Systematic Analysis of Tau Segmentation Methods

Alexander Wurts

October 9, 2019

Abstract

β-amyloid plaques and tau neurofibrillary tangles (NFTs) are hallmark pathologies in Alzheimers disease . The amount of tau (also referred to as the tau burden) is an important metric used to determine stages of Alzheimers disease. Recent work by Signaevsky, et. al. [1] has shown that convolutional neural networks can be used to determine large scale accumulation of tangles in immunohistochemically stained tissue. The goal of this project is to compare the performance of several networks for the segmentation, quantification and classification of tau tangles and -amyloid plaques. The networks to be compared include U-Net, FCNet, SegNet and Mask-RCNN. The goal of this open-ended research is to evaluate the effectiveness of deep convolutional networks in performing unsupervised classification on noisily labeled data.

[1]

Contents

[1]DISTRIBUTION STATEMENT A. Approved for public release. Distribution is unlimited. This material is based upon work supported by the Under Secretary of Defense for Research and Engineering under Air Force Contract No. FA8702-15-D-0001. Any opinions, findings, conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Under Secretary of Defense for Research and Engineering. © 2019 Massachusetts Institute of Technology. Delivered to the U.S. Government with Unlimited Rights, as defined in DFARS Part 252.227-7013 or 7014 (Feb 2014). Notwithstanding any copyright notice, U.S. Government rights in this work are defined by DFARS 252.227-7013 or DFARS 252.227-7014 as detailed above. Use of this work other than as specifically authorized by the U.S. Government may violate any copyrights that exist in this work.

1 Executive Summary

Doctors would like to diagnose tau-opathies from images. There are multiple tau-opathies, of which Alzheimer's disease is a significant one.

Tau protein in a normal brain is bound to axonal microtubules. There are many theories as to how Alzheimers truly works, one of them is the tau theory. In Alzheimer's disease a proportion of tau protein becomes abnormally phosphorylated and is no longer associated with axonal microtubules, but instead accumulates in filaments throughout affected nerve cells [5]. This leads to a disintegration of the microtubules destroying the structure of the cells cytoskeleton which collapses the neurons transport system. Eventually the collapse results in the death of the cell. [7] hypothesizes that the cell death caused by tau is the primary cause of the disease .

Through the use of Deep Learning images can be pre-processed for doctors to increase understanding, and reduce the time required to analyse an image. We focus on systematically finding the best model for segmentation by trying SegNet [2], UNet [15] and Fully Connected Network [12]. Training, validation and testing will be used to determine the best network. Segmentation of the nuclei, and tau clustering can generate quantitative data for clinicopathological correlations as well as additional studies. We introduce two novel improvements, a new model training methodand an analysis of current state of the art segmentation algorithms to determine which works the best for WSI.

Through the use of AI and DL, segmentation can be developed to recognize, classify, and quantify diagnostic elements of tau-opathies. A strategy has been developed and tested using a fully convolutional network (FCN) and showed promising results [16].Testing other advanced neural networks such as SegNet, FCN, and other segmentation networks could lead to better results.

The project used anonymized data supplied by Mass General Hospital. All names and any connection to an individual has been removed. Each position within a WSI is labeled with 3 classes: background, tau, and nuclei. The data was sparsely labeled. In table 1 a large majority of the tau nuclei is not labeled.

The data was sparsely labeled, which upon normal evaluation produced a large amount of false positives. To generate accurate analysis on only the labeled elements, the predictions need to be cropped around the ground truth. To crop the ground truth was expanded on all sides by 20 pixels, then used to crop the prediction. The network prediction was then compared to the ground truth to calculate the various metrics.

Figure 1 shows the best results using Segnet. The top is noticably more red, and the bottom is more blue. This shows the expansion of the disease through the brain. Quantification of the amounts of tau versus nuclei will produce valuable data.

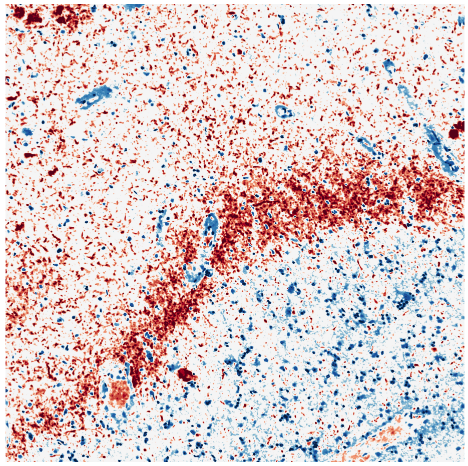


Figure 1: Segnet Red and Blue: Red is tau. Blue is Nuclei.

2 Introduction

Doctors would liketo diagnose tau-opathies from images. There are multiple tau-opathies, of which Alzheimer's disease is a significant one.

