Optical Coherence Tomography Scans and Deep Learning based Alzheimer's Disease Diagnosis

Ahmet Alperen Güngör: agungor@sabanciuniv.edu, *Computer Science & Engineering, Junior*

Serra Yılmaz: serrayilmaz@sabanciuniv.edu, *Computer Science & Engineering, Senior*

Yasemin Turkan: yaseminturkan@sabanciuniv.edu, *Computer Science and Engineering, Supervisor*

Abstract

Alzheimer's disease (AD) is the most common form of dementia. It is an irreversible, progressive brain disorder marked by a decline in cognitive functioning with no treatment. Interest has been overgrown in retinal optical coherence tomography (OCT), a promising tool for early AD diagnosis. By analyzing retinal thickness, layer segmentation, and other morphological features, OCT scans provide valuable information that can aid in the early detection and monitoring of AD. This tool is more accessible, non-invasive, and cost-effective than alternative neuroimaging tools. OCT generates high-dimensional, multimodal imaging data. It is time-consuming and challenging to interpret and classify. Therefore, the integration of computer-aided Machine Learning approaches are required to aid reading and classifying. Machine learning explainability methods can help to highlight relevant patterns and even discover new ones.

In recent years, deep learning techniques, coupled with OCT images, have shown promise in AD diagnosis. Additionally, transfer learning, a powerful approach in deep learning, can be leveraged to enhance the performance of AD detection models using OCT images. Transfer learning involves transferring knowledge from a pre-trained deep neural network to a new task or dataset. In the case of AD diagnosis, a pre-trained model, typically trained on large-scale image classification tasks, can be fine-tuned and adapted to analyze OCT images for AD-related features. This abstract highlights the utilization of transfer learning techniques in the context of OCT-based AD diagnosis, specifically focusing on 3D segmented OCT image datasets. Our study utilizes a dataset comprising 3D OCT images that have been preprocessed and segmented to isolate the relevant retinal layers. The availability of 3D segmented OCT data enables the extraction of volumetric information, allowing for a comprehensive analysis of retinal changes associated with AD.

Keywords: Alzheimer's Disease, optical coherence tomography (OCT), optical coherence tomography angiography (OCTA), Machine Learning

1 Introduction

Alzheimer's disease is one of the most challenging diseases that our world faces, given its far-reaching impact on individuals, families, and societies worldwide. Alzheimer's disease is the most common form of dementia. It is an irreversible, progressive brain disorder marked by a decline in cognitive functioning with no treatment. It is characterized by a massive decrease in brain size due to the accumulation of proteins (amyloid-beta and tau) in the neurons. Eyes extend the brain as both the retina and brain grow from the same neural tube. Postmortem studies in AD also highlighted the collection of these proteins in the retina.

More recently, high-resolution visual imaging techniques, including optical coherence tomography (OCT), have been proposed as tools for evaluating structural changes in the retina of AD patients. Conventional diagnostic methods from medical images greatly depend on physicians' professional experience and knowledge. Artificial intelligence (AI) has improved the performance of many challenging tasks when working with high resolution, complex imaging data. Artificial neural networks are a subset of AI inspired by a simplification of neurons and their connections in the brain. Deep learning (DL) is a multi-layer structure of neural networks that mimics human learning by analyzing data with a given logical structure. This project focuses on using deep learning/transfer learning based analysis of retinal OCT scans for AD detection. Even though this technique is widely used to detect many other retinal diseases from OCT images, there is no application in AD.

Retinal scans obtained from OCTS devices are two-dimensional (2D) and three dimensional (3D) images. Despite their high performance, DL architectures are black-box models. Trusting their predictions is an important factor in using them for decision-making in medicine. Therefore, this research aims to train the model with retinal images and develop algorithms that will help clinicians to review and visualize the decision process. The power of explainability tools can also help to highlight relevant patterns and even discover new ones. We plan to investigate various learning methods to increase learning with small AD datasets by transferring knowledge from other studies and datasets.

In this paper, we are explaining our methods and implementation details in the second section, our results, discussion and future work in the third section. In the fourth section, the results of our study are presented.

