MULTI-SCALE MODELING OF NEURAL STRUCTURE IN X-RAY IMAGERY

Aishwarya Balwani[1,*]   *Joseph Miano*[1,*]   *Ran Liu*[1]   *Lindsey Kitchell*[2]

Judy A. Prasad[3]   *Erik C. Johnson*[2]   *William Gray-Roncal*[2]   *Eva L. Dyer*[1,4,+]

1 - Georgia Institute of Technology   2 - Johns Hopkins University Applied Physics Laboratory
3 - University of North Carolina at Chapel Hill   4 - Emory University

ABSTRACT

Methods for resolving the brain's microstructure are rapidly improving, allowing us to image large brain volumes at high resolutions. As a result, the interrogation of samples spanning multiple diversified brain regions is becoming increasingly common. Understanding these samples often requires multi-scale processing: segmentation of the detailed microstructure and large-scale modelling of the macrostructure. Current brain mapping algorithms often analyze data only at a single scale, and optimization for each scale occurs independently, potentially limiting the consistency, performance, and interpretability. In this work we introduce a deep learning framework for segmentation of brain structure at multiple scales. We leverage a modified U-Net architecture with a multi-task learning objective and unsupervised pre-training to simultaneously model both the micro and macro architecture of the brain. We successfully apply our methods to a heterogeneous, three-dimensional, X-ray micro-CT dataset spanning multiple regions in the mouse brain, and show that our approach consistently outperforms another multi-task architecture, and is competitive with strong single-task baselines at both scales.

Index Terms— Multi-task learning, brain mapping, segmentation, neuroimaging, X-ray microtomography.

1. INTRODUCTION

Studies of brain structure and function often require us to map out and analyse the neuroanatomy across various regions of interest (ROIs) [1, 2, 3]. Consequently, methods in neuroimaging have been making steady progress towards pipelines that can capture multi-area and even whole-brain volumes at high resolutions [4, 5, 6]. The analyses of such large-scale, diversified neural datasets however, bring with them a host of problems [7, 8, 9] such as limited or partial views of the data, imaging and physical artifacts, as well as the need for huge amounts of compute and automation. Furthermore, the heterogeneity in the underlying neuronal components within and across brain areas, in terms of both structure and distribution, makes the characterisation of the individual components as well as that of the various regions they constitute incredibly challenging.

Over the past few years, deep learning approaches have yielded highly accurate, robust, and generalisable solutions for problems of structural identification and characterisation in brain mapping pipelines [10, 11, 12, 13]. However, most of these methods focus on only single-scale structure, either segmenting neural components and processes from high-resolution images of relatively small brain samples [10, 11], or parcellation of ROIs from relatively low-resolution images of large multi-area or whole brains samples [12, 13]. The practice of conducting these two tasks independently, however, is highly inefficient in terms of the time spent creating, training, and deploying such models, especially as datasets get larger. Additionally, very often the two types of segmentations serve as prerequisites for one another; By resolving neural components in a sample, *microstructural segmentation* aids in the overall characterisation of different (macro) brain areas. Likewise, by first resolving and identifying different ROIs, *macrostructural segmentation* is crucial for detailing the distributional properties of (micro) structural compositions across various neurological conditions of age, health, etc. It is, therefore, desirable to solve the two modeling tasks simultaneously in a way that leverages structural features present in the data across multiple spatial scales.

In this work we present a deep learning approach for modelling neural architecture across multiple spatial scales in a data-driven and generalizable manner, using a multi-task learning based framework. Specifically, we leverage soft-parameter sharing to combine a model of the *macrostructure*, that is learnt by discriminating between different brain areas in the sample [14, 15], with that of the *microstructure*, which is learnt by semantically segment different neural components present in the sample, into a single parameterized framework trained end-to-end with a two-step, warm start procedure. We empirically show that our proposed multi-task architecture and training methodology consistently perform at the level of, or better than strong single-task baselines at both scales. We also address certain modifications needed to allow for large-scale deployment of our model, and provide results of the same to establish our methods' generalizability and robustness to neural heterogeneity across a large, three-