Tau protein in a normal brain is bound to axonal microtubules. There are many theories as to how Alzheimers truly works, one of them is the tau theory. In Alzheimer's disease a proportion of tau protein becomes abnormally phosphorylated and is no longer associated with axonal microtubules but instead accumulates in filaments throughout affected nerve cells [5]. This leads to a disintegration of the microtubules destroying the structure of the cells cytoskeleton which collapses the neurons transport system. Eventually the collapse results in the death of the cell. [7] hypothesize that the cell death caused by tau is the primary cause of the disease. The complexity and overlapping nature of many neurological degenerative diseases make diagnosing a very challenging job, demanding high expertise. The only way to confirm diagnosis of tau-opathies is through microscopic analysis of stained postmortem sections. The time and analysis required for such a diagnosis is significant, as depends upon the observer [16]. The current method, the Braak staging system, uses a mainly qualitative approach. The three tiers are, no tau, some tau, and a lot of tau. An accurate quantitative approach is demanded by the evolving research and healthcare standards [1].

Recent advances in technology have shown success in classifying and grading the progression of breast and prostate cancer [11]. Microscopic analysis of postmortem whole slide image (WSI) brain samples remains the only way to confirm diagnosis of tauopathies. Specifically, an understanding of the quantities of various cells, and tau configurations. The previous method, Braak Staging, requires a highly trained doctor to find and count neurofibrillary tangles (NFT) [3]. However due to the significant amount of time required, they could not fully label a WSI. Recently automatic segmentation of neurofibrillary tangles (NFT) has been shown possible with artificial intelligence [16]. AI Segmentation can scan an entire WSI with high accuracy in as little as 30 minutes. A more comprehensive segmentation is desirable. We aim to develop deep learning network that segments all tau, and cell nuclei. By segmenting more elements, more data is available for post processing and aiding in diagnosing.

Through the use of Deep Learning images can be pre-processed for doctors to increase understanding, and reduce the time required to analyse an image. We focus on systematically finding the best model for segmentation by trying SegNet [2], UNet [15] and Fully Connected Network [12]. Training, validation and testing will be used to determine the best network. Segmentation of the nuclei, and tau clustering can generate quantitative data for clinicopathological correlations as well as additional studies. We introduce two novel improvements, a new model training methodand an analysis of current state of the art segmentation algorithms to determine which works the best for WSI.

3 Background

3.1 Alzheimer's Disease

Alzheimers disease (AD) is a type of Dementia caused by a disease of the brain. It is normally characterized by a chronic and continuing degrade in higher brain functions such as memory, thinking, learning capacity and judgement. Dementia affects mainly people over 65, with only 2% of cases starting before 65. Every five years after 65 the prevalence doubles. Cortical amyloid plaques and neurofibrillary tangles are present in the most common manifestations. These appear in one half to three quarters of all cases. Alzhemer's and other forms of dementia are not race, or culture specific [13].

For diagnosis, doctors and other clinicians use impariment of cognitive function, including loss of independent living. As the disease progresses, behavioral and psychological symptoms of dementia (BPSD) are the most relevant. BDSD can consist of agitation, aggression, calling out repeatedly, sleep disturbance, wandering and apathy. All of these provide significant strain on caretakers. For tau-caused Alzheimers there is currently no treatment. Other causes such as oxygen deprivation if caught early can be treated, but this rarely happens. Once diagnosed a person can expect to live 5-7 more years in a developed country [13].

3.2 Image Segmentation

Image segmentation is the process of labeling sections or pixels in an image, to make the image more meaningful and easier to for additional analysis. There are many methods for segmentation, and it is considered an unsolved problem. Uses for segmentation include self-driving cars to process the image input by the cameras [9]. An example of segmentation for cars is shown in Figure 2. In biology segmentation is commonly used for extracting the locations of tumors, or other data from medical data.

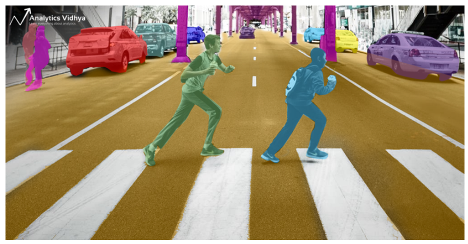


Figure 2:

Some simple methods use a threshold approach to remove light items on a dark background. More complex methods utilize edge detection. Both methods are effective in the correct use cases [4]. The state of the art methods using convolutional neural networks significantly improve results due to their ability to find patterns in very high dimensional data.

3.3 Image Segmentation and Deep Learning

Through the use of AI and DL, segmentation can be developed to recognize, classify, and quantify diagnostic elements of tau-opathies. A strategy has been developed and tested using a fully convolutional network (FCN) and showed promising results [16]. Testing other advanced neural networks such as Segnet FCN and other segmentation networks could lead to better results.

3.4 Deep Learning

Deep learning allows computational models that are composed of multiple processing layers to learn representations of data with multiple levels of abstraction. These methods have dramatically improved the state-of-the-art in speech recognition, visual object recognition, object detection and many other domains such as drug discovery and genomics [11]. Deep learning uses back-propagation to alter internal model parameters to learn a given dataset. After learning on an initial dataset the model can then be applied to new data.

