Multi-task Learning for Neural Image Classification and Segmentation Using a 3D/2D Contextual U-Net Model

---

A Thesis
Presented to
The Academic Faculty
By

Joseph D. Miano

In Partial Fulfillment of the Requirements
for the Bachelor of Science in Computer Science Degree in the
College of Computing

Georgia Institute of Technology
May 2020

Faculty Mentor: Dr. Eva Dyer

Faculty Second Reader: Dr. Constantine Dovrolis

Contents

1 Abstract

We present a 3D/2D Contextual U-Net model and apply it to segment and classify samples from a heterogeneous mouse brain dataset obtained via X-ray microtomography, which spans 4 distinct brain areas: Striatum, Ventral Posterior Thalamic Nucleus (VP), Cortex, and Zona Incerta (ZI). Our multi-task model takes in a 3D volume and outputs both a 2D segmentation of the central plane and the volume's brain area class, which can then be used to generate 3D reconstructions across samples with heterogeneous microstructure distributions. We investigate various properties of the model, including its quantitative segmentation and classification performance across the 4 brain regions, qualitative performance via the generation of 3D reconstructions, and interpretability via an investigation of the network's latent representations. Because our model performs both classification and segmentation, we also investigate how changing their relative weight during training, via a parameter we call λ, affects performance and latent representations. Quantitative and qualitative results demonstrate that our model achieves reasonable segmentation and classification performance and can be scaled to large, heterogeneous brain regions. This technique could be used by neuroscientists seeking to automate the creation of multi-scale brain maps that incorporate both microstructure and brain area information.

2 Introduction

Typical supervised machine learning models are task-specific, meaning they are trained to perform a narrowly defined task by learning from a specific dataset (e.g. automatic classification of a tumor as benign or malignant based on a medical scan). To perform these tasks accurately, large quantities of labeled training data are often required, which may be difficult or costly to collect (e.g., in biomedical imaging settings [1]). Multi-Task Learning (MTL) refers to training machine learning algorithms to perform multiple related tasks, which can help improve the performance across all the tasks [2]. MTL has the advantage of enabling good performance on tasks where training data are limited by training towards similar tasks, which can each benefit the learning of the others. Multi-Task Learning can therefore potentially provide value in biomedical imaging settings, including brain imaging and mapping.

Current computational approaches for high-resolution mapping and segmentation of neural structures, which have been established extensively in the electron microscopy community, have thus far focused on relatively small samples from a single area [3]. Neural segmentation efforts for larger, more diverse volumes, on the other hand, have typically focused on brain area classification rather than microstructure segmentation [4, 5]. Thus, there is a growing need to develop methods for analyzing neuroanatomy that can work flexibly across different brain areas, at both micro and macro scales, and draw meaningful comparisons across different large scale samples.

In this study, we develop and examine the properties of a new MTL convolutional neural network architecture to perform semantic segmentation and image classification of a thalamocortical mouse brain X-ray dataset [6, 7]. Semantic segmentation refers to assigning a class label to each pixel in an image, while image classification refers to assigning a class label to the overall image itself. The architecture we developed to address these joint tasks is a variation of the U-Net, which was originally developed for biomedical image segmentation [8].

The dataset [6] used for this study is a thalamocortical mouse brain volume obtained via X-ray microtomography, which spans from hypothalamus to cortex. X-ray microtomography is a technique that allows micron-resolution imaging of millimeter-scale tissue samples without requiring tissue sectioning [9]. Because of the diversity in microstructure distribution between the 4 brain areas, we hypothesized that a network's ability to classify brain regions could help inform its microstructure segmentation, thus improving final performance and network interpretability. The specific brain regions under consideration are Striatum, Ventral Posterior Thalamic Nucleus (VP), Cortex, and Zona Incerta (ZI). Within each of these 4 brain regions, our segmentation task consists of labeling pixels for 4 classes: background, blood vessels, cells, and myelinated axons. The classification task consists of determining which brain region a given input image is from.

Our main contribution with this study is the development and analysis of an MTL architecture for biomedical classification and segmentation, which we call a 3D/2D Contextual U-Net, that can achieve good performance on relatively little training data and can be applied by researchers to automate multi-scale neural analysis. We investigate the network's classification and segmentation performance across our 4 brain areas of interest. We also investigate how placing more or less relative weight on classification vs. segmentation, using a user-tunable parameter we call λ, affects the performance of both. Furthermore, we investigate the interpretability of networks trained with different λ values by visualizing correlation matrices of their latent representations, which consist of intermediate data output by the bottom layer of the encoding part of the network. Interpretability is especially important in the biomedical domain because patients and providers need a "why" to support potentially life-changing diagnoses.

3 Literature Review

Deep learning refers to the use of layered computational models, often called artificial neural networks, that can learn the data representations required for a given task [10]. By passing data through several layers of linear and non-linear transformations, neural networks can perform a multitude of tasks, such as speech recognition [11], image recognition [12], natural language processing [13], among others. Unlike traditional machine learning techniques, which require humans to define and calculate relevant features from data to train models, neural networks typically operate directly on raw data and can learn the features most useful for performing a specified task [10]. In our work, we implement supervised learning via a neural network architecture based on the U-Net [8], which we train using microstructure-annotated and brain-area-annotated data. Traditional machine learning techniques, which do not leverage deep neural networks, remain prevalent in biomedical imaging analysis work (e.g. with a tool like Ilastik [14] where manual feature selection is needed). Though such tools can be effective, they are limited by the domain knowledge of the user, who must specify which features to include in the machine learning model. By leveraging a deep learning approach, we aim to develop a process that can be ported across many domains and function well without the need for manual feature engineering or selection.

