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

Joseph Miano[δ] *Aishwarya Balwani*[δ] *Ran Liu*[δ] *Lindsey Kitchell*[#]
Judy Prasad[†] *Erik Johnson*[#] *William Gray Roncal*[#] *Eva Dyer*[δ,σ,*]

[δ]Georgia Institute of Technology [#]Johns Hopkins University Applied Physics Laboratory
[†]University of North Carolina at Chapel Hill [σ]Emory University

ABSTRACT

Methods for resolving the brain's microstructure are rapidly approaching capabilities that would allow imaging whole-brain volumes at high resolutions, making it increasingly more commonplace to interrogate the structure of samples spanning multiple diversified brain regions. In such multi-area samples, segmentation of *microstructure* and subsequent modeling and analyses of the *macrostructure* prove to be especially challenging due to the heterogeneity of neural components both within and across different regions of interest. In this work we introduce a deep learning approach that combines aspects of multi-task learning with the commonly used U-Net architecture for segmentation of brain structure at *multiple scales*. In particular, we apply our methods to a recently introduced three-dimensional X-ray micro-CT imaging dataset that exhibits heterogeneity in neural structure across a number of brain regions in the mouse. We show that it is possible to have a scalable, accurate, unified framework that jointly learns to model the micro and macro architecture of the brain, using both structural and contextual information available in the sample.

***Index Terms*—** Multi-scale, microstructure, macrostructure, segmentation, multi-task learning.

1. INTRODUCTION

Studies of brain structure and function often require us to map out and analyze the anatomy of the brain across various regions of interest (ROIs) [1, 2, 3]. As a consequence, imaging methods in neuroscience have been steadily moving towards pipelines capable of rendering multi-area and even whole-brain volumes [4, 5, 6]. In addition to having to deal with huge amounts of data that necessitate ample compute and automation for their study, the analyses of large-scale, diversified neural datasets bring with them a host of other problems including but not limited to imaging and physical artifacts, as well as limited and partial views of the data [7, 8, 9]. Furthermore, the heterogeneity in distributions of underlying neuronal components (e.g., cells, blood vessels, neural processes like myelinated axons and dendrites) both within and across brain areas makes the characterization of their structure, as well as that of the different brain regions they constitute incredibly challenging, consequently hindering much progress across the field of comparative neuroanatomy.

Most current methods for brain mapping have primarily focused on single scale segmentation and analyses, where the goal is to segment either high-resolution images of relatively small brain samples into microstructural components such as neurons and/or neurites and reconstruct the connectome of the imaged sample [10, 11], or relatively low-resolution images of large multi-area samples or entire brains into various pre-defined regions [12, 13]. However, the practice of conducting these two tasks independently is highly inefficient in terms of the time spent creating, training, and deploying such models. Moreover, very often the two types of segmentations also serve as prerequisites for one another, where the end goal of *microstructural* segmentation often is to resolve various neural components so as to aid in the characterization of the topology of different (macro) brain areas, and that of *macrostructural* segmentation is to resolve ROIs by identifying sections that share distributional properties in terms of their (micro) structural compositions to form models of the brain 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 strategy for modeling neural architecture across multiple spatial scales using a multi-task learning based framework. Specifically, we develop a model of the *macrostructure* by learning to discriminate (i.e., classify) between different brain areas in the sample [14, 15] and extend these representation learning methods to also explicitly model the *microstructure* by learning to semantically segment different neural components present in our data. We combine the two tasks into a single architecture via soft-parameter sharing and train end-to-end following a two-step procedure, thus producing a parameterized framework that can perform multi-scale modeling and segmentation of neural data in a data-driven and generalizable manner.