2 Methods

2.1 Researched Models:

We researched multiple articles in order to find proper Convolutional Neural Network algorithms to build from scratch. Our aim involves classification of 3D input of OCT data - which is not very common.

We analyzed the following articles, and summarized & discussed potential benefits and limitations:

- *Aggregated Residual Transformations for Deep Neural Networks:* This article (Xie et al., 2016) elaborates on the concepts of the Inception and ResNet modules. The Inception model employs split-transform-merge operations to process input data. By splitting the input into lower-dimensional embeddings and transforming them using specialized filters, the model enhances learning capabilities. The subsequent merging of the transformed features through concatenation provides a comprehensive representation. However, adapting this approach to new datasets requires careful tuning of parameters, such as filter numbers and sizes. Instead of pursuing deeper networks, this study suggests exploring wider architectures. Additionally, incorporating 1x1 convolutions within a convolutional neural network (CNN) reduces input dimensions and enables the learning of patterns across channels. The utilization of varying convolution sizes, including 3x3 and 5x5 convolutions, facilitates the network's ability to capture diverse spatial patterns. Moreover, the integration of max pooling helps create smaller outputs by selecting the maximum value within each pooling region.

  The ResNeXT model may increase the ability to detect patterns - which is very crucial for detection of Alzheimer's Disease - ; however, it poses an important question regarding the suitability of this model for datasets with similar input size and type. If the data exhibits such characteristics, the additional flexibility provided by this model might not justify its increased complexity.

- *Learning Spatio-Temporal Features with 3D Residual Networks for Action Recognition:* This study (Hara et al., 2017) focuses on addressing the limitations faced in training deep 3D convolutional neural networks (CNNs) for action recognition. While deeper models yield higher accuracy, training such networks becomes challenging due to the large number of parameters involved. To mitigate this issue, the researchers propose the utilization of shortcut connections, inspired by ResNet architectures, to allow signals to bypass certain layers. These shortcuts alleviate training difficulties and enhance the network's ability to capture spatio-temporal features effectively. However, the choice of optimization algorithm, specifically the use of SGD instead of Adam, is questioned in the article. Although Adam is known for its computational speed and overall performance, the authors might have opted for SGD due to its suitability for larger datasets.

  The 3D ResNet model raises an additional question regarding the necessity of employing residual networks given a limited number of training samples. While depth is not solely determined by the number of samples, the complexity and specificity of image features play a crucial role. Extremely deep networks may require more memory and longer training time, potentially leading to overfitting when training on small datasets.

- *EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks:* This article (Tan & Le, 2019) explores an innovative approach to scale convolutional neural networks (CNNs) by optimizing three dimensions: depth, width, and resolution. Traditionally, scaling was focused on a single dimension, leading to inefficiencies. The study finds that scaling each dimension with a constant ratio yields improved efficiency. Scaling the depth dimension enables better capture of complex features, but it also poses the challenge of the vanishing gradient problem. To mitigate this issue, the authors suggest incorporating techniques such as residual networks, batch normalization, weight initialization, gradient clipping, and even LSTM. Batch normalization, in particular, reduces the dependence of gradients on parameter scale and initial values, facilitating the use of higher learning rates and regularizing the model.

This model may help our research in terms of computational resources, since our equipment is not the latest technology (for example, we preferred ResNet18 and ResNet10 models instead of ResNet50, since computational resources were not enough). In addition, by batch normalization and allowing faster convergence, computation time can be reduced as well.

- PointNet (Qi et al., 2017) is a deep neural network that classifies and segments 3D point clouds. Initially, our study aimed to use transfer learning techniques to expedite our progress using the PointNet algorithm. The goal was to reuse the pretrained PointNet model as the starting point for our new model to avoid building a model from scratch. The adjustment of the hidden layers and the output layer according to our classifications was what we were aiming for. However, the saved network of the model containing the weights and biases of the artificial neurons could not be located. When the model is trained, it is decided that it was unproductive to wait for the model to be trained since it took too much time. In the end, it was decided to give up on using PointNet for transfer learning and we started to search for other solutions on edge detection.
- MedNet (Chen et al., 2019) is a deep neural network model that uses datasets from various medical challenges in order to perform segmentation on 3D medical imaging. The model uses an encode-decode segmentation network called Med3D - which is trained with large amounts of 3D medical datasets. This pre-trained backbone of Med3D can be utilized for different segmentation tasks - which is ultimately what we try to achieve. By