Supervised learning, one of the major categories of machine learning, refers to training a model to perform a target task via labeled training examples. In many supervised learning problems, there is a single task that the network should perform. In order to extend beyond a single task per model, Multi-task Learning (MTL) was popularized in the 1990s and characterized as an approach to help improve model generalization performance by training on related tasks in addition to the target task [2]. Deep neural networks have been used extensively for MTL over the past several years [15]. In neural networks, MTL can be accomplished by adding branches towards different outputs, where each output corresponds to a different task. Thus, a neural network learns a shared representation between tasks, which can help improve performance across all tasks [2]. One way to accomplish MTL using a neural network is via hard parameter sharing [15], which refers to learning joint network weights for all the tasks and branching the network after some number of layers in order to differentiate tasks. We leverage hard parameter sharing to extend the capabilities of a traditional U-Net in order to enable a single network to perform both classification and segmentation.

The U-Net is an example of a fully convolutional neural network well-suited to image segmentation where little data is available [8]. A fully convolutional network, unlike networks typically used for image classification, does not contain fully connected layers; this means the network has relatively fewer weights to learn and can thus be trained and tested more quickly. The original U-Net takes in a 2D image and outputs a 2D segmentation, which means that each pixel in the input image is assigned a class label. The U-Net builds on the work done by Long et al. [16] on fully convolutional networks for semantic segmentation by introducing an ascending, upsampling path for the data to traverse after the downsampling path. Furthermore, the U-Net connects the downsampling and upsampling paths via "copy and crop" connections, which allow localization by combining high-resolution features from the downsampling path to be incorporated into the upsampling path. This basic architecture is useful for our task because it can distinguish between similar brain microstructures by leveraging broader image context (e.g. macroscale shapes) and higher-resolution localization (e.g. shape edge).

To extend the capabilities of the original U-Net in order to fit our use-case, we leverage a branched 3D/2D architecture that can perform both image classification and image segmentation. Our 3D/2D U-Net architecture is based on the work done by Guo et al. [17]; the network takes in a 3D image as input and outputs a 2D segmentation of the central input image plane. The advantage of such a construction is two-fold. First, we gain the advantage of 3D context around the input image, which can be helpful in discriminating between structures that look similar in a single 2D plane (e.g. cells and blood vessels). Second, the training data does not need to be fully annotated in 3D—individual 2D planes can be annotated by a human, which means our training data can span further within each brain region, capturing diversity while minimizing time-consuming human annotation work. We extend this 3D/2D U-Net architecture, which itself can perform only segmentation, by adding a branch with classification capabilities by using hard parameter sharing [15].

With our MTL architecture, we aim to enable automatic classification and segmentation of heterogeneous neural datasets. Our technique draws upon the traditional U-Net [8] and 2.5D U-Net [17], extending their functionality via a contextual classification branch, thus enabling a single network to perform classification and segmentation. Segmenting brain microstructure automatically is important because it enables the creation of large-scale 3D reconstructions, which can be useful for disease modeling and connectomics.

4 Methods and Materials

Our goal is to develop and apply a novel neural network architecture to a thalamocortical mouse brain X-ray microtomography dataset [6] to classify brain areas and segment microstructures. Several distinct brain regions are observable within our dataset, each containing visible microstructure (like cells, blood vessels, and axons). We implemented what we call a 3D/2D Contextual U-Net, which is a neural network architecture that can perform both image classification and image

segmentation. The 3D/2D in the name refers to the fact that the network takes in a 3D input volume and outputs a 2D segmentation; the U-Net is contextual because it outputs both microstructure segmentations and macroscale brain area context predictions (i.e., which brain area a given image is sampled from). All the code for this project is written in Python and using PyTorch, which is an open-source deep learning library for Python. We leverage Nvidia GPUs to accelerate data processing and model training/evaluation. The code is organized as a mix of Jupyter notebooks and separate Python functions.

Ground truthing and Pixel-level Annotations. We manually created ground truth pixel-level annotations of 4 major brain areas [18]: Cortex, ZI, Striatum, and VP. We used ITK-SNAP, which is a tool that facilitates 3-dimensional semantic segmentation of biomedical images [19], for this task. To standardize the ground truth segmentations, we extracted 4 volumes of size (x:y:z) = (257:257:361), one for each of the 4 brain areas. The coordinates of each of these volumes are as follows, formatted as (xstart_xend, ystart_yend, zstart_zend): Cortex (4600_4857, 900_1157, 110_471), ZI (1543_1800, 650_907, 110_471), Striatum (3700_3957, 500_757, 110_470), VP (3063_3320, 850_1107, 110_471). Within each of these (257:257:361) brain area volumes, we densely annotated (starting at index z = 0) slice z = 30, 60, 90, 120, 150, 180, 210, 240, 270, 300, 330. Because each annotated z slice has an area of (x:y) = (257:257), we extract 4 (128:128) images from each z slice by omitting the bottom row of pixels and leftmost column of pixels; these (128:128) images are the annotated training and validation data used by the 3D/2D Contextual U-Net. The data were split into training and validation as follows: slices z = 30 through 240 inclusive are training data and z = 270 through 300 inclusive are validation data. Because 4 samples are extracted per z slice, this means for each of the 4 brain areas we have 32 training samples and 8 validation samples. When training the overall U-Net on data from all 4 brain areas, we therefore have 128 training samples and 32 validation samples.

