Classification of Macular Degenerative Pathologies from OCT images with Speckle Noise

Brent Savage
MEng in Photonics and Optical Sciences
Duke University
Durham, NC 27708
brent.savage@duke.edu

Abstract

Age-related macular degeneration (AMD), globally, is a leading cause of blindness with several different related pathologies. The need for rapid retinal imaging and classification techniques are becoming ever important. This project attempts to demonstrate a machine learning pipeline architecture for the classification of three types of pathologies associated with AMD captured via Optical Coherence Tomography (OCT). These OCT images are simulated with three different levels of speckle noise, an inherent noise source associated with this imaging modality. These images go through speckle noise suppression via a denoising autoencoder (UNet), then passed through a classification network (VGG-11). While training on this classification data was limited, it still shows promise to be an effective classification method, given proper time. Despite this lack of training, it was determined that a medium amount of speckle noise added to the images resulted in ideal speckle suppression and subsequently classification.

1 Instruction

Age-related macular degeneration (AMD) is an eye disease that accounts for 6-9% of legal blindness globally, and is the leading cause of blindness in individuals over the age of 55 [1]. AMD, or indications of AMD, include several different pathologies. For the sake of this project, these pathologies include choroidal neovascularization (CNV), drusen, and diabetic macular edema (DME). Each of these pathologies manifests themselves in different manners, have their own symptoms, progress at different rates, and require different treatments. Given this, one could see the need for rapid retinal imaging and classification techniques for early detection and ultimately vision loss prevention.

1.1 Optical Coherence Tomography

Currently, the clinical gold standard in retinal imaging is Optical Coherence Tomography (OCT). OCT is an interferometric imaging technique that collects back-scattered light from biological samples. OCT is rapid, noninvasive, and can provide single-cell resolution of two-dimension cross-sectional images and/or three-dimensional tomographic maps . As the name suggests, this modality is highly dependent on the coherence properties of the light source used. Shown below are examples of typical OCT setups in both the time- and Fourier-domain, along with cross-section (B-scan) image [3].

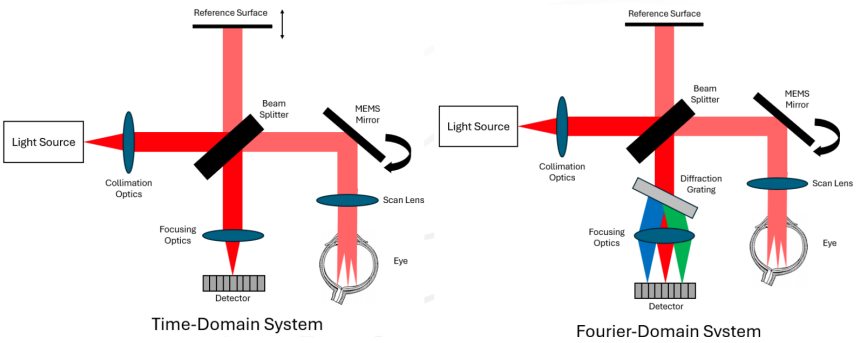

Figure 1: Typical OCT setups in the time- and Fourier-domains

1.2 Coherence and Speckle Noise

Coherence can be described as a fixed phase relationship between electromagnetic field values at either different location (spatial) or different times (temporal). This property is important for an imaging modality such as OCT as it allows for interference to be occur and cross-correlation signal to be detected. The level of coherence can be modified by changing physical characteristics to your light source, such as spatial filtering to change spatial coherence and changing bandwidth for temporal coherence. Tuning coherence allows you to reach certain imaging specifications, namely resolution. Additionally, coherence is strongly tied to the idea of speckle noise.

Speckle can be described as the interference between different parts of an object due to scattering. In an image, this manifests as random fluctuations in intensity, or what would be called speckle noise. Speckle noise can be quantified using the speckle contrast, which is defined in terms of the standard deviation and mean intensity of the image background [4].

$$C_s = \frac{\sigma_n}{\langle I \rangle} \tag{1}$$

1.3 Dataset Description

The original dataset for this project comes from data in "Large Dataset of Labeled Optical Coherence Tomography (OCT) and Chest X-Ray Images" from Daniel Kermany et al., University of California San Diego, Guangzhou Women and Children's Medical Center [3]. The OCT subset of this data includes a total of 109,309 B-scan images, including 1,000 test images and 108,309 training images. These images are broken up into four main categories, based on AMD pathology, including 37,455 choroidal neovasularization (CNV) images, 11,598 diabetic macular edema (DME) images, 8,866 drusen (DRUSEN) images, and 51,390 normal or healthy (NORMAL) images. It is important to note that these images vary in size, orientation and speckle noise level. Shown below are example images of each category. The original dataset used for this project can be found at the following link [5]:

`https://data.mendeley.com/datasets/rscbjbr9sj/2`

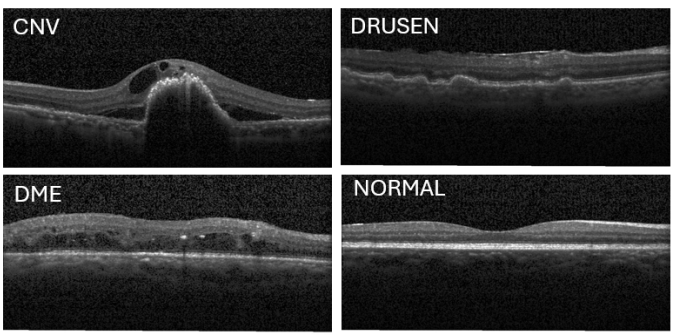

Figure 2: Four OCT image categories from the original dataset [5]

2 Related Work

Previous work has been done on classifying and segmenting OCT images of age-related macular degeneration (AMD). In this machine learning architecture, previous authors have proposed using complementary mask guided convolutional neural networks (CM-CNN) to perform classification of different AMD pathologies, along with a class-activation map guided UNet (CAM-UNet) to segment lesions associated with certain pathologies, namely CNV and drusen [2]. The author of this report is attempting to replicate certain parts of this work, specifically the classification task, along with simulating noise inherent to OCT capture, namely speckle noise, to learn optimal conditions for data acquisition.

3 Methods

The primary goal of this project is to create a machine learning architecture that can take an OCT B-scan image, simulate the inherent speckle noise associated with its measurement, suppress said noise, and ultimately classify the image based on its pathology. As will be described in the following sections, three levels of speckle noise (low, medium, and high) were simulated using random phasors. These speckled images were then cleaned up using a denoising autoencoder (U-Net) and classified into the different pathology classes. It was also determined which amount of speckle noise was optimal for this classification task.

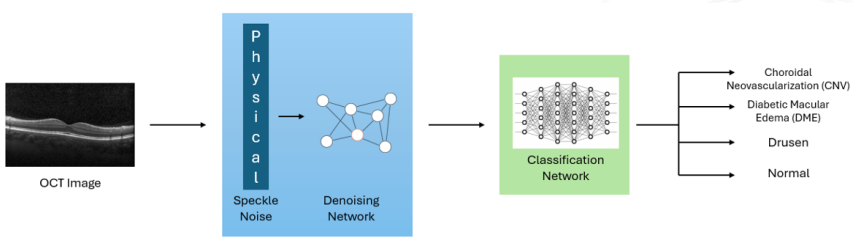

Figure 3: Machine learning architecture for this project

3.1 Resizing the Dataset

As mentioned in a previous section, the original dataset included images of varying sizes, orientations, and speckle noise levels. Before any other preprocessing, these images needed to be resized so that they could be passed into a neural network. It was determined that the images would be resized to be square images (256x256 pixels) in order to speed up computation. For this resizing process, an aspect ratio threshold of 1.5:1 was set, meaning that if the original image was too far away from a square shape as to cause large distortion or information loss during resizing, the images were discarded and not included in the new dataset. After resizing, the new dataset had the following statistics: 42,345 total images, with 41,345 training images and 1,000 test images. In terms of pathologies, this new dataset included 9,494 CNV images, 10,144 DME images, 5,371 DRUSEN images, and 16,336 NORMAL images.