The amount of data required to train a deep learning network can be on the order of thousands of images. ImageNet contains 14 million photos and 22 thousand categories [21]. Training neural networks on this data takes a significant amount of time, but produces great results as seen with AlexNet [10] and VGG [17]. Constructing a large high quality dataset is the first part in any deep learning problem.

3.5 Convolutional Neural Networks

Convolutional neural networks (CNNs) are an implementation of deep learning. They can be extremely accurate at image classification and segmentation when trained on enough data. CNNs typically take a long time to train; much of the time used to perform matrix multiplications. In tasks such as image classification neural networks can achieve accuracies of up to 98% [18]. In a task like cell labeling humans are very accurate, but not necessarily fast. CNN can be trained to be very fast and accurate, providing decision support to a doctor within minutes, compared with human examination of a brain whole slide image (WSI) which takes much longer.

CNNs utilize convolutional layers which are matrices that are multiplied by an input matrix. The convolutional matrix is filter that activates when a specific pattern matches the filter. The filter is then translated across an array and produces an activation matrix. As a large network trains the initial layers tend to form simple edge, and pattern filters. Deeper layers use lower layers to form more complex filters before a final layer connects the features generated by the network to the output classes.

3.5.1 Layers

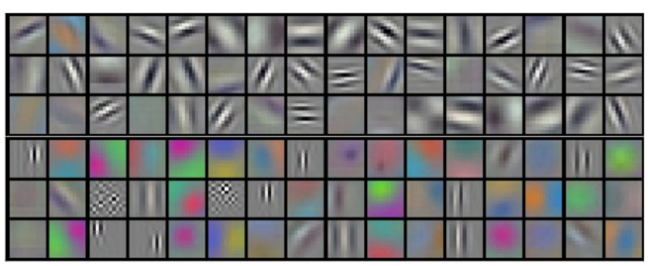


Figure 3: Low Level Convolutional Filter Examples [8]

Convolutional Layers perform a majority of the heavy computing. They are a set of small filters that slide across the image producing activations and results based on their input. Intuitively the network will activate on simple features such as edges, or certain colors. Deeper in the network the filters could recognize entire structures such as letters or numbers [8].

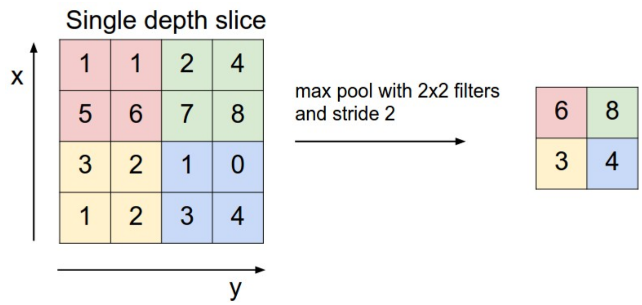


Figure 4: Max Pooling Layer [8]

Pooling Layer Pooling layers reduce the size of their input while trying not to lose any valuable information. Pooling layers reduce the size of input by sliding a function across their input. Figure 4 shows a Max Pool layer which takes the maximum for each subsection of the input [8].

Fully Connected Layer Fully Connected layers connect to every neuron in the preceding layer. The activations can be calcuated by a matrix multiplication followed by the bias offset [8].

ReLU: Rectifier Linear Unit ReLU layers are activation layers. When greater than a preset threhsold they output the input, when less they output 0. Mathematically they are respresented by $max(0, x)$ if the threshold is 0 [8].

3.6 Transfer Learning

Transfer learning is using a network trained on another dataset and task as the starting point for a different domain and task. Given a source domain, D_S and source task, T_S a target domain D_T and target task T_T transfer learning aims to use the information in D_S on T_S to improve the accuracy of D_T in S_T [22].

Transfer learning has been commonly used in image classification and segmentation. Many low level image features are the same among all types of images. Deep learning takes a lot of data, and the acquiring enough domain specific data to train a neural network can be very challenging. Transfer learning provides a starting point to work from. It allows problems without enough data to train a full network to fine-tune a network. This type of fine tuning is called supervised domain adaptation [23].

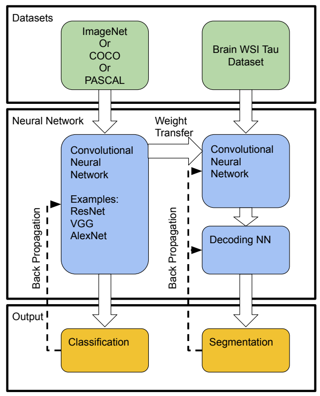


Figure 5: Trasnfer Learning Pipeline

3.7 Deep Learning in Medical Image Analysis

Deep learning has pervaded every aspect of medical image analysis. They are used for everything from neurological data to musculoskeletal. For Brain Magnetic resonance imaging (MRI) data. deep learning is used for many diseases including Schizophrenia and Huntingtons. The state of the art research is using Deep Learning to create connections between MRI brain scans and various diseases. The logic is that if there is a pattern between certain MRI brain activity and a disease a deep learning network such as a CNN may be able to find it [11].