3D/2D Contextual U-Net Architecture. Our 3D/2D Contextual U-Net is composed of 2 main sections—an encoding section where the input image is downsampled via convolution and pooling operations and a decoding section where the low-dimensional latent representation is upsampled, mainly via transpose convolutions. A unique aspect of the 3D/2D Contextual U-Net is that it takes as input a 3D image and yet outputs a 2D segmentation. Thus, the convolution operations done on the encoding path of the U-Net are 3D. We project the latent representation from 3D into 2D at the bottom of the encoding section of the U-Net and apply 2D convolution operations in the decoding path, which leads to the 2D segmentation output. Furthermore, we add bridge connections between the encoding and decoding paths. Unlike the original 2D U-Net, these bridge connections project the data from 3D into 2D so that it can be meaningfully concatenated with the data in the decoding path. In addition to these encoding and decoding paths, we add what we call a contextual branch at the bottom of the encoding path. Thus, there is a fork in the network at the bottom of the encoding path where it splits into the decoding segmentation path and the contextual branch. The contextual branch consists of further convolutional layers that terminate with a classification label for the overall input image, while the decoding path produces a pixel-wise segmentation.

Model Training Procedure. Before we can begin network training, we must first pre-process the data from 4 different brain area volumes: Striatum, Ventral Posterior Nucleus (VP) of the Thalamus, Cortex, and Zona Incerta (ZI). These 4 brain volumes are pulled from a larger thalamocortical dataset [6] that was generated via X-ray microtomography at the Advanced Photon Source in Argonne National Laboratory. The 3D volume data for each brain area must be first processed into the individual 128x128x5 raw images and corresponding human-annotations to train the network. Because we are using a 3D/2D Contextual U-Net, there are 5 planes of depth to each raw input image, where the central plane corresponds to the human annotation. The purpose of this is to allow the network to leverage 3D structural information from the input image, which is useful for distinguishing between similar microstructures like cells and blood vessels. Furthermore, we augment the training data with horizontal and vertical flips as well as 90, 180, and 270-degree rotations. Once the data are preprocessed, training can begin.

In order to train the U-Net, we specify a loss function and optimizer. The optimizer we use is Adam [20], which is a well-known stochastic optimization algorithm for neural networks. We use Adam because it has been shown to perform favorably during training when compared to other common optimizers like AdaGrad, Stochastic Gradient Descent (SGD), and RMSProp [20]. In order to account for pixel-level class imbalances (e.g. there are far more background pixels than vessel pixels), we reweight each class proportional to the number of instances of the class. For our loss function, we use 2 cross entropy loss terms, one for segmentation and one for classification. The classification cross entropy loss term has a scalar multiple (λ), which controls the relative weight of classification towards the overall loss. We train and test the performance of the U-Net for the following values of λ: 0, 10^{-9}, 10^{-8}, 10^{-7}, 10^{-6}, 10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 1, 10, and 100. For each of these values of λ, we tune the learning rate and batch size via random search. For each λ and