3.2 Adding Speckle Noise

Once a resized dataset was obtained, varying levels of speckle noises needed to be added to these images. For this project, it was determined that speckle noise would be simulated using a random phasors method. A phasor is a complex value that is commonly used to represent electromagnetic fields, and takes the following form:

$$\text{phasor} = A \cdot e^{i\phi} \tag{2}$$

where A represents amplitdue, and ϕ represents phase. For this simulation, A had a constant value, while ϕ had a uniform distribution from - π to π. In order to simulate speckle noise in an image, a sum of N phasors is added to each pixel in the image. The exact formulation for the transformation from the original image to a speckled image is the following [6]:

$$\text{Speckled Pic} = \alpha\sqrt{\text{Original Pic}} + \text{phasorSum} \tag{3}$$

where α is a pre-factor that can be adjusted to produce different speckle noise levels, as it determines the balance between the contributions of the original image versus the random phasor sum. For this project, three different levels of speckle noise were simulated: a low level (α = 10), a medium level (α = 5), and a high level (α = 2.5). The resulting images after speckle addition are shown below.

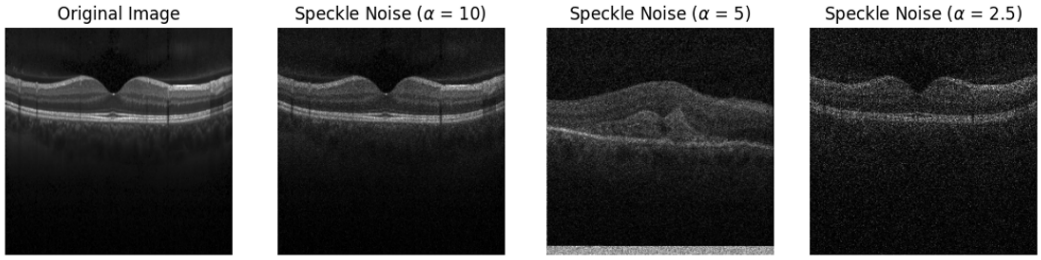

Figure 4: Original versus speckle noise images at three different noise levels

3.3 Denoising Autoencoder

Before the speckled images can be passed into a classification network, there must be an attempt to clean up said speckle noise. For this project, speckle noise suppression was done using a denoising autoencoder network (U-Net). This U-Net style architecture takes an input image, downsamples is using a series of convolutional and max pooling layers, then subsequently upsamples the image back to its original size, with the ultimate goal of outputting a less noisy image. This process can be done in one of two ways. First, the images can be fed into this network to train without labels, allowing the network to learn a reprentation of the data from a corrupted version of it. Second, noisy images, along with clean images as labels can be fed into the network for training, allowing the network to learn a series of transformations go between the corrupted data to the actual representation of said data [7]. For this project, the latter approach was taken, using the various speckle noise level data as the input into the network, with the original resized data being the labels. Shown below is the exact architecture for this denoising autoencoder, including intermediate dimensions throughout the network, and representative input and output images.

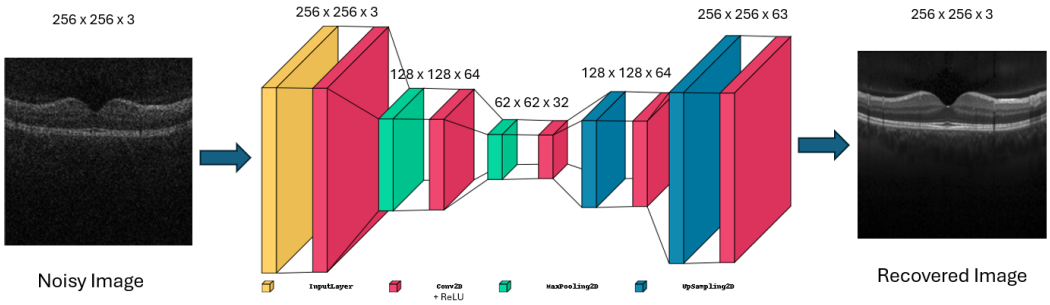

Figure 5: Denoising Autoencoder (U-Net) architecture with representative inputs and outputs

3.4 VGG-11 Classification Network

After being passed through the denoising autoencoder, recovered images are then passed through a VGG-11 network for classification. Visual geometry group, or VGG, networks are a very common convolutional neural network (CNN) architecture for large-scale multiclass classification tasks [8]. In the case of this specific type of VGG, a VGG-11 consists of 8 convolutional layers, 5 max-pooling layers, and 3 fully connected layers. The output of this network is a series of 4 integers associated with a type of class. The labels CNV was [1 0 0 0], DME was [0 1 0 0], DRUSEN was [0 0 1 0], and NORMAL was [0 0 0 1]. It is important to note that these labels were added manually by the author. Each level of speckle noise images that were recovered from the denoising autoencoder were fed into this network, with their corresponding labels for training. A visual representation of the

VGG-11 network is shown below, including intermediate dimensions throughout the network, and representative input images.