We empirically show that our proposed multi-task architecture consistently performs competitively when compared to strong single-task baselines at both scales. Our work also

*Corresponding author. Email: evadyer@gatech.edu

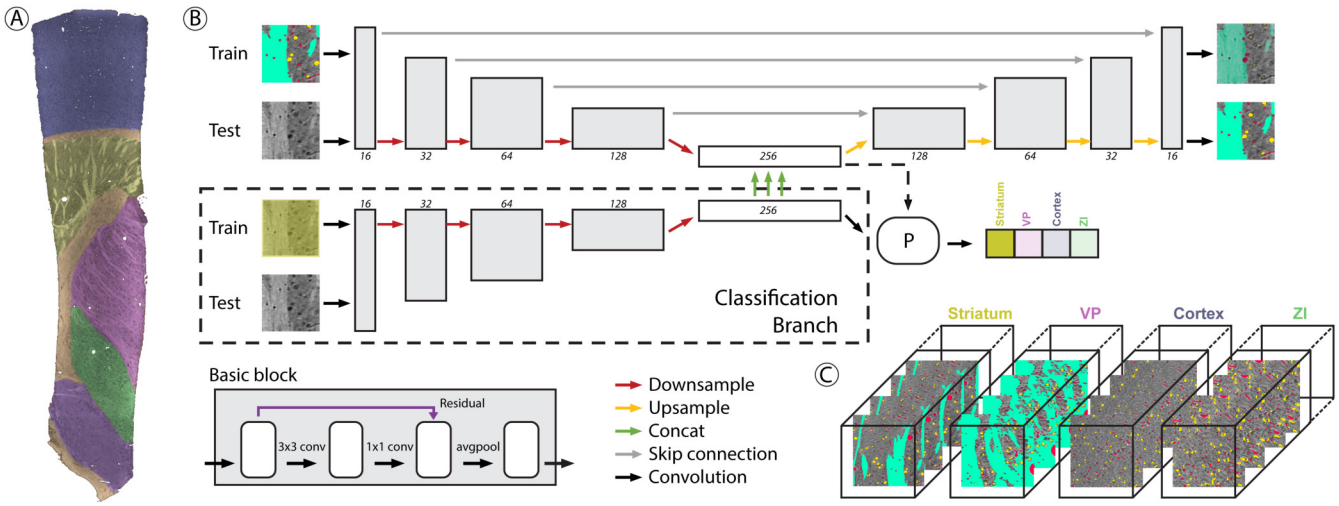

Fig. 1. *Overview of multi-task architecture for modeling brain architecture across spatial scales.* (A) shows a 2D slice of the data with macroscale annotations (different colours represent different ROIs). (B) depicts the Double U-Net architecture; Top = Segmentation half of the network. Bottom (dotted region) = Classification branch. The high-level, multiscale latent representation shared by the two halves is shown in the middle (before and after the green arrows). To the bottom left, we show the basic block of our U-Net, which consists of a 3x3 convolutional filter, followed by a 1x1 convolutional filter and average pooling. We further include a residual connection between the input and output of the 1x1 filter. The projection head following the classification branch is denoted by the P-block. (C) shows examples of microscale annotations across different ROIs.

addresses modifications needed to allow for large-scale deployment of our model, and we provide results of the same to establish our methods' generalizability and robustness to neural heterogeneity across the three-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 the task significantly.

2. METHODS

Dataset and annotations. To build a good model of the neuroanatomy at multiple scales, we work with a newly introduced, publicly available three-dimensional X-ray microCT dataset [16] spanning 5.9 gigavoxels (5805x1420x720 voxels) imaged at 1.56um isotropic resolution. The dataset studies the Agmon-Connors slice [17], a well-known thalamocortical sample encompassing a range of ROIs from the somatosensory cortex to the ventral posterior thalamic nucleus in the mouse brain, and is ideal for our purposes of studying and quantifying multi-scale heterogeneity of neural structure across diverse brain areas. The dataset also provides us with sufficiently diverse and high-quality annotations of the micro and macrostructure; At the *macroscale* [18], the annotations 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 (∼78um) slices along the depth of the imaged dataset (Fig. 1A). Those at the *microscale* [19] identify myelinated axons, blood vessels (BV), and cell bodies densely labelled in blocks of 256x256 pixels across eleven evenly spaced (∼46um) slices along the depth of the imaged sample (Fig. 1C) in the CTX, STR, VP, and ZI.

For the purposes of jointly modeling the neural structure across multiple spatial scales, we work within the regions 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, 64 images for training, validation, testing respectively. At the macrostructural level we once again sample 128x128 images, restricting ourselves to only those which are contained strictly within the four ROIs, across all slices which have area-level annotations. The final macroscale train, validation and test datasets have 1109, 197, and 165 images respectively across the four ROIs.