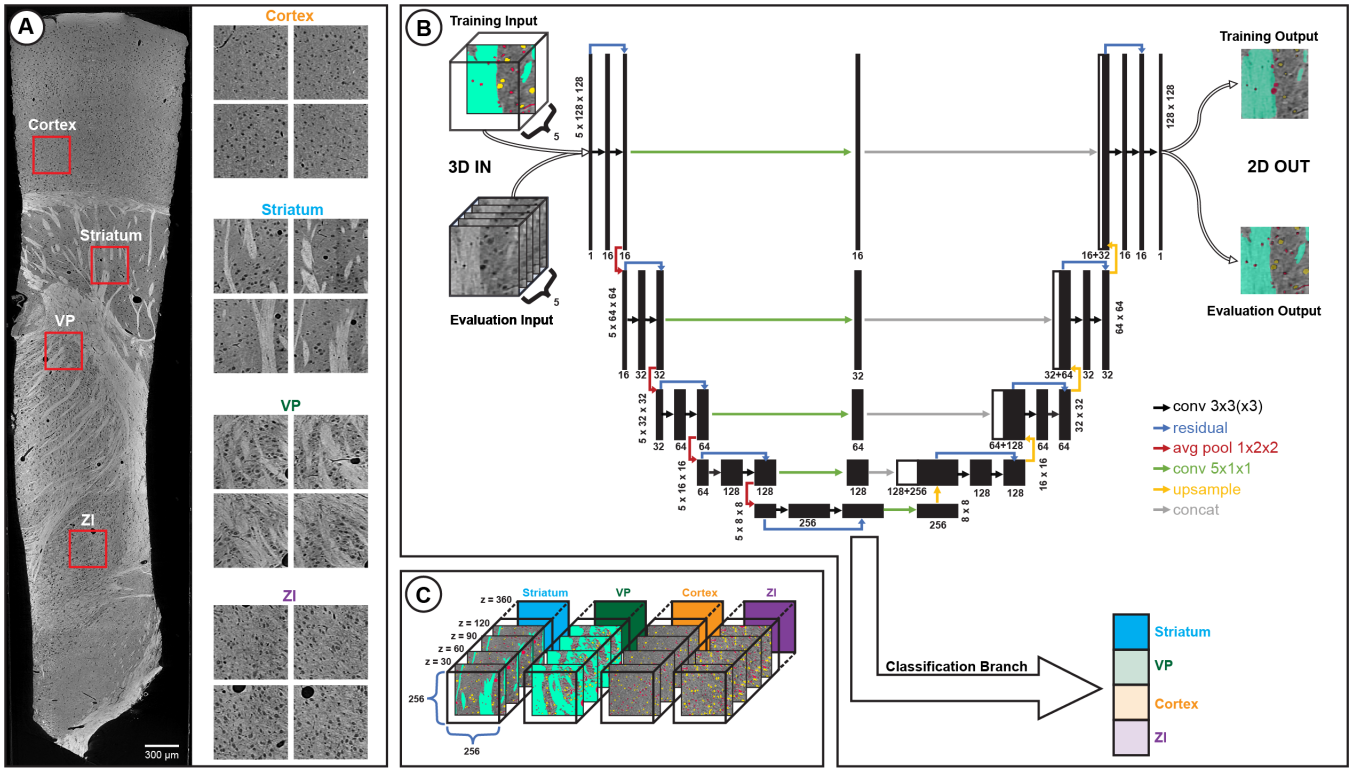

Figure 1: *Overview figure showing data and model architecture.* On the left in (A), we show a slice of the full thalamocortical dataset with to-scale cutouts of each volume used for training. On the right in (A), we show larger examples of 256x256 pixel slices from each brain region of the training dataset. In (B), we show the 3D/2D Contextual U-Net architecture, which accepts 3D (128x128x5) volumes as input and produces both classification and 2D segmentation (128x128) outputs. The central plane of each input volume is annotated in the training data. (C) shows the standardized annotation procedure we used to prepare the training data. Each brain area volume used for training data is 257x257x361. Annotated planes have a 30 pixel gap between them. Training data consists of planes z = 30, 60, 90,..., 240. Validation data consists of planes z = 270 and z = 300. Each annotated z plane is processed into four 128x128x5 volumes.

hyperparameter combination, the best model, according to the weighted segmentation F1 score, is chosen over 300 epochs of training (where each epoch consists of passing all the training data through the network once).

Performance Evaluation. The methodology to evaluate the model's performance consists of both quantitative and qualitative measurements. From a quantitative perspective, we compare the model's output on the validation dataset, which it did not see during training, with the human ground-truth annotations. The pixel-level differences between the model's outputs and the ground truth allow us to compute F1 scores across our various brain areas and classes. We measure both the classification and segmentation mean F1 scores for our 3D/2D Contextual U-Net model at varying values of λ. We also gain insight into how the input data are transformed by the network by visualizing the latent representation correlation matrices taken from its base. From a qualitative perspective, we process entire 3D image volumes to pass 128x128x5 chunks through the network, get segmentation outputs, and reconstruct the 3D segmentation volumes. The dense microstructure reconstructions generated by this process give us a qualitative view of how well the network is segmenting and serves as a use-case demonstration for this model.

5 Results

Evaluating our Best Performing Model. We evaluated the performance of our best (segmentation) performing U-Net model by examining qualitative and quantitative metrics. As shown in Figure 2(A), the model's output segmentation is visually very similar to the ground truth annotation; the model is mostly correct when discriminating between cell and blood vessel objects in the example shown, with errors mainly occurring around the edges of objects. It appears that the U-Net frequently fills in more of each object than the annotator who created these particular annotations; the pixels near the edges of these objects are usually the most ambiguous, and it is possible that different annotators would fill the same objects to differing degrees. The model uncertainty plot on the bottom row of Figure 2(A) exemplifies the uncertainty present when labeling pixels close to object edges.