adding classification layers on top of the backbone, a high-accuracy classification even with low amounts of data can be achieved. In order to do so for our Alzheimer's Disease classification, first we created 3 different environments to work with the model & dataset. We downloaded the data and pre-trained model to a Google Colab cloud & 2 laptops with (above) average GPUs. Most of the modules that the pre-trained model utilize were outdated, therefore we modified the code to work in Anaconda spyder with the latest torch and also fixed some issues regarding cuda and deprecations. We trained the base model with their original brain segmentation dataset once more for 200 epochs, which resulted in a 0.05 error rate. This is a good accuracy for their own task, but we also need to make sure it works well with our classification task as well. The results of the segmentations can be seen in the figures 8, 9 and 10. The image on the left is the outcome of the MedNet segmentation, the middle image is OCT segmentation result and the right image is the raw data. Below are the dice results of the segmentation:

- mean dice for class-1 is 0.912173004275817
- mean dice for class-2 is 0.15975092559917503
- mean dice for class-3 is 0.5305641949272135
- mean dice for class-4 is 0.9184622021687835

Our OCT and OCTA data are for classification task of Alzheimer's Disease; however, the model performs with high accuracy even on a small amount of classification data (however, since the amount of data is not very high, it may have a little correlation with overfitting on validation set). In order to achieve high accuracy on the classification task, the authors suggest training the data initially on the segmentation task before training the data for the classification task. Their backbone provides multiple ResNet models with varying depths, we used ResNet50 initially; however, our hardware could not handle it, and therefore preferred to use ResNet18 for the first segmentation training, and ResNet10 model for further steps. With only 80 training data and with 5 epochs, our model achieved 0.108 error rate, which is a really impressive number considering the lack of data/ epoch number.

2.2 Canny Edge Detection for Layer Segmentation of OCT Images

In this particular phase of the project, the primary objective was to detect the boundaries of retinal layers within optical coherence tomography (OCT) images utilizing various methods. The thickness measurements of these layers in OCT images have been identified as significant indicators of certain diseases, such as Alzheimer's Disease (Yang et al., 2010).

To address this objective, the chosen approach involved implementing the Canny Edge Detection algorithm. This algorithm was specifically selected as an edge detector due to its ability to generate thinner and cleaner lines within the images through a multi-stage process. Edge detection plays a crucial role in extracting computable, locally meaningful, and identifiable features from an image while simultaneously reducing the amount of data that needs to be processed. Among the available edge detection algorithms, the Canny Edge Detector is preferred for its utilization of non-maximum suppression to refine edges by removing ridges and thick lines, double thresholding to identify strong, weak, and irrelevant edges, and hysteresis to transform weak edges into strong ones based on their relationship to strong edges (OpenCV: Canny Edge Detection, n.d.).

Initially, the direct application of the Canny method from the cv2 library on the OCT images was attempted. However, due to the horizontal lines present in the OCT scans, the results seemed unsatisfactory. Figure 2 shows the obtained outcome. Consequently, preprocessing steps were implemented, starting with noise reduction. Given the considerable amount of noise outside our region of interest in the OCT scans, a colorbar was utilized to visualize the color map (Figure 3). This facilitated the identification of pixels considered noise and subsequently assigned a value of 0, represented by the color black. After excluding unrelated pixels by setting them to 0, the Canny method from the cv2 library was applied once more, but the results remained suboptimal. Thus, it was decided that more preprocessing on the image was necessary.