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. in two key ways. First, we bolster the core U-Net architecture comprised of a fully convolutional encoder and decoder (Fig. 1B) with residual connections [20], thereby allowing better propagation of signal and gradients throughout the network and helping it achieve high levels of accuracy [21]. Next, we add a second encoder to the core U-Net, and use it as our *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 different brain areas, therefore requiring the use of two separate encoders for the different tasks. However, since the tasks are learnt on the same data and both 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 concatenating the latent vectors of the classification branch with that of the core Double U-Net encoder. Finally, in addition to replicating the architecture of the U-Net encoder, we also append an additional *projection head* to the latent representations learnt by the classification branch to encourage better representations that achieve higher accuracies on downstream tasks, in accordance with the current representation learning literature [22].

Training procedure for multi-task learning. The model is trained end-to-end in two steps. We first 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 (brain area predicted by the model). The second term measures the mean squared error (MSE) between $\mathbf{x}$ and the reconstructed image g_r produced by the core 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 $g_r(\mathbf{x})$, i.e., the core branch of Double U-Net).

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	0.91	0.54	0.63	0.66	0.69
U-Net	**0.93**	**0.60**	0.70	0.78	0.75
MT U-Net	0.79	0.51	0.68	**0.91**	0.72
Double U-Net	0.92	0.59	**0.70**	0.79	**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 gets most of its macroscale predictions right, we find that there often are small clusters throughout the output volume that are misclassified. We therefore enforce local spatial consensus in 3D amongst macroscale predictions by extending the 2D k-Nearest Neighbours (kNN) based cleanup procedure in our previous work [15] (Methods – Algorithm 1) and apply it to the predicted brain-area segmentation outputs of our model. The idea behind the algorithm is simple – assuming that each ROI is expected to be one large connected component (as is the case with most brain areas), we only keep the labels of the points in the *largest* component predicted as belonging to a particular ROI, and re-label all the points which don't fall in this category with the label of the majority of their k spatially nearest neighbours having labels. We find that the algorithm correctly remaps many small incorrectly classified clusters to the right brain areas and boosts the overall macroscale segmentation accuracy significantly. Quantitative results for the application of the clean-up algorithm are reported in Sec. 3.

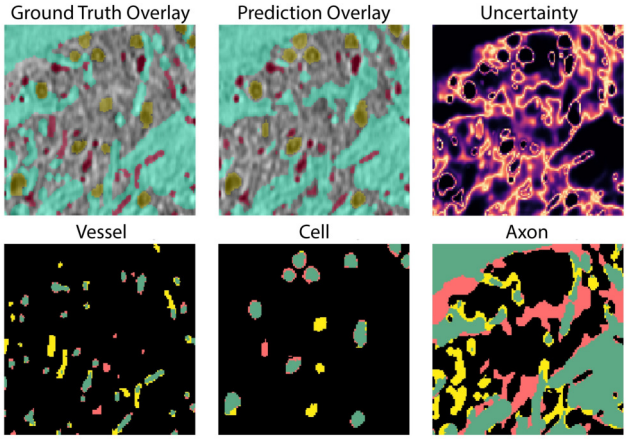

Fig. 2. *Visualization of predicted microstructural segmentations.* Top: From left to right, we show examples of 1) Annotated ground truth of microstructural labels, 2) Microstructural segmentation results produced by the Double U-Net, 3) Uncertainty of the model's predictions as quantified by the difference between the top two predicted probabilities for every pixel, Bottom: Per pixel true positives (green), false positives (red), and false negative (yellow) in the predictions for blood vessels, cells, and axons, respectively.

