AFRL-RH-FS-TR-2019-0015

Machine Learning for Laser Lesions

Zachary D. Stoecker-Sylvia, M.S.
Edward Pier, Ph.D
Oceanit

Adam R. Boretsky, Ph.D.
SAIC

Joel N. Bixler, Ph.D.
711th Human Effectiveness Wing
Airman Systems Directorate
Bioeffects Division
Optical Radiation Branch

June 2019

Final Report – 26 April 2018 - 6 June 2019

DESTRUCTION NOTICE – Destroy by any method that will prevent disclosure of contents or reconstruction of this document.

Air Force Research Laboratory
711th Human Performance Wing
Airman Systems Directorate
Bioeffects Division
Optical Radiation Bioeffects Branch
JBSA Fort Sam Houston, Texas 78234

NOTICE AND SIGNATURE PAGE

Using Government drawings, specifications, or other data included in this document for any purpose other than Government procurement does not in any way obligate the U.S. Government. The fact that the Government formulated or supplied the drawings, specifications, or other data does not license the holder or any other person or corporation; or convey any rights or permission to manufacture, use, or sell any patented invention that may relate to them.

This report was cleared for public release by the 88th ABW Public Affairs Office and is available to the general public, including foreign nationals. Copies may be obtained from the Defense Technical Information Center (DTIC) (http://www.dtic.mil).

"Machine Learning for Laser Lesions"

(AFRL-RH-FS-TR- 2019 - 0015) has been reviewed and is approved for publication in accordance with assigned distribution statement.

Digitally signed by SHORTER.PATRICK.D.1023156390
Date: 2019.09.04 10:15:33 -05'00'

PATRICK SHORTER, Major, USAF
Chief, Optical Radiation Bioeffects Branch

MILLER.STEPHANIE.A.1230536283

Digitally signed by MILLER.STEPHANIE.A.1230536283
Date: 2019.09.23 13:07:35 -05'00'

STEPHANIE A. MILLER, DR-IV, DAF
Chief, Bioeffects Division
Airman Systems Directorate
711th Human Performance Wing
Air Force Research Laboratory

This report is published in the interest of scientific and technical information exchange, and its publication does not constitute an official position of the U.S. Government.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
13-06-2019	Final Technical Report	26 April 2018 – 6 June 2019

4. TITLE AND SUBTITLE

Machine Learning for Laser Lesions

5a. CONTRACT NUMBER
FA8650-14-D-6519

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Zachary D. Stoecker-Sylvia, Edward Pier, Adam R. Boretsky, Joel N. Bixler

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER
H0RY

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Air Force Research Laboratory
711th Human Performance Wing
Airman Systems Directorate
Bioeffects Division
Optical Radiation Bioeffects
JBSA, Fort Sam Houston, Texas 78234
4141 Petroleum Rd
Fort Sam Houston, Texas 78234

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

711th Human Effectiveness Wing
Airman Systems Directorate
Bioeffects Division
Optical Radiation Branch
JBSA, Fort Sam Houston, Texas 78234

10. SPONSOR/MONITOR'S ACRONYM(S)
711 HPW/RHDO

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
AFRL-RH-FS-TR-2019-0015

12. DISTRIBUTION / AVAILABILITY STATEMENT

Distribution Statement A. Approved for public release; distribution is unlimited. TSRL-PA-19-0200. The opinions expressed on this document, electronic or otherwise, are solely those of the author(s). They do not represent an endorsement by or the views of the United States Air Force, the Department of Defense, or the United States Government.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Oceanit, using its RetinaView software and data collected by the U.S. Air Force developed and trained convolutional neural networks to detect and classify laser lesion injuries in three imaging modalities: Fundus imagery, hyperspectral imagery, and OCT imagery. Networks using the commonly collected Fundus and OCT imagery can detect lesions with 97% accuracy, providing a viable tool for non-expert clinicians with no experience in laser lesion injuries to detect them without additional imaging equipment. Further experiments show it is possible to predict the age of a given lesion, allow for the potential to tie an injury to a specific engagement or event.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
Unclassified					Dr. Joel Bixler
a. REPORT	b. ABSTRACT	c. THIS PAGE	U	28	19b. TELEPHONE NUMBER *(include area code)*
U	U	U			(210) 539-8172

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. 239.18

This Page Intentionally Left Blank

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF EQUATIONS

1 SUMMARY

The objective of this research effort is to take advantage of recent advances in machine learning and multi-modal imaging technologies in order to create a means of detecting and classifying retinal injuries that is superior in speed and accuracy to human observation. The primary aim of this effort is to detect and classify retinal images of laser injures using novel application of machine learning algorithms.