3.8 Models

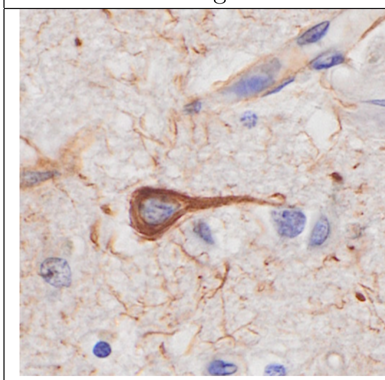
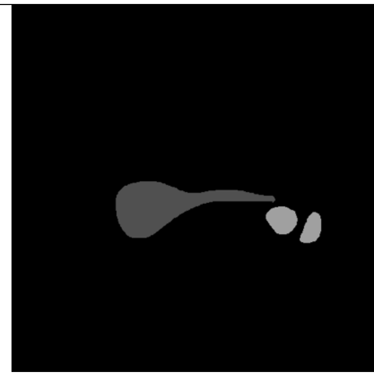
3.8.1 UNet

UNet consists of a contracting path on the left and an expanding path on the right. The left side follows the typical architecture of a CNN. The right side consists of an up-sampling layer followed by a convolution layer to reduce the number of features. Finally the remaining features are mapped to specific classes for the output. UNet specializes in learning with a small dataset. It uses a significant amount of data augmentation to learn the most out of a small dataset, as small as thirty photos.

3.8.2 SegNet

SegNet has two sides. The convolution downsampling into a big feature space, and a upsampling half to create the segmentation map. The first is topologically identical to VGG16. The decoder uses pooling indices and upsampling to reduce the number of learnable features. The results are then convolved

Table 1: Example Data: Left image is the raw data. The right image is the pixel-wise labels. Black is background, dark gray is tau, and light gray is nuclei.

Image	Labeled Mask
	

with trainable features to produce dense feature maps. The novelty of SegNet lies is in the manner in which the decoder up samples its lower resolution input feature map(s). The design goal of SegNet focused on scene understanding segmentation, and use memory and processor time efficiently.

3.8.3 Fully Connected Network

FCN matches the standard CNN architecture more closely. It consists of convolutional layers transforming an image into feature matrices, followed by a large pixel wise prediction layer to form the final segmentation.

4 Methods and Materials

4.1 Materials

The project used anonymized data supplied by Mass General Hospital. All names and any connection to an individual has been removed. Each position within a WSI is labeled with 3 classes: background, tau, and nuclei. The data was sparsely labeled. In Table 1 all strands of tau, and most nuclei are not labeled.

4.2 Programming Language and Libraries

Python 3.7 [20] was used as the primary language. Pytorch [14] was used for training, testing and evaluating networks on GPUs, Python Image Library (PIL) for image processing, and Numpy [19] for data processing. All training, testing and evaluation was performed on the MIT Supercloud supercomputer.

4.3 Network Training

WSI Images have a resolution 120,000 by 120,000 pixels averaged over the collection of images. Many small slices of 512 by 512 were randomly selected from the large image for training and testing. The resulting dataset consisted of 838 images. It was split into three sections, half for testing, one third for testing and one sixth for validation. The training data was then augmented by flipping vertically, horizontally, and both to quadruple the size of the training data to 1676.

Stochastic gradient descent was used with a cross entropy loss function. The learning rate was set at 0.08 and multiplied by 0.01 every epoch. The networks were trained for 20 epochs. The networks were initiallized trained on large segmentation datasets.

Every network trained was saved. Post-training analysis was performed to calculate the intersection over union, true positive rate, false positive rate, and ROC curve.

Pre-trained base networks were used for every network. SegNet [2] used VGG19 [17] and FCN [12] used Resnet101 [6]. UNet did not have a pre-trained base.

4.3.1 FCN Training

In order to train the FCN, the dataset needed to be reduced to images that had labeled examples of both classes. This balanced the data allowed the network to learn the classes better. Without balanced classes the network did not learn nuclei very well and often not at all.

4.3.2 Segnet Training

Segnet was trained only on labeled tau and bvackground, but learned to distinguish between tau, nuclei and background. Once trained the output for the background class was thresholded at multiple values to segment the three classes. All results above a value were labeled nuclei, and below a different value were labeled tau. The middle values were labeled background. The learned behavior eliminated the need for labeled nuclei data.

4.4 Segmentation Mask Thresholding

When training on data with only tau labeled the network starts to differentiate between background and nuclei. After training for 15-20 epoch the network background class activates more on nuclei than normal background. The difference allows a threshold to be placed on the segmentation result to effectively segment the nuclei.