---

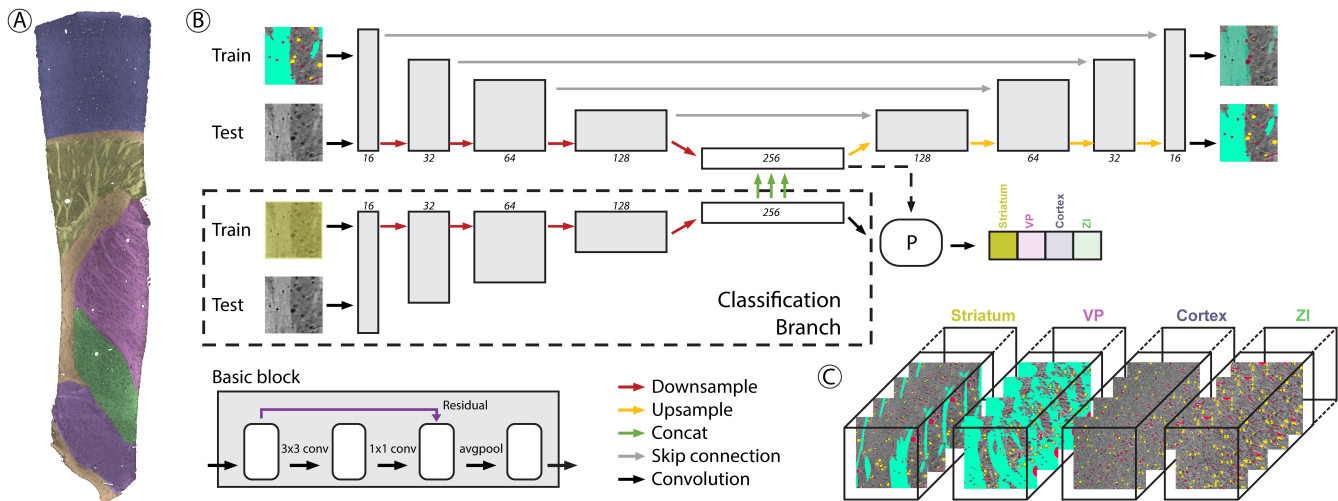*Equal contribution. [+]Corresponding author – evadyer@gatech.edu

Fig. 1. *Overview of multi-task architecture for modeling neuroanatomy across spatial scales.* (A) shows a 2D slice of the data with macroscale annotations (different colours represent different ROIs). (B) depicts our proposed Double U-Net architecture; Top = Microscale segmentation half of the network. Bottom (dotted region) = Macroscale feature extractor and projection head. The high-level, multi-scale latent representation shared by the two halves is shown in the middle (on either side of the green arrows). To the bottom left, we show the basic block of our U-Net, which consists of a stack of 3x3 convolutional filters, followed by a stack of 1x1 convolutional filters, and average pooling. We also include a residual connection between the outputs of the 3x3 filters and corresponding inputs of the pooling layers. The projection head following the classification branch comprises of a series of three convolutional layers (kernel sizes = 5x5, 4x4, 1x1), and is denoted by the P-block. The non-linearity used throughout the entire architecture is leaky ReLU (negative slope = 0.01). (C) shows examples of microscale annotations across our four ROIs (green = myelinated axons, red = blood vessels, yellow = cells).

dimensional imaged sample. To this end, we also demonstrate the use of a nearest-neighbour based clean-up algorithm that enforces local consensus amongst macroscale predictions and boosts our performance on its segmentation significantly.

2. METHODS

Dataset and annotations. To build a model of the neuroanatomy at multiple scales, we work with a new, publicly accessible three-dimensional X-ray microCT dataset [16] that studies the Agmon-Connors slice [17]. The images span multiple ROIs from the somatosensory cortex to the ventral posterior thalamic nucleus of the mouse brain (∼5.9 gigavoxels at 1.56μm isotropic resolution), and come with annotations of both, the micro and macrostructure, making the dataset a good test bed for analysing multi-scale heterogeneity of neural structure across diverse brain areas. The *macroscale* annotations [18] identify six different ROIs, viz. the cortex (CTX), striatum (STR), ventral posterior (VP), zona incerta (ZI), hypothalamus (HYP), and white matter (WM) across eight evenly spaced (∼78μm) slices along the depth of the imaged dataset (Fig. 1A). The *microscale* annotations [19] identify myelinated axons, blood vessels (BV), and cell bodies densely labelled in blocks of 256x256 pixels across eleven evenly spaced (∼46μm) slices along the depth of the imaged sample (Fig. 1C) in the CTX, STR, VP, and ZI.