The subsequent steps involved blurring the image using Gaussian Blurring with a kernel size of 3 to reduce unwanted noise such as random brightness or incorrect color spots. Blurring also results in decreasing the size of the image, thereby reducing algorithmic complexity. After blurring, edge detection was initiated using a horizontal Sobel filter with a kernel size of 5, as the desired edges were primarily horizontal. Figure 4 illustrates the outcome of applying the horizontal Sobel filter. Non-maximum suppression was subsequently performed utilizing the theta value derived from the Sobel filter convolution, as it is shown in Figure 5. In the double thresholding step, specific hyper-parameters were chosen, with a low-threshold ratio of 0.03 and a high-threshold ratio of 0.06. The resulting image after thresholding and hysteresis can be seen in Figure 6. Finally, to facilitate understandability, the detected edges were overlaid on the original image, as demonstrated in Figure 7.

3 Discussion and Conclusion

As a result of the Edge Detection part of the project, some boundaries of the layers are detected successfully. But the lines are discrete and not continuous as desired. The lines are not explicitly seen. Also there should be 9 boundaries in total. So, some further improvements are needed. In Qi Yang (2010)'s paper, there is a second method that they implemented which is axial gradients alongside Canny Edge Detection. So, this second method can also be implemented to refine the discontinuous edges.

Utilizing the pretrained Med3D backbone, we have successfully validated the model's performance in transfer learning, even under constraints of limited data and computational resources. The next step is to conduct transfer learning using the extensive UK Biobank dataset - comprising approximately 80k OCT data points - in order to enhance the decoder. Then, we will train the final algorithm for our classification task, firmly believing in the model's capacity to achieve remarkable accuracy rates for Alzheimer's classification.