4.5 Analyzing Entire Whole Slide Image

Segmenting an entire image at 120,000 by 120,000 pixels is not possible given the GPUs and memory constraints. The WSI was divided into 512 by 512 pixel sections offset by 492 per section. The edges were trimmed to eliminate erroneous results along the edge of each section. Each segment was run through the neural network and the output as a three layer Numpy array. The Numpy array was then transformed into a tiff. The tiff could then be viewed using mapping software for analysis.

4.6 Data Storage and Code Backup

All code and data is stored on the MIT Supercloud Supercomputer. A shared directory was created to ease collaboration between myself and Lincoln Laboratory. The supercomputer does not backup their data system, so daily backups of the code were saved and moved to my local computer. Git was used for version control.

4.7 Testing Protocol

4.7.1 Evaluation procedure

The data was sparsely labeled, which upon normal evaluation produced a large amount of false positives. To generate accurate analysis on only the labeled elements, the predictions need to be cropped around the ground truth. To crop, the ground truth was expanded on all sides by 20 pixels, then used to crop the prediction. The network prediction was then compared to the ground truth to calculate the various metrics.

4.7.2 Segmentation Network Success Measures

A confusion matrix of the number of true positives, false positives, true negatives, and false negatives was constructed for every network at every epoch during the training process.

The true positive rate, false positive rate, intersection over union (IoU), and dice loss score will be used to determine success. Determining which network is the best based on these metrics is the goal. We cannot compare results to [16] because the data is not available to the public. A good qualitative segmentation, differences above 0.1 for IoU between all networks to determine which one is the best.

4.7.3 ROC Curve

ROC curves show the values of the true positive rate (TPR) and the false positive rate (FPR) for varying threshold values. A ROC curve that is above the line $y = x$ represent results that have a TPR greater than FPR. The goal is the maximize TPR and minimize FPR.

4.8 Results Colorization

The results were thresholded to label tau, background and nuclei. Additional thresholds were added to differentiate between the strength of prediction. The thresholds were determined based on the optimal point in the ROC curve.

5 Results

5.1 Timing

Table 2: Timing

Network	Training Time 20 Epochs (Minutes)	WSI Evaluation Time (Minutes)
Segnet	21.5	30.5
FCN	25.2	60.0
UNet	5.45	62.7

The time to train the networks ranged from 21 to 25 minutes. The segmentation software used 2 Nvidia Volta GPUs and 16 processors. The processing time decreases when more GPUs are used.

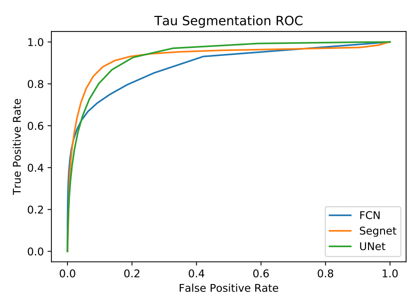
5.2 Quantitative Results

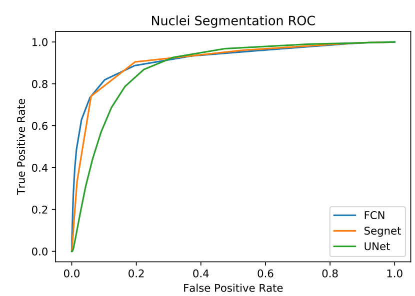
Table 3: Best Intersection of Union and Dice Loss

Network and Class	IoU	Dice Loss
SegNet tau	0.611	0.234
FCN tau	0.520	0.297
UNet tau	0.544	0.272
Segnet Nuclei	0.538	0.309
FCN Nuclei	0.542	0.306
UNet Nuclei	0.387	0.441

The IoU and Dice loss scores were chosen based on the best TPR and FPR values. Segnet had a better IoU score for tau, and within 0.004 for nuclei. Both networks had an average IoU greater 0.5. The numerical values are good for determineing which network is better. However, since the data was sparsely labeled the absolute value cannot be compared to other papers. UNet performed very poorly on nucleus segmentation. For tau, UNet did well.

Table 4: ROC Curves by Network and Class: The orange, blue, and green lines are Segnet, FCN, and UNet respectively. The left graph is for the tau class, and the right graph is for Nuclei.

<table>
<tr><th>tau ROC</th><th>Nuclei ROC</th></tr>
<tr><td>


</td><td>


</td></tr>
</table>

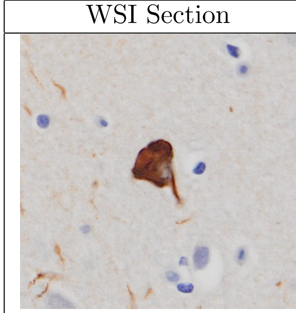
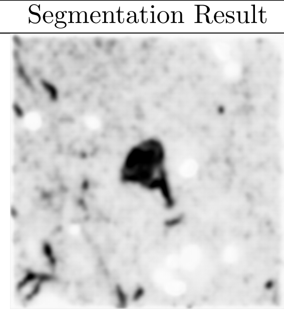
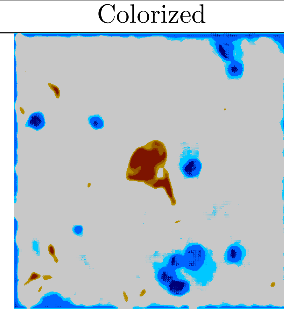
5.2.1 ROC Curves

Table 4 shows the ROC curve for tau and Nuclei with FCN and Segnet. UNet is higher than FCN with tau, but lower at nuclei than the other two networks.