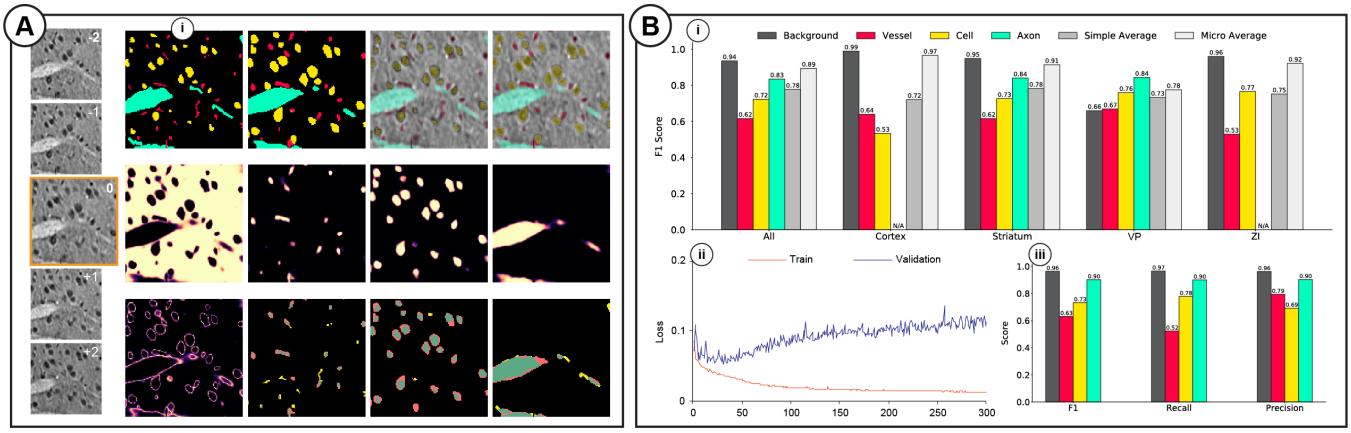

Figure 2: *Segmentation results showing an example model input and output and model performance.* On the left of (A), we show the 5 planes of an example input validation volume for Striatum, with the central plane (0) to be segmented. Row (i) shows, from left to right: ground truth annotation, model output, ground truth overlay, and model output overlay. Row (ii) shows model output probability maps for, from left to right: background, blood vessel, cell, and axon. Row (iii) shows, from left to right: model uncertainty, blood vessel accuracy, cell accuracy, and axon accuracy. We computed model uncertainty as the difference between the most probable and second most probable class output by the model for each pixel. Each accuracy plot is color coded to show green for true positives, yellow for false negatives, and red for false positives. In B(i) we show our model's F1 score achieved on the validation dataset for each brain area and output class. In B(ii), we show a training curve for our model over 300 epochs. B(iii) shows inter-rater reliability (IRR) between two instances of the same Striatum dataset for two annotators.

In order to quantitatively evaluate our model's performance, we examined F1 scores across each of the 4 brain areas, broken out by microstructure class (see Figure 2B(i)). The model performed best on axon (where applicable) and background labeling across the 4 brain areas, with an overall average F1 score of 83% for axons and 94% for background. Blood vessels were the most difficult microstructure to segment, with an overall average score of 62%. This lower performance is likely due to their small size and ambiguous appearance–it is often difficult to visually distinguish them from background.

Each 3D/2D Contextual U-Net model was trained using a weighted cross entropy loss combining segmentation and classification performance. In order to account for the microstructure class imbalances (e.g., there are far more pixels labeled as other or axon than as blood vessel), we weighted the cross entropy loss for each class as the reciprocal of the number of instances of that class. To control the relative weight of segmentation and classification loss in the overall loss function, the two are combined according to a user-specified parameter, λ. Thus, when this parameter is set to zero, the model will only aim to solve the semantic segmentation problem (baseline without multi-task). The training curves in 2B(ii)) demonstrate that our model fits the data relatively quickly; though we train for 300 epochs, the lowest validation loss is achieved around the 40^{th} epoch. The model is likely able to fit the dataset quickly due to its small size. In order to select the best model for each value of λ, we used a weighted average F1 score (with the same class balancing as in the loss) on the validation set to save the best-performing parameter weights.

We computed various inter-rater reliability (IRR) metrics (see Figure 2B(iii)), including F1, recall, and precision, in order to establish a baseline to which we can compare the model's performance. Two different human annotators annotated the same Striatum volume, and the slices where annotations overlapped were used for the IRR computation. In order to accelerate the process, the second annotator, whose data was not used to train or test the model, completed their annotations by using model outputs from a previous model over the full 3D volume as a starting point and correcting them for each slice. As shown by the IRR results, F1 scores for the Striatum IRR and for the U-Net model are very similar across all 4 microstructure categories, indicating that the model is achieving near human-level performance, given that many pixels are ambiguous even for human annotators. The F1 IRR scores were 96%, 63%, 73%, and 90% for background, blood vessels, cells, and axons, respectively; the F1 U-Net scores were 95%, 62%, 73%, and 84% for background, blood vessels, cells, and axons, respectively. However, we note that due to the ambiguity of many of these microstructures, it is possible that using the U-Net outputs as a baseline could have influenced the second human annotator's prior beliefs about what each pixel should be, thus causing a higher similarity between those annotations and the U-Net outputs. We investigated this question by treating the second set of human annotations as the ground truth and computing the best U-Net's F1 scores. The F1 scores for this side experiment were 96% for background, 57% for blood vessels, 74% for cells, and 92% for axons, which are similar to the numbers we reported for IRR. These results help increase our confidence that the results reported for IRR are an accurate representation of the difficulty of our segmentation task; this means that our best U-Net is indeed performing close to what is possible for humans, given the ambiguity and subjectivity involved when annotating certain microstructure pixels.