Given our aim of jointly modelling neural structure across multiple spatial scales, we choose to work with data from the four ROIs where the micro and macrostructural annotations

overlap. At the microstructural scale, we split all 256x256 annotated blocks into non-overlapping 128x128 blocks, resulting in a total of 512, 128, and 64 images for training, validation, and testing respectively. We similarly sample 128x128 images at the macrostructural scale from all slices with area-level annotations, but restricting ourselves to only those which are contained strictly within an ROI, resulting in a total of 1109, 197, and 165 images for training, validation, and testing respectively.

Double U-Net architecture. To perform both brain-area classification and microstructural semantic segmentation simultaneously, we modify the classic U-Net architecture proposed by Ronnerberger et al. [20] in two key ways. First, we bolster the core U-Net (Fig. 1B - Top) with additional residual connections [21] that allow for better propagation of signal and gradients throughout the network and help it achieve high levels of accuracy [22]. Next, we add a second encoder to the core U-Net, and use it as our *macroscale feature extractor* or *classification branch* (Fig. 1B - Bottom), thus giving our new architecture its name – the **Double U-Net**.

The rationale behind our choice of architecture is intuitive; Features used for microstructural segmentation need to be more sensitive and "local" as compared to those required for discriminating between entire brain areas, thereby prompting the use of two separate encoders for the different tasks. However, since both tasks are learnt on the same data and would benefit from having access to features describing different aspects of the microstructure, we allow the encoders to share information at a sufficiently high level by concate-

nating the latent vectors of the classification branch with that of the core Double U-Net encoder. Finally, we append the bottleneck of the encoder used to model the macroscale neuroanatomy with a series of convolutional layers that we call the *projection head* (i.e., P-block). The use of such layers is in step with the current representation learning literature, having been shown to encourage learning of better representations which achieve higher downstream task accuracies [23].

Training procedure for multi-task learning. Our model is trained end-to-end in two steps. First, we train the network using only *macroscale* information, wherein given an input image $\mathbf{x}$, the loss function minimized is

$$\mathcal{L} = \ell(\mathbf{y}_c, g_c(\mathbf{x})) + \lambda \cdot \text{MSE}(\mathbf{x}, g_r(\mathbf{x})) \quad (1)$$

Here $\ell(\cdot)$ is the cross-entropy between $\mathbf{y}_c$ (brain area label of $\mathbf{x}$) and $g_c(\mathbf{x})$ (brain area predicted by the model). The second term measures the mean squared error (MSE) between $\mathbf{x}$ and the reconstructed image $g_r(\mathbf{x})$ produced by the segmentation branch of Double U-Net. λ is a tunable hyper-parameter. The network is then trained over 100 epochs in the first step, and the model yielding the lowest validation loss is saved. Next, we freeze the weights of the classification encoder and proceed to (re)train the core U-Net using the *microscale* information over 300 epochs with the segmentation loss

$$\mathcal{L} = \ell(\mathbf{y}_s, \widehat{\mathbf{y}}) \quad (2)$$

where $\ell(.)$ is the pixel-wise cross-entropy between $\mathbf{y}_s$ (microscale annotation of $\mathbf{x}$) and $\widehat{\mathbf{y}}$ (semantically segmented output given by the core branch of Double U-Net, i.e., g_r).

Our proposed architecture and two-step training procedure give us the following advantages: i) We leverage macroscale annotations to perform microscale segmentation, thus improving performance on the latter, significantly harder task. ii) Pre-training the model with self-supervised reconstruction allows us to train the model in a generalizable and robust manner with *limited* microscale data.

Architecture	Axon	BV	Cell	Bg	Avg
SegNet	0.66	0.54	0.63	0.91	0.69
U-Net	0.78	0.60	0.70	0.93	0.75
MT U-Net	0.79	0.51	0.68	0.91	0.72
Double U-Net	0.79	0.59	0.70	0.92	0.75

Architecture	STR	VP	CTX	ZI	Avg
U-Net Encoder	0.97	0.92	0.99	0.81	0.93
MT U-Net	0.90	0.86	0.96	0.64	0.84
Double U-Net	0.95	0.92	0.99	0.81	0.92

Table 1. Comparison of f1 scores across architectures for microscale (top), and macroscale (bottom) segmentation on the test set.