Currently, detection of laser injuries relies on retinal imaging via fundus photography and optical coherence tomography, or direct observation using clinical methods such as direct or indirect ophthalmoscopy or slit lamp examination with fundus lenses. Many clinicians are unfamiliar with the clinical diagnosis of retinal lesions and must seek the assistance of researchers with expertise in this type of injury. Consulting with researchers causes a time delay in diagnosis that in turn delays treatment.

The use of machine learning for feature recognition and computer vision has become a topic of great interest over the past decade, with applications ranging from facial recognition to object detection for self-driving cars. To achieve this, a machine learning model is first presented with a set of data designed to train the model to recognize certain features. The model is then given a second set of data designed to test and evaluate its effectiveness at detecting the features for which it was trained. Recently, machine learning has been applied to medical imaging with goals such as the identification of tumors in CT scans, but to date no work has been done on the development of machine learning algorithms for detecting laser lesions in retinal images.

2 BACKGROUND AND PURPOSE

2.1 Laser Lesion Injuries

Currently, the identification of laser injury for both clinical and research purposes is subjective, and clinicians usually lack the training to diagnose retinal laser injury due to the rarity of such damage. Current research methods assess laser damage by establishing consensus among researchers who directly observe tissue samples. This methodology can result in significant variation among observers, particularly in the case of studies that aim to determine the thresholds for minimally visible lesions. As more powerful laser systems become commonplace, accurate assessment of retinal laser damage and identification of the laser source will become a critical challenge to enabling adequate protection. The purpose of the present work was to determine if machine learning is better able to detect and classify laser damage.

2.2 Datasets

In 2014, the U.S. Air Force 711th Human Performance Wing, Optical Radiation Bioeffects branch (711HPW/RHDO) conducted retinal imaging in the rhesus macaque to study the appearance of suprathreshold laser lesions in the macula from both photothermal (532 nm, 100 ms) and photomechanical (532 nm, 9 ns) injuries using five different imaging systems [2]. Three of these are clinically approved: Heidelberg Spectralis spectral domain (SD) optical coherence tomography (OCT) scanning laser ophthalmoscopy SD-OCT-SLO, Heidelberg HRT3 confocal scanning laser

ophthalmoscope (cSLO), and Topcon Fundus camera, and two are experimental: multispectral using a fundus camera and hyperspectral using a modified PSI Inc. line scanning laser ophthalmoscope (LSLO). The macula of each eye of 16 subjects was given a 5x5 grid of laser lesions for a total of 800 lesions. Half of the subjects received photothermal lesions and half photomechanical. The eyes were imaged before exposure, immediately after exposure, then one, three and ten days after exposure with a variety of imaging techniques.

This was known as the BSELL (Biomarkers and Sensitive Evaluation of Laser Lesions) study [2]. We extracted training, test, and validation data from BSELL for the work described here. We considered only the fundus, hyperspectral, and OCT imagery.

The animals in the BSELL study were procured, maintained, and used in accordance with the Federal Animal Welfare Act, "Guide for the Care and Use of Laboratory Animals," prepared by the Institute of Laboratory Animal Resources National Research Council, nd DoD Regulation 40-33 Secnavinst 3900.38C AFMAN 40-401(1) DARPAINST 18 USUHSINST 3203 “The Care and Use of Laboratory animals in DOD Programs.” The 711th Human Performance Wing, Fort Sam Houston TX has been fully accredited by the Association for Assessment and Accreditation of Laboratory Animal Care, international (AAALAC) since 1967.