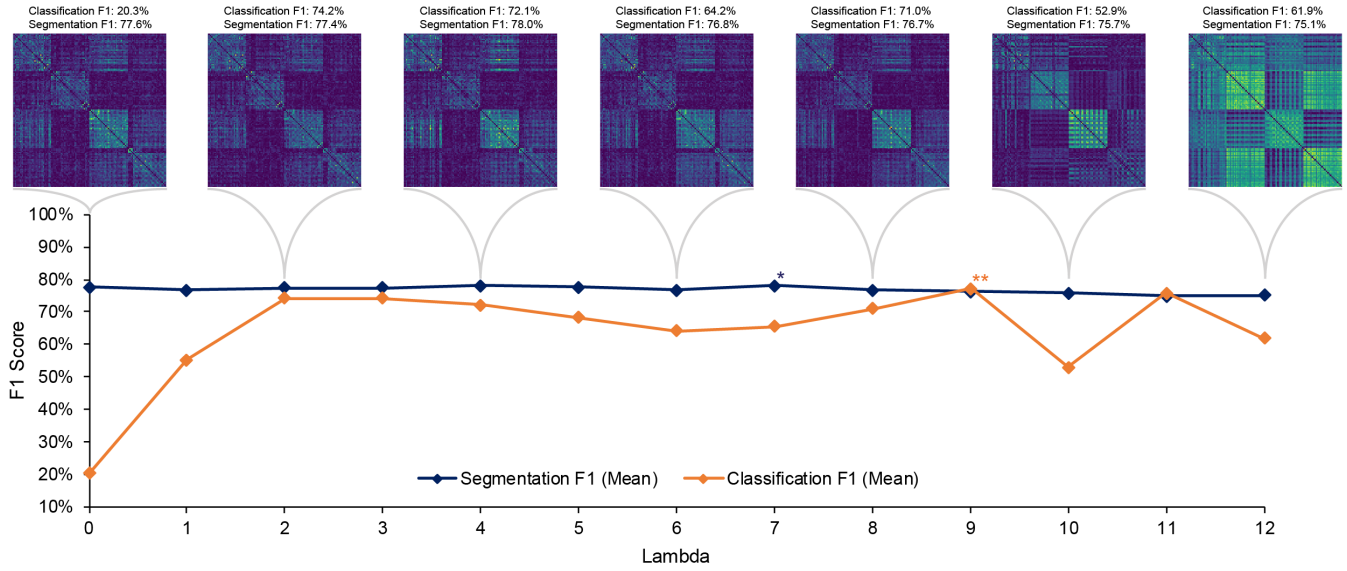

Figure 3: *Multi-task results for segmentation and classification with varying* λ. λ controls the relative ratio of classification to segmentation weight in the overall model loss during training. The λ values displayed on the x-axis are mapped using a log base 10 scale as follows: $0 \rightarrow 0$, $1 \rightarrow 10^{-9}$, $2 \rightarrow 10^{-8}$, $3 \rightarrow 10^{-7}$, ..., $9 \rightarrow 10^{-1}$, $10 \rightarrow 1$, $11 \rightarrow 10$, $12 \rightarrow 100$. Along the top of the figure, we show the correlation matrix (with zeroed diagonals) of the latent representations taken from the base of each U-Net trained at the corresponding λ value. The * corresponds to the highest overall mean segmentation F1, which occurs at $\lambda = 7$ and has a value of 78.1%. The ** corresponds to the highest overall mean classification F1, which occurs at $\lambda = 9$ and has a value of 77.1%.

Model Performance at Different Values of λ. The segmentation F1 score is a simple average of the F1 scores for each distinct microstructure class (background, cell, blood vessel, and myelinated axon) across all 4 standard brain volumes under consideration (Striatum, VP, Cortex, and ZI). To compute the classification F1 scores, we leveraged additional human-annotated brain area classification data [21] for the 4 brain areas under consideration. This additional data enabled a more robust evaluation of classification performance for each brain area; this is because each validation dataset used for segmentation is part of the same standard brain volume as the training set. While this does not pose too much of a problem for segmentation evaluation, it makes overfitting likely when evaluating classification performance, since there are fewer classification data points than segmentation data points.

As shown in Figure 3, we evaluated segmentation and classification performance for various values of λ. Our original hypothesis was that increasing the value of λ would improve classification performance, and, after some threshold, begin decreasing segmentation performance. This is because increasing the value of λ decreases the relative importance of segmentation performance in favor of classification performance. While we did observe that some λ values greater than 0 achieved greater segmentation performance than when λ was 0, the difference was only 0.5% (see Figure 3). Increasing the value of λ too much led to a decrease in segmentation performance for the U-Net, but did not necessarily lead to a corresponding increase in classification performance. It seems that once a non-zero λ is used, classification performance quickly jumps up relative to zero λ.

Generating Large Scale 3D Reconstructions. We further evaluated the qualitative performance of our U-Net by generating dense 3D reconstructions of the visible brain microstructure across various brain regions. To generate these reconstructions, we divided the input raw image volumes into non-overlapping 128x128x5 patches, passed each of those patches through the U-Net to get a 128x128 segmentation output per patch, and stitched them together into 3D microstructure volumes. As shown in Figure 4, the 3D/2D Contextual U-Net achieves reasonable 3D reconstructions of the brain microstructure with clear visual differences between each brain region. These 3D reconstructions further confirm the difficulty of segmenting these blood vessels. Though larger blood vessels are well-connected in 3D, some smaller vessels have breaks in them. This is likely due to the method we used to stitch together 3D reconstructions; our U-Net outputs individual 128x128 segmented patches, which we stitched together into a 3D reconstruction. Strategies aimed at mitigating this limitation are addressed in the Discussion section.