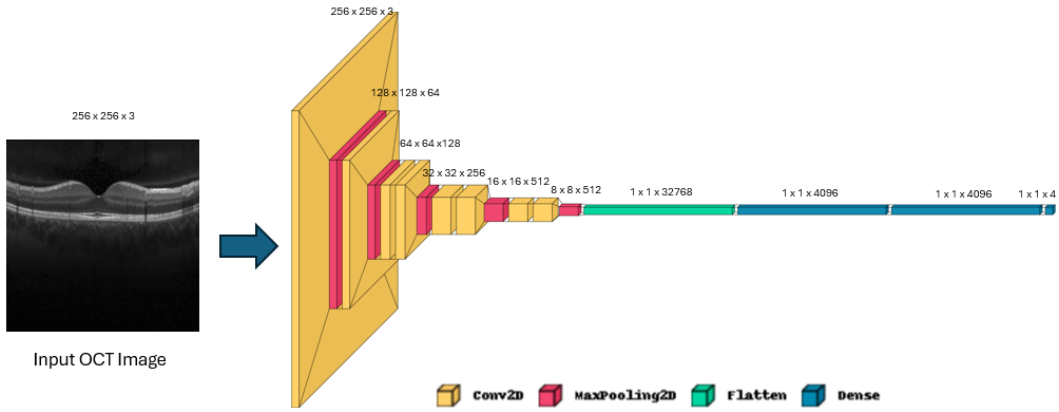

Figure 6: Classification network (VGG-11) architecture with representative inputs

4 Results

4.1 Speckle Noise Supression

The datasets at the three speckle noise levels of interest were used to train the denoising autoencoder. The autoencoder was compiled using the Adam optimizer, at a learning rate of 0.001, and a loss function of 0.001. The results of this training are shown below, with a direct comparison between the original image from the resized dataset, the image after the speckle noise was added, and the recovered image after it is passed through the autoencoder. Along with this visual representation is the associated mean squared error (MSE) values, quantifying the difference between the recovered image from the autoencoder and the original image used as a label. While the autoencoder performed fairly similarly for all three speckle noise levels, it was found that it performed best for the medium speckle noise level, as the mean squared error for this level was the lowest. Upon visual inspection of the recovered image, it can be seen that while the network successfully suppresses the random fluctuations of the speckle noise, it leaves a Gaussian-like blur across the image that could be of concern for classification as the finer details of each pathology get lost in the blur.

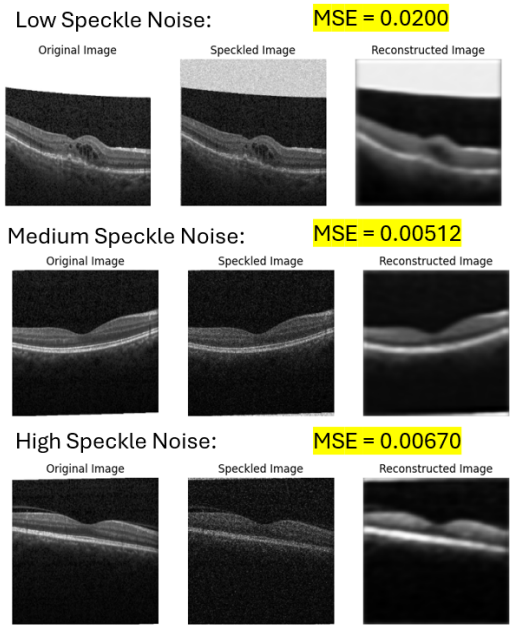

Figure 7: Reconstructed image results from denoising autoencoder

4.2 Classification Results

The recovered images, along with their associated labels were used to train the VGG-11 classification model. The model was compiled with the Adam optimizer,at a learning rate of 0.001, using a categorical cross-entropy loss function. As a disclaimer, the author experienced significant computation time and memory (RAM) issues while training this model, and therefore was only able to run 3 epochs on it. Therefore, the following results from the training of this model are fairly limited. Shown below is a comparative bar chart of the maximum accuracy values acheived by the trained model for each speckle noise level. The accuracy values are very low due to the level of epochs ran, and not representative of the capabilities of this classification architecture. With more time and epochs, the author expects much higher accuracy values and more conclusive data to make conclusions on how speckle noise level affects this accuracy. However, with the limited data, it can be concluded that once again the medium speckle noise level was optimal in terms of classification accuracy.