Clean-up algorithm to improve macroscale segmentation performance. Even though our proposed architecture has high macroscale prediction accuracy, we find that there often
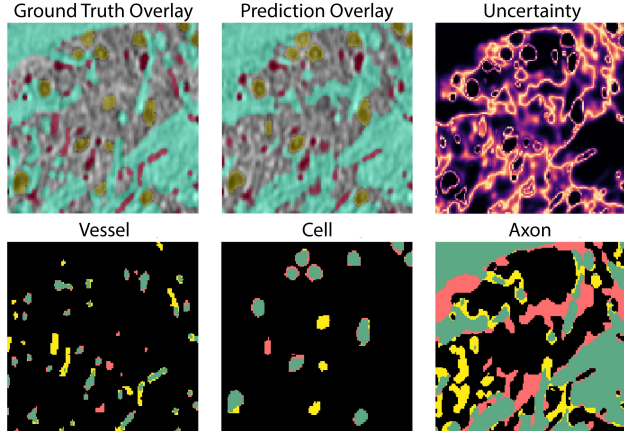

Fig. 2. *Visualization of predicted microstructural segmentations.* Top: From left to right, we show examples of i) Annotated ground truth of microstructural labels, ii) Microstructural segmentation results produced by our model, iii) Uncertainty of the model's predictions as quantified by the difference between the top two predicted probabilities for every pixel; Bottom: From left to right, per pixel true positives (green), false positives (red), and false negatives (yellow) in the predictions for blood vessels, cells, and axons.

are small clusters throughout the output volume that are misclassified. We therefore enforce local spatial consensus in 3D amongst macroscale predictions by extending a 2D k-Nearest Neighbours (kNN) based cleanup procedure [15] (Methods, Algorithm 1) and apply it to the predicted brain-area segmentations of our model. The working of the algorithm is as follows: Assuming that the ith ROI has m_i connected components, we only maintain the set of points in the m_i largest components predicted with the label i, which we call our "seeds". Once we have the set of seeds for all our ROIs, we relabel *all* points which are not seeds using the kNN algorithm. The seeds act as our labels, while the features used are the 3D position vectors of the points. We find that the algorithm correctly remaps many small previously incorrectly classified clusters to the right ROIs, and boosts the overall macroscale segmentation accuracy appreciably, especially in ZI.

Large-scale 3D reconstructions using the Double U-Net. To create 3D reconstructions spanning large volumes, we developed a pipeline that allows us to deploy our models at scale by taking as input any arbitrary 3D volume. As a first step, the pipeline divides the input into 128x128 blocks that are compatible with the Double U-Net model, and feeds them to both encoders simultaneously. We also fix the "field size", which is a hyper-parameter that specifies the stride of the blocking operation and gives us the ability to trade-off between the speed and quality of macroscale segmentation of the final 3D-reconstructed output. With a field size of 128, the input volume is divided such that there is no overlap between the 128x128 images extracted from the 3D input volume, thus speeding up the overall algorithm. However, this also implies
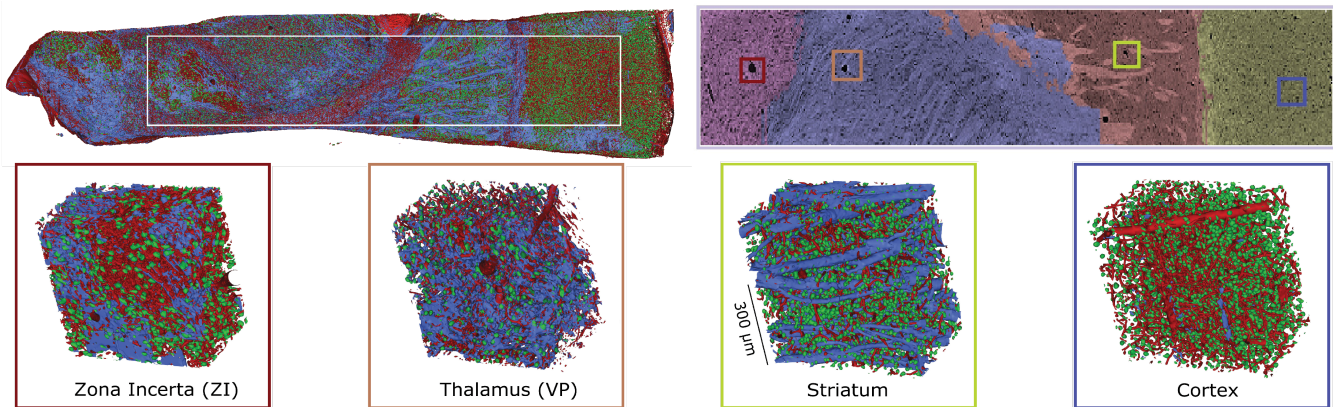
Fig. 3. *Visualization of pixel-level semantic segmentations and brain area predictions at scale.* Top: Left - Results for large-scale microstructural segmentation of blood vessels (red), cells (green), and myelinated axons (blue). Right - Results of macrostructural segmentation (post clean-up) on the area highlighted by the rectangle in the slice to the left. Bottom: Zoomed in versions of microscale segmentations of the cuboids across the different ROIs, highlighted in macrostructural panel.