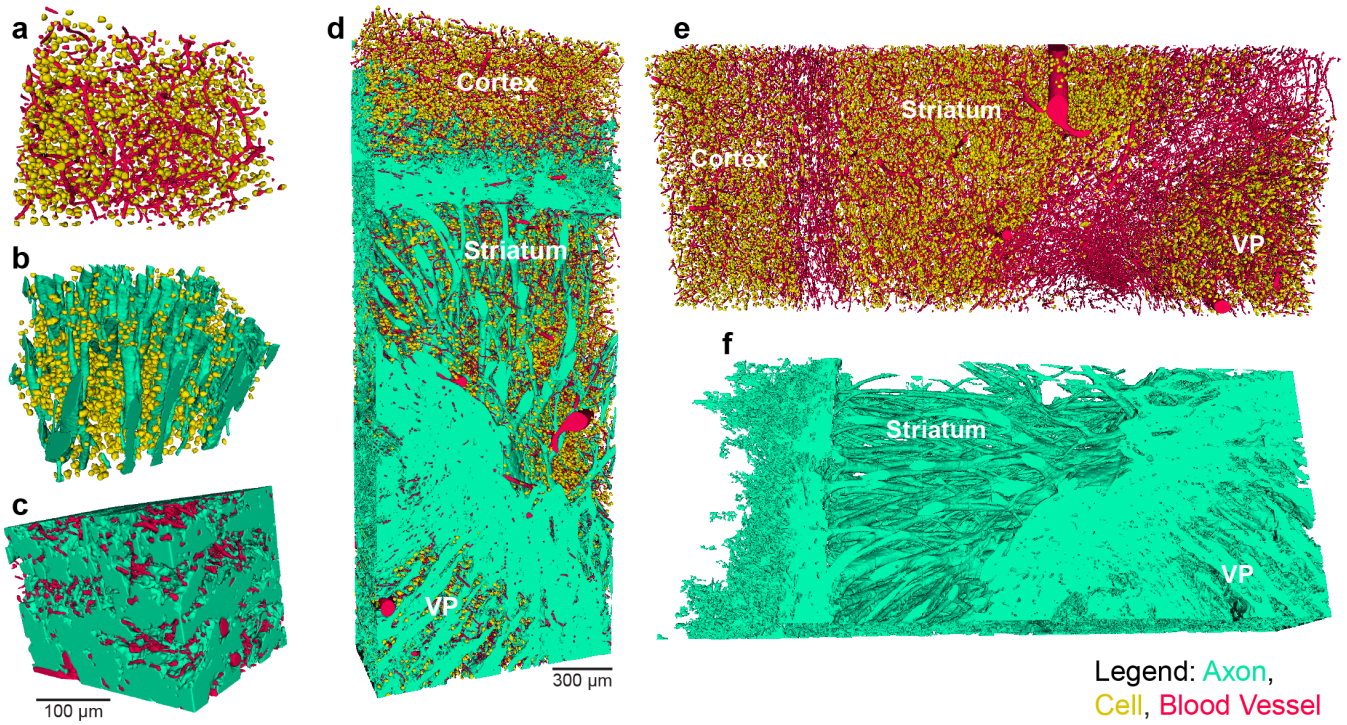

Figure 4: *3D reconstructions generated by combining outputs from a 3D/2D Contextual U-Net.* The color scheme used throughout these images is red for blood vessels, yellow for cells, and aqua for myelinated axons. (a), (b), and (c) show Cortex (cells and vessels), Striatum (myelinated axons and cells), and VP (myelinated axons and vessels) regions, respectively. (d), (e), and (f) show the same larger Cortex to VP brain region with a focus on different microstructures.

6 Discussion and Conclusion

In this thesis, we introduced and evaluated a 3D/2D Contextual U-Net architecture on a thalamocortical mouse brain dataset. We showed that the network can achieve reasonable quantitative performance, close to human-level IRR, with relatively little training data on the segmentation task. The network's qualitative performance was also strong, with the 3D reconstructions demonstrating that our method recovers fine microstructure detail for cells, vessels, and myelinated axons; blood vessels proved to be the most difficult structures to segment because they are often thin and difficult to distinguish from background. We also introduced a loss weighting parameter, λ, and investigated how segmentation performance and classification performance are influenced by varying λ. We found that some non-zero values of λ slightly improve segmentation F1 score relative to the 0 baseline and that small values of lambda quickly improve classification F1 score. Furthermore, we found that the visual "blockiness" of the correlation matrices associated with the latent representations increased as λ increased; this makes sense, because a greater λ causes the model to weight classification more than segmentation, meaning that data samples from different brain area classes should be more separable by the time they reach the bottom of the encoding path.