4 Appendix

Figure 1. Original Image

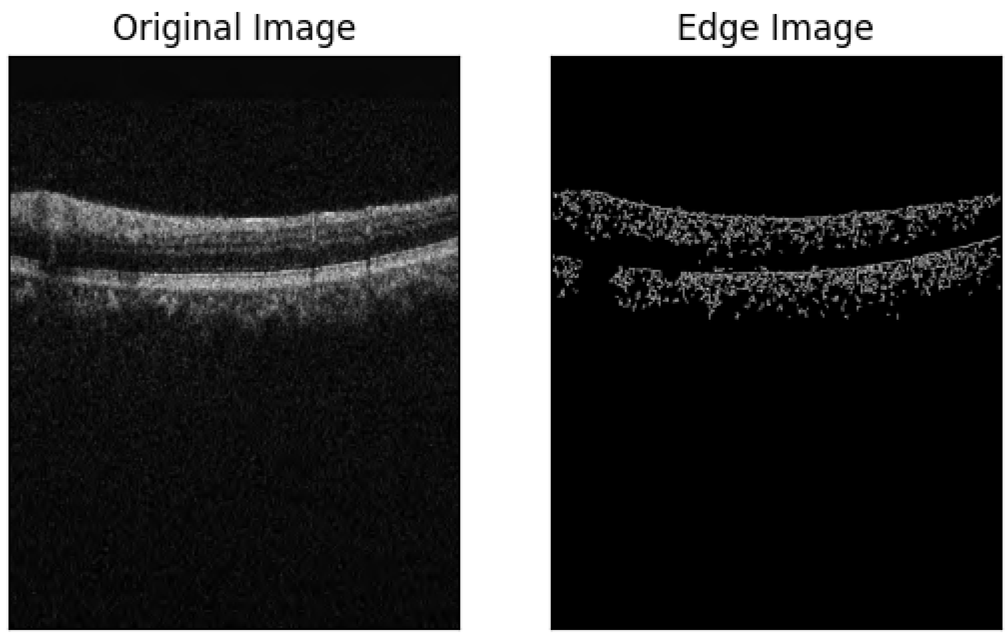

Figure 2. Canny method in cv2 library

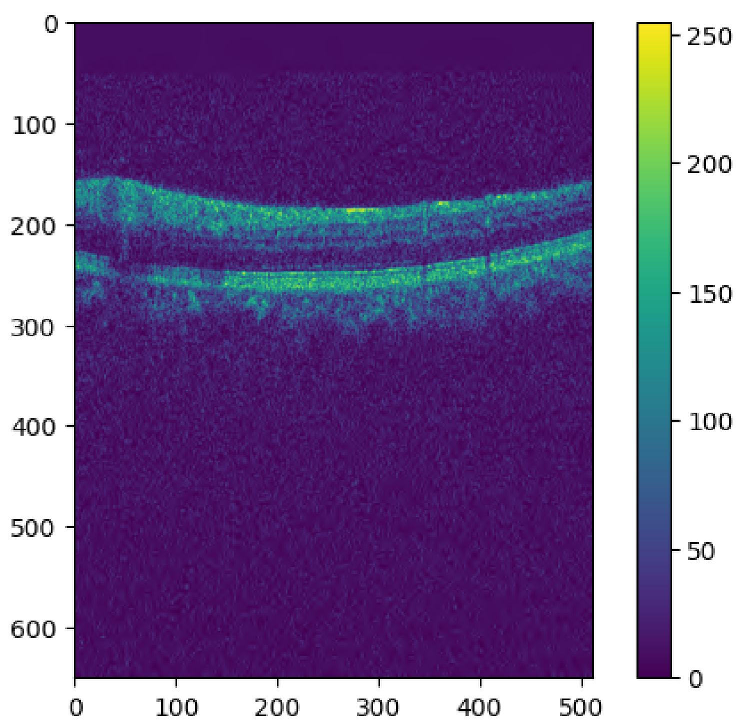

Figure 3. Colormap

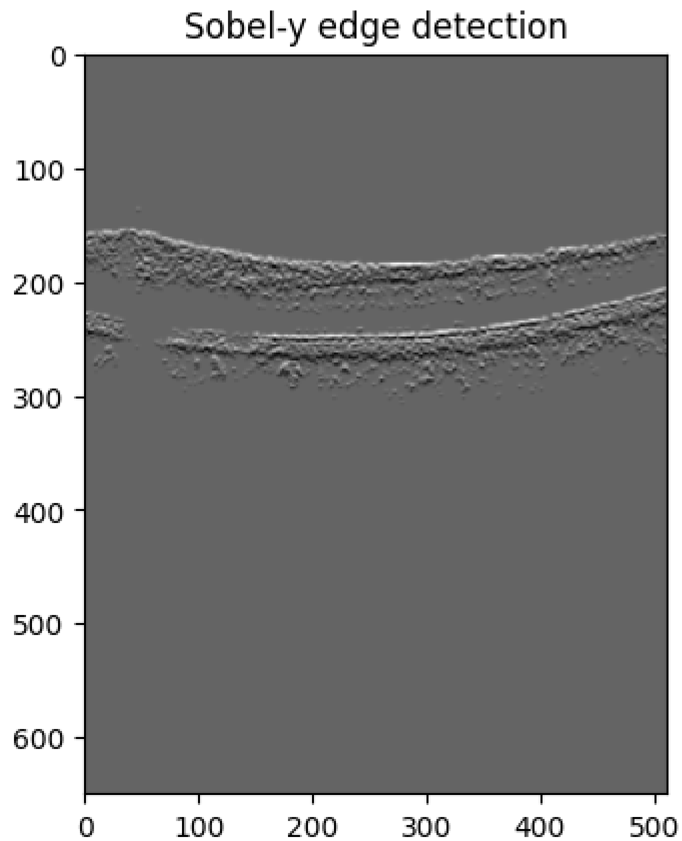

Figure 4. Horizontal Sobel Filter

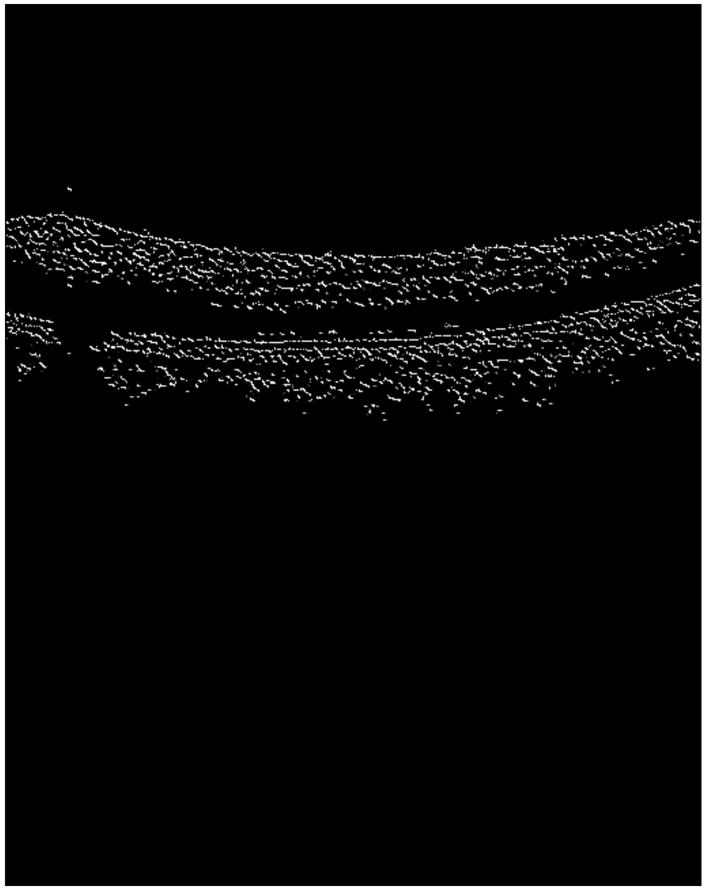

Figure 5. Non-maximum suppression

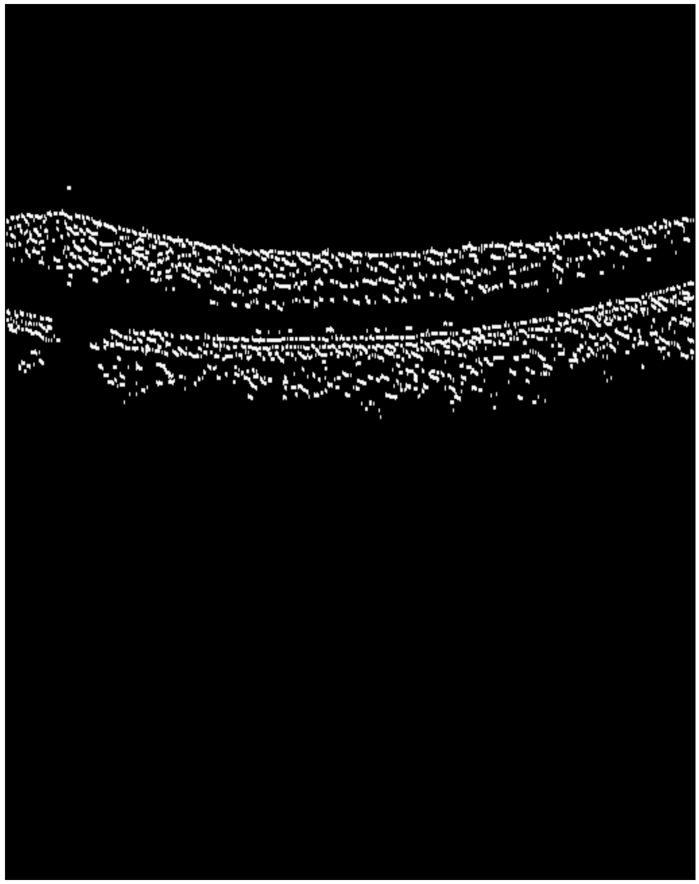

Figure 6. Thresholding & Hysteresis

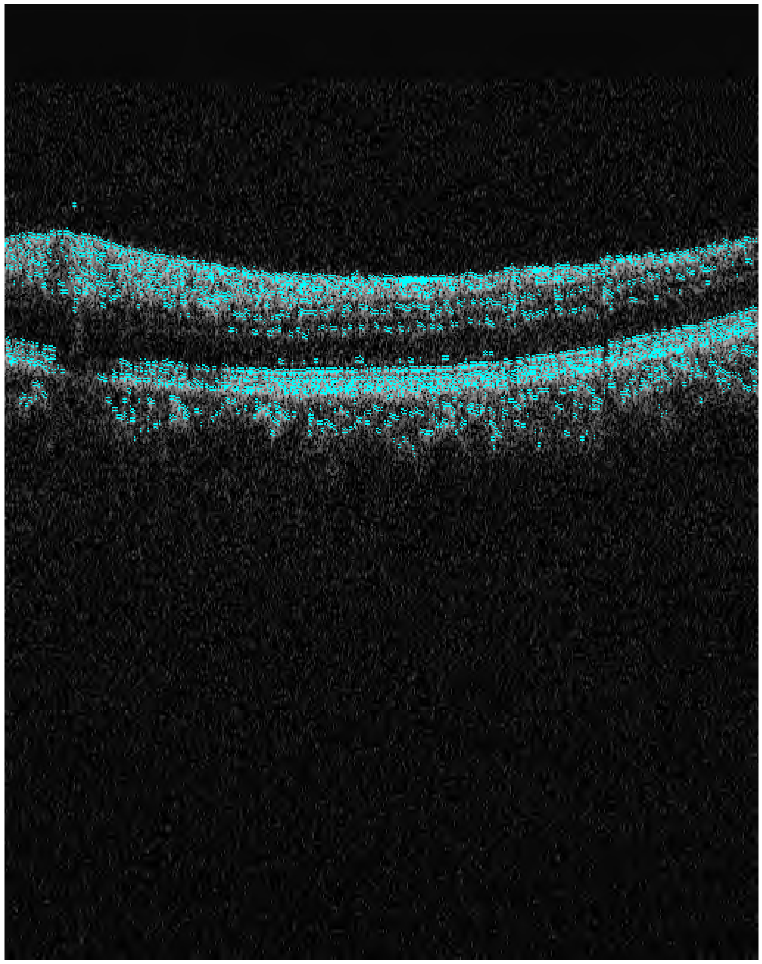

Figure 7. Final Result

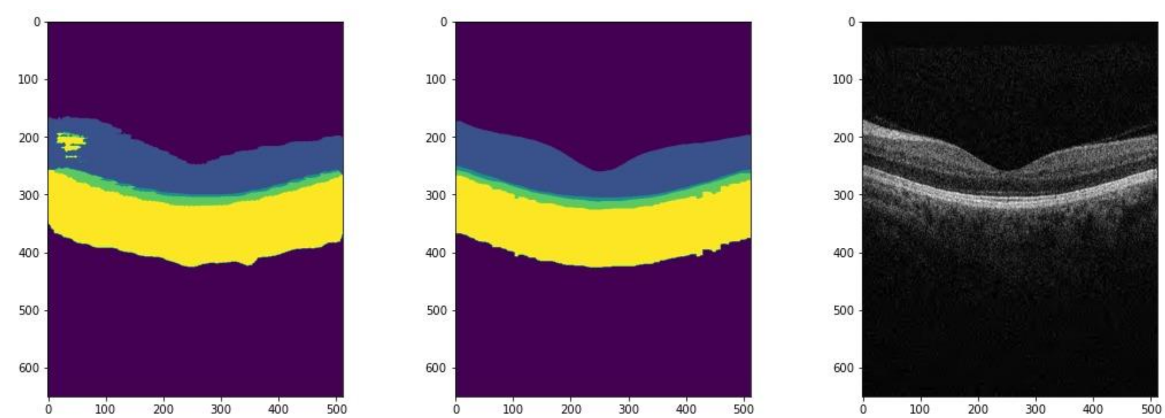