that each pixel in a 128x128 patch is given the exact same class label, thus leading to very low-resolution macroscale segmentation. On the other hand, a smaller field size increases the resolution of the classification output at the expense of added computation. Results of our model deployed at scale (field size = 8) are shown in Fig. 3 and discussed in Sec. 3.

3. RESULTS & DISCUSSION

Segmentation across scales. We compared our Double U-Net model and training methodology with both single and multi-task architectures (Table 1). The single task baselines used were the U-Net encoder (which acts as a discriminative CNN) in the case of macroscale segmentation (i.e., brain-area classification) and the SegNet [24], U-Net in the case of microscale segmentation (i.e., pixel-wise classification of neural structure). The multi-task architecture compared against (MT U-Net) was a fully convolutional network constructed and trained similarly to the Y-Net [25], a hard-parameter sharing multi-task deep network previously introduced in the context of breast cancer diagnosis. Our methods outperformed the MT U-Net and were competitive with single-task baselines across both tasks and across all classes.

On the microstructural front, we found that our model provides very high accuracy in foreground segmentation, with average prediction of foreground pixels being $\sim 92\%$. The hardest components to segment accurately were cells and blood vessels, which are found in lesser proportions than myelinated axons and background tissue, and also share a lot of similarities with one another in terms of morphology, thus making it hard for the model to distinguish from one another (see Fig. 2). In a similar vein, on the macrostructural front, we found that ZI was the hardest class to learn, which we hypothesize is due to the fact that the region shares a lot of similarity with both the VP and CTX, thus making it an easy ROI for a deep network to misclassify.

3D Reconstructions and testing at scale. We deployed our model at scale on a multi-area sample spanning all ROIs (Fig. 3 Left & Middle; Sample volume = $3500 \times 520 \times 20 = 36.4M$ voxels) and to contiguous 3D volumes drawn from each of the ROIs (Fig. 3 Right; Sample volume = $257 \times 257 \times 361 \approx 23.84M$ voxels). We found that we could accurately reconstruct the microstructure, and in the case of the multi-area sample, with the help of the kNN based clean-up, produced highly accurate macrostructural reconstructions as well. The highest improvement post clean-up was in ZI (accuracy improved from 0.67 to 0.80), boosting our overall accuracy from 0.88 to 0.93. We demonstrate that the proposed multi-task approach can be used to effectively pull out ROIs and segment 3D data in large heterogeneous multi-region brain samples.

4. CONCLUSION & FUTURE WORK

In this work we developed a multi-task deep learning approach to model neural micro- and macrostructure simultaneously. Solving this task successfully requires a holistic understanding of the data, i.e., a good model of the data's structural features at different scales. We empirically establish our approach by testing it out on a large 3D X-ray dataset that spans many diverse brain regions. By providing an example of how techniques in multi-task representation learning can be leveraged to solve a widely pervasive problem, we are hopeful that this study will further inspire exciting research at the intersection of comparative neuroanatomy and machine learning. Ideas from active, semi-supervised, and self-supervised representation learning could lead to even better models of neural architecture across neurological states.