Our 3D/2D Contextual U-Net we introduced in this paper can be a useful tool for neuroscientists who want to perform rapid large-scale classification and segmentation of heterogeneous brain-image datasets. The automated segmentation and classification across large, diverse brain regions can be used to create multi-scale brain maps that incorporate both the microscale microstructure information and the macroscale brain area information. Such maps could be a useful tool in the study of brain anatomy and disease progression. Because the model performs well with relatively little training data, it could be used by other researchers with cost or time constraints who may want to adapt our method by using training data specific to their task. Potential future directions of this work include leveraging more training data and measuring its impact on performance, scaling our technique to even larger brain samples, and investigating post-processing strategies to improve the 3D reconstructions by, for example, enforcing blood vessel connectivity in 3D space.

References

[1] L. Yang, Y. Zhang, J. Chen, S. Zhang, and D. Z. Chen, "Suggestive annotation: A deep active learning framework for biomedical image segmentation," in *International conference on medical image computing and computer-assisted intervention*, pp. 399–407, Springer, 2017.

[2] R. Caruana, "Multitask learning," *Machine learning*, vol. 28, no. 1, pp. 41–75, 1997.

[3] V. Kaynig, A. Vazquez-Reina, S. Knowles-Barley, M. Roberts, T. R. Jones, N. Kasthuri, E. Miller, J. Lichtman, and H. Pfister, "Large-scale automatic reconstruction of neuronal processes from electron microscopy images," *Medical image analysis*, vol. 22, no. 1, pp. 77–88, 2015.

[4] A. de Brebisson and G. Montana, "Deep neural networks for anatomical brain segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pp. 20–28, 2015.

[5] R. Mehta and J. Sivaswamy, "M-net: A convolutional neural network for deep brain structure segmentation," in *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*, pp. 437–440, IEEE, 2017.

[6] J. Prasad, A. Balwani, E. Johnson, J. Miano, V. Sampathkumar, V. de Andrade, M. Du, R. Vescovi, C. Jacobsen, D. Gursoy, N. Kasthuri, and E. Dyer, "A three-dimensional thalamocortical dataset for characterizing brain heterogeneity," *under review in Nature Scientific Data*, 2020.

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

[8] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*, pp. 234–241, Springer, 2015.

[9] E. L. Dyer, W. G. Roncal, J. A. Prasad, H. L. Fernandes, D. Gürsoy, V. De Andrade, K. Fezzaa, X. Xiao, J. T. Vogelstein, C. Jacobsen, *et al.*, "Quantifying mesoscale neuroanatomy using x-ray microtomography," *Eneuro*, vol. 4, no. 5, 2017.

[10] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[11] D. Amodei, S. Ananthanarayanan, R. Anubhai, J. Bai, E. Battenberg, C. Case, J. Casper, B. Catanzaro, Q. Cheng, G. Chen, *et al.*, "Deep speech 2: End-to-end speech recognition in english and mandarin," in *International conference on machine learning*, pp. 173–182, 2016.

[12] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, pp. 1097–1105, 2012.

[13] Y. Goldberg, "A primer on neural network models for natural language processing," *Journal of Artificial Intelligence Research*, vol. 57, pp. 345–420, 2016.

[14] C. Sommer, C. Straehle, U. Koethe, and F. A. Hamprecht, "Ilastik: Interactive learning and segmentation toolkit," in *2011 IEEE international symposium on biomedical imaging: From nano to macro*, pp. 230–233, IEEE, 2011.

[15] S. Ruder, "An overview of multi-task learning in deep neural networks," *arXiv preprint arXiv:1706.05098*, 2017.

[16] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3431–3440, 2015.

[17] S.-M. Guo, L.-H. Yeh, J. Folkesson, I. Ivanov, A. P. Krishnan, M. G. Keefe, D. Shin, B. Chhun, N. Cho, M. Leonetti, *et al.*, "Revealing architectural order with quantitative label-free imaging and deep neural networks," *BioRxiv*, p. 631101, 2019.

[18] J. Prasad, A. Balwani, E. Johnson, J. Miano, V. Sampathkumar, V. D. Andrade, K. Fezzaa, M. Du, R. Vescovi, C. Jacobsen, K. P. Kording, D. Gursoy, W. G. Roncal, N. Kasthuri, and E. Dyer, "A three-dimensional thalamocortical dataset for characterizing brain heterogeneity: Microstructure Annotations (NumPy)," `https://figshare.com/articles/A_three-dimensional_thalamocortical_dataset_for_characterizing_brain_heterogeneity_Microstructure_Annotations_NumPy_/12153516`, 2020.

[19] P. A. Yushkevich, J. Piven, H. Cody Hazlett, R. Gimpel Smith, S. Ho, J. C. Gee, and G. Gerig, "User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability," *Neuroimage*, vol. 31, no. 3, pp. 1116–1128, 2006.

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

[21] J. Prasad, A. Balwani, E. Johnson, J. Miano, V. Sampathkumar, V. D. Andrade, K. Fezzaa, M. Du, R. Vescovi, C. Jacobsen, K. P. Kording, D. Gursoy, W. G. Roncal, N. Kasthuri, and E. Dyer, "A three-dimensional thalamocortical dataset for characterizing brain heterogeneity: Region of Interest Annotations (Nrrd)," `https://figshare.com/articles/A_three-dimensional_thalamocortical_dataset_for_characterizing_brain_heterogeneity_Region_of_Interest_Annotations_Nrrd_/12153549`, 2020.