Figure 8. Segmentation Result Example 1

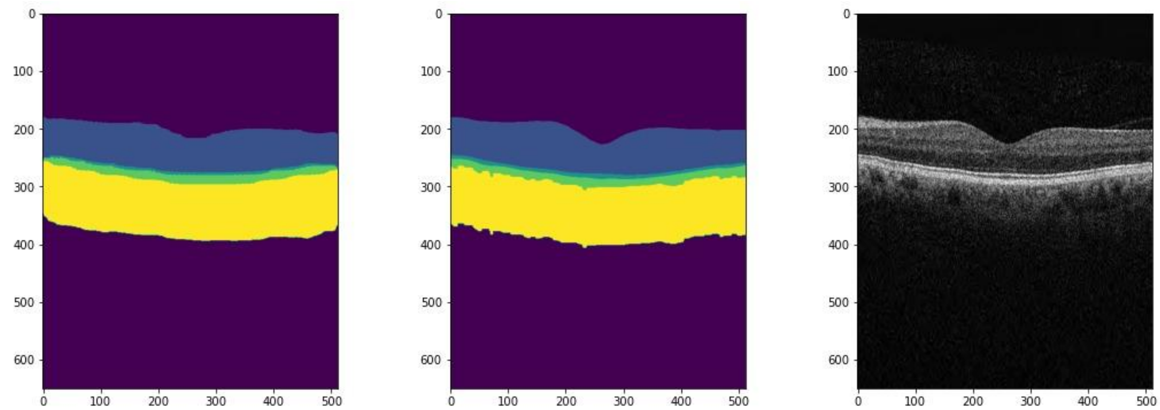

Figure 9. Segmentation Result Example 2

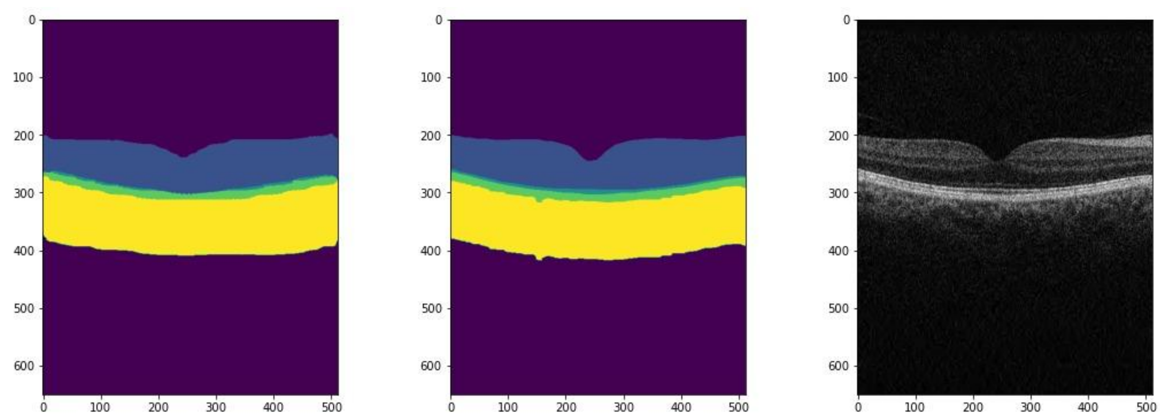

Figure 10. Segmentation Result Example 3

References

Hara, K., Kataoka, H., and Satoh, Y. (2017). Learning Spatio-Temporal Features with 3D Residual Networks for Action Recognition. *arXiv e-prints*, doi:10.48550/arXiv.1708.07632.

OpenCV:Canny Edge Detection. (n.d.). OpenCV: Canny Edge Detection. https://docs.opencv.org/3.4/da/d22/tutorial_py_canny.html