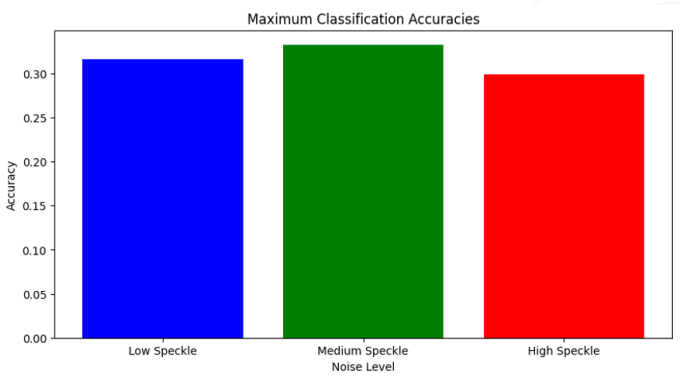

Figure 8: Maximum classification accuracy from each speckle noise level

5 Discussion

Overall, the combination of the denoising autencoder and the VGG-11 classification network provided an architecture for multi-class classification of pathologies from OCT images corrupted with speckle noise. In both the case of speckle noise suppression/image reconstruction and classification, the medium speckle noise level performed the best. However, this result should be taken with an appropriate level of consideration due to the lack of training that was able to be achieved on the classification network. Further training of this network would be needed to make more definitive conclusions.

In the future, the author would suggest many modifications to the process used to complete this project. First and foremost, the collection and dataset of OCT images should include background images under their corresponding acquisition environments to better quantify the speckle noise of the image based on the speckle contrast parameter. The author did attempt to create a speckle contrast quantifier, but found it extremely difficult to consistently isolate the background from the retina across the entire dataset for accurate calculations. Additionally, the author suggests exploring other options for cleaning up the speckle noise (e.g. mean filter, etc.), as to not incur the Gaussian-like blur that could have affected classification results. Optimization of the speckle noise (α value) could be integrated into a CNN network itself as a trainable parameter for more accurate and complete physical layer optimization. Lastly, the author suggests testing other types of CNN architectures for the classification network (e.g. ResNet, MobileNet, etc.)

Acknowledgments

Thank you to Prof. Horstmeyer, and TAs Amey Chaware and Xi Yang for their continued assistance and support throughout this semester.

References

[1] Fleckenstein, M., Keenan, T.D.L., Guymer, R.H. et al, “Age-related macular degeneration," Nat Rev Dis Primers **7**, 31 (2021). https://doi.org/10.1038/s41572-021-00265-2

[2] S. Diao et al, “Classification and segmentation of OCT images for age-related macular degeneration based on dual guidance networks," Biomedical Signal Processing and Control, vol. 84, p. 104810 (2023). https://doi.org/10.1016/j.bspc.2023.104810

[3] Nabila Eladawi et al, “Optical coherence tomography: A review," pp. 191-221 (2020). https://doi.org/10.1016/b978-0-12-817440-1.00007-3.

[4] T. Yoneyama and Y. Sakamoto, “Speckle control for electro-holographic display using high-brightness yellow phosphor light source in projector," Optical engineering, vol. 62, no. 08 (2023). https://doi.org/10.1117/1.oe.62.8.083103.

[5] D. Kermany, K. Zhang, and M. Goldbaum, “Labeled Optical Coherence Tomography (OCT) and Chest X-Ray Images for Classification," data.mendeley.com, vol. 2, (2018). https://doi.org/10.17632/rscbjbr9sj.2.

[6] B. Savage “ECE 542 HW2," (2023). https://github.com/brentsavage/Classification-of-macular-degenerative-pathologies-from-OCT-images-with-speckle-noise/blob/main/Resources/ECE 542 HW%232.pdf

[7] K. Bajaj, D. K. Singh, and Mohd. A. Ansari, “Autoencoders Based Deep Learner for Image Denoising," Procedia Computer Science, vol. 171, pp. 1535–1541 (2020). https://doi.org/10.1016/j.procs.2020.04.164.

[8] “VGG-16 | CNN model," GeeksforGeeks (2020). https://www.geeksforgeeks.org/vgg-16-cnn-model/