Interestingly, it was noted that the motivation behind clDice was to improve topology preservation in semantic segmentation tasks, rather than to directly improve centerline detection [7]. In tasks where only centerline detection is relevant (e.g. for connectivity mapping), it is possible that the loss function could be further refined. For example, using the ρ-Dice formulation and max-pooling for dilation, a model could be trained to directly predict the foreground skeleton while tolerating slight errors in spatial trajectory. Such an approach would have similarities



(a)

(b)

(c)

Fig. 4. a) 3D view of full volume algorithm-predicted tracings, with the shortest 10% of fiber segments filtered out. b) Histograms and c) Empirical cumulative distributions of fiber lengths recovered by ground truth tracings from the labeled subvolume (Region 1, , Fig. 1(b)) and algorithm-predicted tracings from a separate subvolume of identical size (Region 2, , Fig. 1(c)).

TABLE I

DICE, CLDICE, ρ-DICE, AND ADJUSTED RAND INDEX (ARI) ARE REPORTED AS MEAN AND STANDARD DEVIATION FOR 10 TRIALS OF EACH EXPERIMENT. 3D U-NET + CLDICE PERFORMS BEST ACROSS EACH METRIC, AS INDICATED BY BOLD.

	Dice		*clDice*		*ρ-Dice ($\rho = 3$)*		*ARI*	
	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.
3D U-Net	0.580	0.027	0.720	0.020	0.761	0.015	0.575	0.027
CasNet	0.598	0.002	0.763	0.010	0.750	0.004	0.593	0.002
3D U-Net + clDice	**0.628**	0.018	**0.785**	0.018	**0.787**	0.021	**0.623**	0.018

to previous work treating centerline prediction as regression on distance from the true centerline [16]. Alternatively, because max-pooling could mask splitting errors in predicted centerlines, indirect foreground dilation could be achieved via soft-skeletonization of the background segment.

There are conceptual similarities between the CasNet and 3D U-Net + clDice approaches: the centerline detection network in CasNet, which is trained to predict the skeleton of the segmentation output, can be seen as a "black box" approximation to morphological thinning, while clDice uses min- and max-pooling operations to achieve the same effect. However, clDice has no tunable parameters associated with skeletonization and therefore may be less prone to overfitting.

Since the virally labeled PV-expressing cells have many sub-classes with different projection patterns, there are variations in neurite morphology that may be missed by only training on the labeled data. Previous experiments on the PVGPe dataset showed an improvement in segmentation performance by pretraining a 3D U-Net on a self-supervised learning task using unlabeled data. Combining the topologically-aware approaches demonstrated in this paper with self-supervised learning on unlabeled data could yield further improvements with more generalizability.

V. CONCLUSION

In this paper, we explored deep learning-based axon tracing with topological information incorporated into the loss function. We evaluated three approaches using a modified 3D U-Net architecture trained on a mouse brain dataset imaged with light sheet microscopy and achieved a 10% increase in axon tracing accuracy over previous methods. Furthermore, the addition of centerline awareness in the loss function outperformed baseline approach across all metrics, including a boost in the Rand Index by 8%.

ACKNOWLEDGMENT

The authors would like to thank Webster Guan, Juhyuk Park, and Adam Michaleas. This material is based upon work supported by the National Institutes of Health (NIH) U01MH117072. Any opinions, findings, conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of NIH.

REFERENCES

[1] J. W. Lichtman, H. Pfister, and N. Shavit, "The big data challenges of connectomics," *Nature neuroscience*, vol. 17, no. 11, pp. 1448–1454, 2014.

[2] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3D U-Net: learning dense volumetric segmentation from sparse annotation," in *International conference on medical image computing and computer-assisted intervention*. Springer, 2016, pp. 424–432.

[3] K. Lee, J. Zung, P. Li, V. Jain, and H. S. Seung, "Superhuman accuracy on the snemi3d connectomics challenge," *arXiv preprint arXiv:1706.00120*, 2017.

[4] M. Hernandez, A. Brewster, L. Thul, B. A. Telfer, A. Majumdar, H. Choi, T. Ku, K. Chung, and L. J. Brattain, "Learning-based long-range axon tracing in dense scenes," in *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*. IEEE, 2018, pp. 1578–1582.

[5] A. Mosinska, M. Koziński, and P. Fua, "Joint segmentation and path classification of curvilinear structures," *IEEE transactions on pattern analysis and machine intelligence*, vol. 42, no. 6, pp. 1515–1521, 2019.

[6] G. Cheng, Y. Wang, S. Xu, H. Wang, S. Xiang, and C. Pan, "Automatic road detection and centerline extraction via cascaded end-to-end convolutional neural network," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 6, pp. 3322–3337, 2017.

[7] S. Shit, J. C. Paetzold, A. Sekuboyina, I. Ezhov, A. Unger, A. Zhylka, J. P. Pluim, U. Bauer, and B. H. Menze, "cldice-a novel topology-preserving loss function for tubular structure segmentation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 16 560–16 569.

[8] T. Ku, J. Swaney, J.-Y. Park, A. Albanese, E. Murray, J. H. Cho, Y.-G. Park, V. Mangena, J. Chen, and K. Chung, "Multiplexed and scalable super-resolution imaging of three-dimensional protein localization in size-adjustable tissues," *Nature biotechnology*, vol. 34, no. 9, pp. 973–981, 2016.

[9] J. Schindelin, I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, S. Preibisch, C. Rueden, S. Saalfeld, B. Schmid *et al.*, "Fiji: an open-source platform for biological-image analysis," *Nature methods*, vol. 9, no. 7, pp. 676–682, 2012.

[10] T. Klinghoffer, P. Morales, Y.-G. Park, N. Evans, K. Chung, and L. J. Brattain, "Self-supervised feature extraction for 3d axon segmentation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2020, pp. 978–979.

[11] A. Wolny, L. Cerrone, A. Vijayan, R. Tofanelli, A. V. Barro, M. Louveaux, C. Wenzl, S. Strauss, D. Wilson-Sánchez, R. Lymbouridou *et al.*, "Accurate and versatile 3d segmentation of plant tissues at cellular resolution," *Elife*, vol. 9, p. e57613, 2020.

[12] T.-C. Lee, R. L. Kashyap, and C.-N. Chu, "Building skeleton models via 3-d medial surface axis thinning algorithms," *CVGIP: Graphical Models and Image Processing*, vol. 56, no. 6, pp. 462–478, 1994.

[13] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in *2016 fourth international conference on 3D vision (3DV)*. IEEE, 2016, pp. 565–571.

[14] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[15] L. Hubert and P. Arabie, "Comparing partitions," *Journal of Classification*, vol. 2, no. 1, p. 193–218, Dec 1985.

[16] A. Sironi, V. Lepetit, and P. Fua, "Multiscale centerline detection by learning a scale-space distance transform," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 2697–2704.