5.3 SegNet

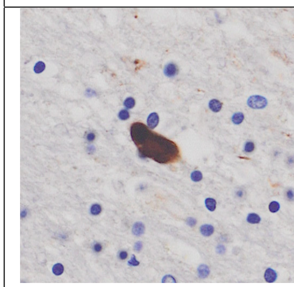
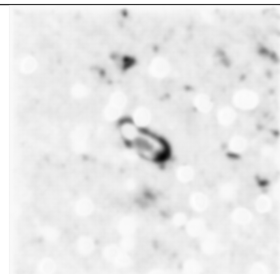
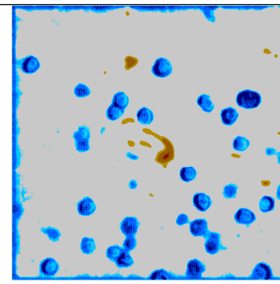
5.3.1 Qualitative Results

Table 5: Segnet Segmentation Example: tau is brown and Nuclei is blue. Darker colors represent higher network confidence in segmentation.

WSI Section	Segmentation Result	Colorized
		

The resulting correctly segments the tau and nuclei in the example images. The raw output shows black areas where the network predicts there are tau, and white areas where the network predicts there are nuclei. The blue border on the image is caused by incomplete border data. The first 20 pixels of the images around the border can be ignored. The network has a few false postivies. The blue area immediately to the right of the central tau area segments the white area around the nuclei wrongly as nuclei. The vary dark tau region in the

Table 6: Segnet Bad Segmentation Example

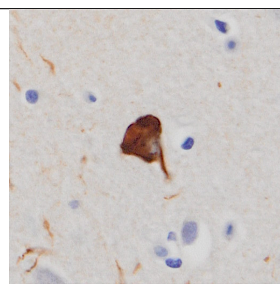
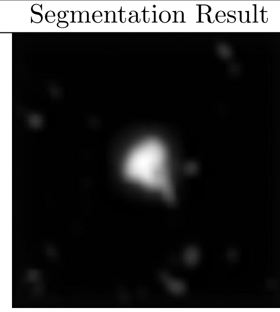
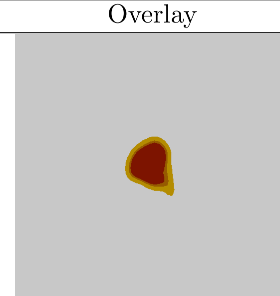
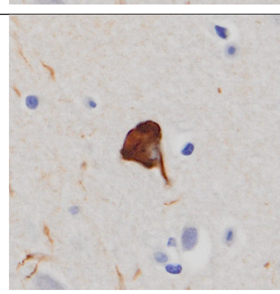
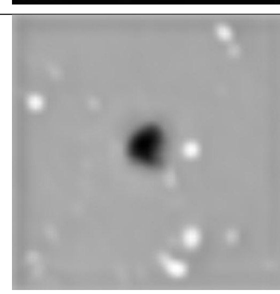
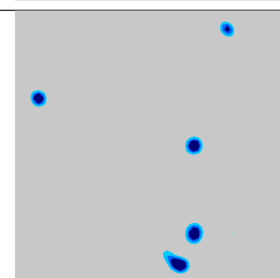
WSI Section	Segmentation Result	Colorized
		

center of the image was labeled incorrectly as background when it was infact tau. The nuclei segementation segmented a majority with minimal error. The region along the edge of the image labeled nuclei is noise caused by the edge.

5.4 Fully Connected Network (FCN)

5.4.1 Qualitative Results

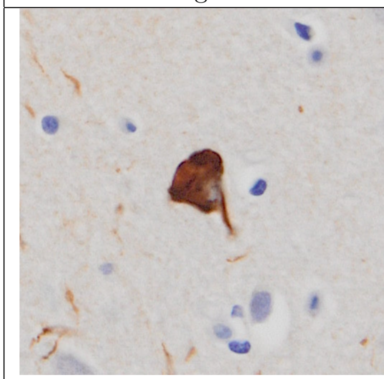
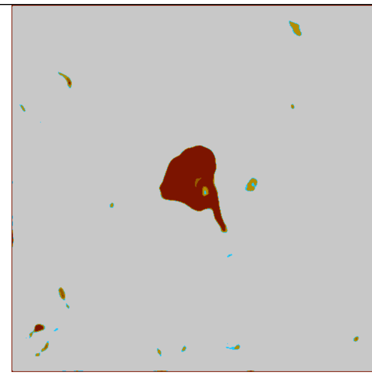
Table 7: FCN Results

Class	WSI Section	Segmentation Result	Overlay
Class 1 tau			
Class 2 Nuclei			

FCN did not fully segment the middle tau shape. The smaller sections of tau on the bottom edge of the picture were not labeled either. The nuclei segmentation was accurate on multiple cells, but missed some.