Tan, M. and Le, Q. V. (2019). EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks". *arXiv e-prints,* doi:10.48550/arXiv.1905.11946.

Turkan, Y., Tek, B. (2022). A Survey on Automated Diagnosis of Alzheimer's Disease Using Optical Coherence Tomography and Angiography. *arXiv:2209.03354v1 [eess.IV].*

Qi, Y., Reisman, C., Wang, Z., Fukuma, Y., Hangai, M., Yoshimura, N., Tomidokoro, A., Araie, M., Raza, A., Hood, D., Chan, K. (2010). Automated layer segmentation of macular OCT images using dual-scale gradient information. *Optical Society of America,* 18 (20), 2-4.

Qi, C. R., Su, H., Mo, K., Guibas, L. J. (2016). PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation. *arXiv e-prints*, 1-3. doi:10.48550/arXiv.1612.00593.

Xie, S., Girshick, R., Dollár, P., Tu, Z., and He, K. (2016). Aggregated Residual Transformations for Deep Neural Networks. *arXiv e-prints*, doi:10.48550/arXiv.1611.05431.

Chen, S., Ma, K., and Zheng, Y. (2019). Med3D: Transfer Learning for 3D Medical Image

Analysis. *arXiv e-prints,* doi.org/10.48550/arXiv.1904.00625.