Large-scale 3D reconstructions using the Double U-Net. To be able 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. In this step, we make use of a "field size" 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

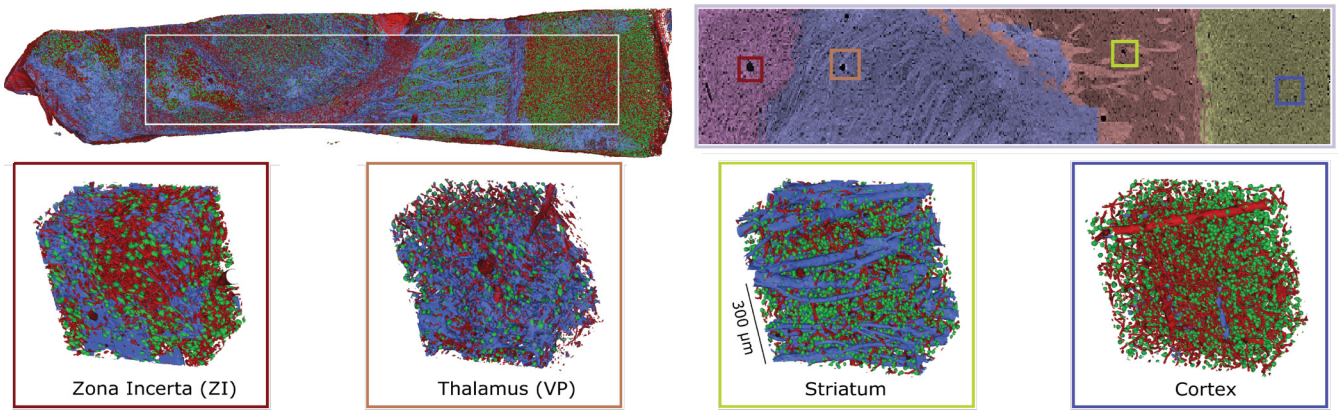

Fig. 3. *Visualization of pixel-level semantic segmentations and brain area predictions at scale.* Top: Left - Results for large-scale microstructural segmentation (blood vessels, cells, and myelinated axons in red, green, and blue respectively). Right - Results of macrostructural segmentation (post 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.

overlap between the 128x128 images extracted from the 3D input volume, thus speeding up the overall algorithm. However, this also implies that each pixel in a 128x128 patch is given the same class label, thus leading to very low-resolution macroscale segmentation (i.e., brain-area classification). On the other hand, a smaller field size increases the resolution of the classificatoutput at the expense of added computation. Results of our model deployed at scale (field size = 8) are shown in Fig. 3 and further 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 [23], 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 [24], a hard-parameter sharing multi-task deep network previously introduced in the context of breast cancer diagnosis. Our methods clearly outperformed the MT U-Net and were extremely 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. These issues are further evidenced by the results in Fig. 2, where we see entire cell bodies being misclassified as vessels and vice versa. 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 its topological features with both, the VP and CTX, thus making it an easy class for a deep network to misclassify.

3D Reconstructions and testing at scale. We deployed our model at scale i) *spatially* (i.e., in the XY plane) on a multi-area sample spanning all our ROIs (Fig. 3 Left & Middle; Sample volume = $3500 \times 520 \times 20 = 36.4$M voxels) and ii) *depth-wise* (i.e., along Z) on 3D volumes drawn from each of the ROIs (Fig. 3 Right; Sample volume = $257 \times 257 \times 361 \approx 23.84$M voxels) to produce multi-scale reconstructions. 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, where we went from an accuracy of 0.67 to 0.80, boosting our overall accuracy from 0.88 to 0.93.

4. CONCLUSION & FUTURE WORK

In this paper, we developed a multi-task learning approach to model and segment neural micro and macrostructure simultaneously. We applied our methods to a new, mesoscale X-ray micro-tomography dataset and empirically establish their ability to learn good models of neural structure at multiple scales. We consequently achieve high accuracies on tasks requiring a holistic understanding of the structural features of the data using a limited number of training samples. We are hopeful that this study will further inspire exciting research at the intersection of neuroimaging, comparative neuroanatomy, and machine learning, especially in the directions of active, semi-supervised, and unsupervised learning representation learning across various neurological states.

5. ACKNOWLEDGMENTS

E.L.D., A.H.B., and R.L. were supported by a generous gift from the Sloan Foundation, in addition to award NSF IIS-1755871 which supported J.M. as well. E.L.D., E.C.J., L.K., and W.G.R. were supported by NIMH grant R24MH114799.

6. REFERENCES

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

[2] 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.

[3] 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.

[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 and Giovanni Montana, "Deep neural networks for anatomical brain segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition 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] Kaiming He et al., "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

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

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

[23] Vijay Badrinarayanan, Alex Kendall, and Roberto Cipolla, "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.

[24] 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.