5.5 UNet Failure

Table 8: UNet Results

Original	Prediction
	

Unet failed to segement the three classes effectively. It was able to differentiate between tau and background, as well as nuclei and background, but not between tau and nuclei. The network was trained with various loss functions, learning rates, data augmentation, data class balance, and with singular classes, but could not learn effectively. What information it did learn did not generalize well. It learned to match the ground truth, but not expand to segment all occurrences of tau and nuclei. As shown in Table 5.5 it segments the tau correct, but not the nuclei.

6 Discussion

6.1 Imagery

Visually the segmentation accuracy of Segnet is extremely accurate and comprehensive. When viewed zoomed out it is easy to see the patterns of tau and nuclei throughout the image and where they cluster.

The results from Table 5 show accurate segmentations on both tau and nuclei. This corresponds with a IoU above 0.5, and a low dice score. The results from Segnet were qualitatively much better than FCN net even though the quantitative measures were quite similar. UNet failed to distinguish between tau and nuclei enough to produce usable results. The FCN tau results could be overlayed ontop of the Segnet results as an additional class to segment neurofibrillary tangles. When entire WSI images were segmented the color scheme used to differentiate between background, tau and nuclei enabled faster analysis of the data.

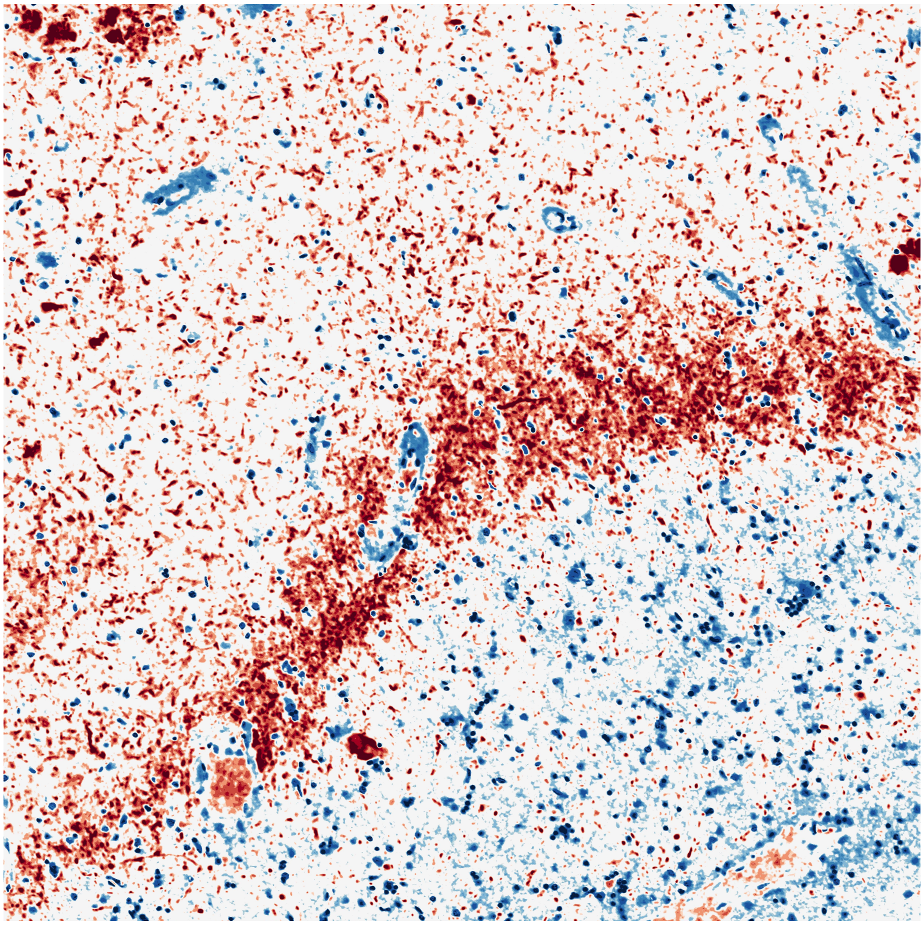


Figure 6: Segnet Red and Blue: Red is tau. Blue is Nuclei.

6.2 Quantitative

Based on the ROC curve along with IoU and Dice loss Segnet provided the best segmentation. The quantitative results are good for determing which network is the best. The qualitative results are better for determining the utility of the network.

6.3 Interpreting Results

Figure 6 shows the value of segmenting an entire WSI. There is a large band of tau across the center of the image. Below the band there is less tau, but more nuclei, and above there is less nuclei and more tau. The larger denser sections of tau and nuclei are easily distinguishable.