5. ACKNOWLEDGMENTS

6. REFERENCES

[1] Philbert Tsai et al., "Correlations of neuronal and microvascular densities in murine cortex revealed by direct counting and colocalization of nuclei and vessels," *Journal of Neuroscience*, vol. 29, no. 46, pp. 14553–14570, 2009.

[2] Jeff Lichtman and Winfried Denk, "The big and the small: challenges of imaging the brain's circuits," *Science*, vol. 334, no. 6056, pp. 618–623, 2011.

[3] Eva Dyer et al., "Quantifying mesoscale neuroanatomy using X-ray microtomography," *Eneuro*, vol. 4, no. 5, 2017.

[4] Anan Li et al., "Micro-optical sectioning tomography to obtain a high-resolution atlas of the mouse brain," *Science*, vol. 330, no. 6009, pp. 1404–1408, 2010.

[5] Hui Gong et al., "Continuously tracing brain-wide long-distance axonal projections in mice at a one-micron voxel resolution," *Neuroimage*, vol. 74, pp. 87–98, 2013.

[6] Michael Economo et al., "A platform for brain-wide imaging and reconstruction of individual neurons," *Elife*, vol. 5, pp. e10566, 2016.

[7] Kyle Milligan et al., "Brain mapping at high resolutions: Challenges and opportunities," *Current Opinion in Biomedical Engineering*, vol. 12, pp. 126–131, 2019.

[8] Maxime Lafarge et al., "Domain-adversarial neural networks to address the appearance variability of histopathology images," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, pp. 83–91. Springer, 2017.

[9] Daniel Tward et al., "Diffeomorphic registration with intensity transformation and missing data: Application to 3d digital pathology of alzheimer's disease," *Frontiers in Neuroscience*, vol. 14, 2020.

[10] Verena Kaynig et al., "Large-scale automatic reconstruction of neuronal processes from electron microscopy images," *Medical Image Analysis*, vol. 22, no. 1, pp. 77–88, 2015.

[11] Jan Funke et al., "Large scale image segmentation with structured loss based deep learning for connectome reconstruction," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 7, pp. 1669–1680, 2018.

[12] Alexander de Brebisson et al., "Deep neural networks for anatomical brain segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2015, pp. 20–28.

[13] Asim Iqbal et al., "Developing a brain atlas through deep learning," *Nature Machine Intelligence*, vol. 1, no. 6, pp. 277, 2019.

[14] Aishwarya Balwani and Eva Dyer, "Modeling variability in brain architecture with deep feature learning," in *2019 53rd Asilomar Conference on Signals, Systems, and Computers*. IEEE, 2019, pp. 1186–1191.

[15] Aishwarya Balwani and Eva Dyer, "A deep feature learning approach for mapping the brain's microarchitecture and organization," *BioRxiv*, 2020.

[16] Judy Prasad et al., "A three-dimensional thalamocortical dataset for characterizing brain heterogeneity," *Sci Data 7, 358*, 2020.

[17] Ariel Agmon and Barry Connors, "Thalamocortical responses of mouse somatosensory (barrel) cortex in vitro," *Neuroscience*, vol. 41, no. 2-3, pp. 365–379, 1991.

[18] Judy Prasad et al., "A three-dimensional thalamocortical dataset for characterizing brain heterogeneity: Region of Interest Annotations (Nrrd)," 2020.

[19] Judy Prasad et al., "A three-dimensional thalamocortical dataset for characterizing brain heterogeneity: Microstructure Annotations (NumPy)," 2020.

[20] Olaf Ronneberger et al., "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2015, pp. 234–241.

[21] Kaiming He et al., "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.

[22] Syuan-Ming Guo et al., "Revealing architectural order with quantitative label-free imaging and deep neural networks," *BioRxiv*, p. 631101, 2019.

[23] Ting Chen et al., "A simple framework for contrastive learning of visual representations," *arXiv preprint arXiv:2002.05709*, 2020.

[24] Vijay Badrinarayanan et al., "Segnet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 12, pp. 2481–2495, 2017.

[25] Sachin Mehta et al., "Y-net: joint segmentation and classification for diagnosis of breast biopsy images," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2018, pp. 893–901.