6.4 Use for Doctors

Now that a computer can process the data a doctor no longer needs to spend the time manually labeling the image. The Braak staging process can now be modified to include actual numbers instead of none, some and many. The process has been automated to save time, and allow them to only review the finished results. An automatic diagnosis feature is possible based on the results using predetermined thresholds of the ratio between tau and nuclei. More research into the formations of tau and nuclei in the image enabled by this research will allow better understanding of Alzheimers and could help treat patients with the disease.

7 Conclusion

After an analysis of segmentation methods on brain WSI's Segnet has created the best results as shown through quantitative and qualititative analysis methods. FCN performs better when the tau clusters are large and distinct, but fails to identify small strands of tau. UNet segments tau very well, but does not recognize Nuclei. Doctors can now use the automatically labeled data to improve analysis and decrease time spent labeling. Future research can use the quantification of tau and nuclei thia paper produces to increase our understanding of the disease and how to improve diagnosis and treatment. There is now significant potential for using the segmentation results to segment additional patterns that are now more visible.

References

[1] Shaimaa Al-Janabi, André Huisman, and Paul J Van Diest. Digital pathology: current status and future perspectives. *Histopathology*, 61(1):1–9, 2012.

[2] Vijay Badrinarayanan, Alex Kendall, and Roberto Cipolla. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *CoRR*, abs/1511.00561, 2015.

[3] Heiko Braak and Eva Braak. Neuropathological stageing of alzheimer-related changes. *Acta neuropathologica*, 82(4):239–259, 1991.

[4] Rajeshwar Dass and Swapna Devi. Image segmentation techniques 1. 2012.

[5] M Goedert, MG Spillantini, and RA Crowther. Tau proteins and neurofibrillary degeneration. *Brain pathology*, 1(4):279–286, 1991.

[6] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, June 2016.

[7] Khalid Iqbal, Alejandra del C Alonso, She Chen, M Omar Chohan, Ezzat El-Akkad, Cheng-Xin Gong, Sabiha Khatoon, Bin Li, Fei Liu, Abdur Rahman, et al. Tau pathology in alzheimer disease and other tauopathies. *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease*, 1739(2-3):198–210, 2005.

[8] Andrej Karpathy. Convolutional neural networks (cnns / convnets).

[9] Çağrı Kaymak and Ayşegül Uçar. A brief survey and an application of semantic image segmentation for autonomous driving. In *Handbook of Deep Learning Applications*, pages 161–200. Springer, 2019.

[10] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 25*, pages 1097–1105. Curran Associates, Inc., 2012.

[11] Geert Litjens, Thijs Kooi, Babak Ehteshami Bejnordi, Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Mohsen Ghafoorian, Jeroen Awm Van Der Laak, Bram Van Ginneken, and Clara I Sánchez. A survey on deep learning in medical image analysis. *Medical image analysis*, 42:60–88, 2017.

[12] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. *CoRR*, abs/1411.4038, 2014.

[13] World Health Organization. *Neurological disorders: public health challenges*. World Health Organization, 2006.

[14] Adam Paszke, Sam Gross, Soumith Chintala, Gregory Chanan, Edward Yang, Zachary DeVito, Zeming Lin, Alban Desmaison, Luca Antiga, and Adam Lerer. Automatic differentiation in PyTorch. In *NIPS Autodiff Workshop*, 2017.

[15] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. *CoRR*, abs/1505.04597, 2015.

[16] Maxim Signaevsky, Marcel Prastawa, Kurt Farrell, Nabil Tabish, Elena Baldwin, Natalia Han, Megan A Iida, John Koll, Clare Bryce, Dushyant Purohit, et al. Artificial intelligence in neuropathology: deep learning-based assessment of tauopathy. *Laboratory Investigation*, page 1, 2019.

[17] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. In *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, 2015.

[18] Hugo Touvron, Andrea Vedaldi, Matthijs Douze, and Hervé Jégou. Fixing the train-test resolution discrepancy. *CoRR*, abs/1906.06423, 2019.

[19] S. van der Walt, S. C. Colbert, and G. Varoquaux. The numpy array: A structure for efficient numerical computation. *Computing in Science Engineering*, 13(2):22–30, March 2011.

[20] Guido Van Rossum and Fred L Drake Jr. *Python tutorial*. Centrum voor Wiskunde en Informatica Amsterdam, The Netherlands, 1995.

[21] Kaiyu Yang, Klint Qinami, Li Fei-Fei, Jia Deng, and Olga Russakovsky. Towards fairer datasets: Filtering and balancing the distribution of the people subtree in the imagenet hierarchy. 2017.

[22] Jing Zhang, Wanqing Li, and Philip Ogunbona. Cross-dataset recognition: A survey. *CoRR*, abs/1705.04396, 2017.

[23] Jing Zhang, Wanqing Li, Philip Ogunbona, and Dong Xu. Recent advances in transfer learning for cross-dataset visual recognition: A problem-oriented perspective. *ACM Computing Surveys (CSUR)*, 52(1):7, 2019.