2.3 RetinaView

RetinaView [3] is a collection of software for organizing, viewing, and manipulating collections of retina images. It was developed by Oceanit under an SBIR contract with the US Air Force. RetinaView has many useful features, including:

- A Data Manager that does database queries using an intuitive graphical interface. This allows the user to organize files into hierarchical directory-like structures that can be rearranged at will.
- A Warper tool for registering images using thin plate spline warps. The registration process is semi-automated, requiring manual placement and coarse alignment of control points and providing automated refinement of their positions.
- A viewer tool that lets the user adjust brightness and contrast and mark Regions of Interest (ROIs) on images. It can use Warper registrations to transform ROIs from one image to another.
- An OCT Slicer that lets the user slice OCT data volumes along straight or curved surfaces shaped like the inner-limiting membrane (ILM) or retinal pigmented epithelial (RPE) layers. This makes it simple to extract images of blood vessels for registration or laser lesions at particular depths. The tool renders these slices in two or three dimensions.
- A Hyperspectral viewer that can register the images from a BSELL hyperspectral scan, create false color images and extract spectra from regions of interest.
- A Combiner that can average sets of registered images and create videos from individual frames.

As part of a previous effort, Oceanit used RetinaView to register nearly every image in the BSELL dataset and to mark ROIs for each laser lesion in the fundus images. In the current effort we used

components of RetinaView to preprocess the data for the machine learning work that we describe below.

2.4 Terminology and Practice for Machine Learning with Convolutional Neural Networks

The broad machine learning technology of artificial neural networks has been used for a variety of learning tasks due to their flexible learning potential and optimized training algorithms. Neural networks learn by iteratively optimizing weights on connections between a large network of vertices (neurons). These weights are combined with an activation function at each neuron to create a mathematical function whose result turns inputs (in the case of imagery a two- or three-dimensional array of image channel values) into either a real number associated with a single regression output or a vector of classification output neurons. For a classification task, the output neurons correspond to the network's judgement on the presence or absence of the class. In instances where classes are incompatible, such as "lesion" or "no-lesion", the output vector should contain a one for the correct class and zeros for any incorrect classes. By comparing the results from one of these computations to the desired value, the network can calculate a small modification to each connection weight to more correctly classify that instance. By performing this task on many inputs and attempting to optimize for a representative population of examples, the network learns a general-purpose solution.

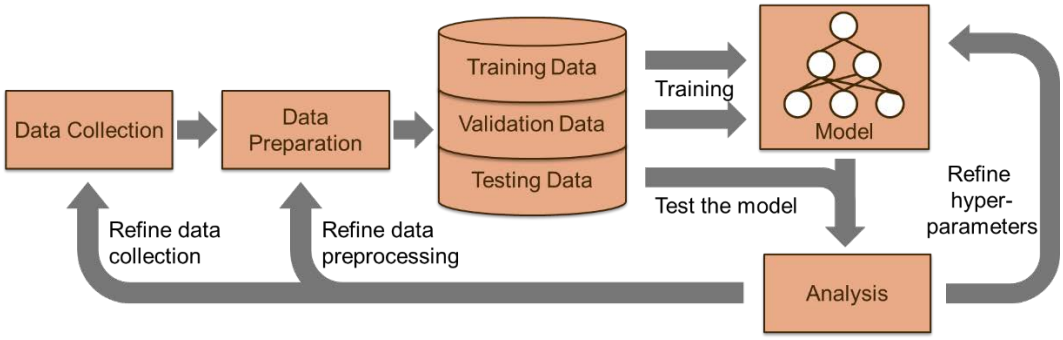

Figure 1. A broad overview of training a neural network. Further data collection during the project's period was not feasible and due to the limited data, validation and testing datasets were merged into one for the work.

For detecting objects in imagery, as was done for this project, a form of these networks called "convolutional neural networks" is the primary method of detection. These neural networks, instead of having fully connected layers (i.e. every neuron on one layer connected to every neuron on the next), use layers of convolutional filters that scan across the previous 2D or 3D layer in smaller segments and compute a single output. This scan is repeated by shifting in one or more axes by a number of pixels, called the convolutional stride. The weights of these filters are thus applied and updated simultaneously for their applicability to all segments across the image while keeping the relationships they find localized to the filter's size.

Additionally, convolutional layers are frequently followed by a pooling layer that uses a similar convolutional style of processing, but instead of performing a calculation based on relationships between the contained pixels, it propagates the highest value. These pooling layers effectively discard weak values and focus later processing on whether any of the pixels provide a strong response to the filter.

Finally, convolutional networks used for classification flatten their multi-dimensional convolutional filter results into a one-dimensional vector and use a standard dense net to produce a final classification.

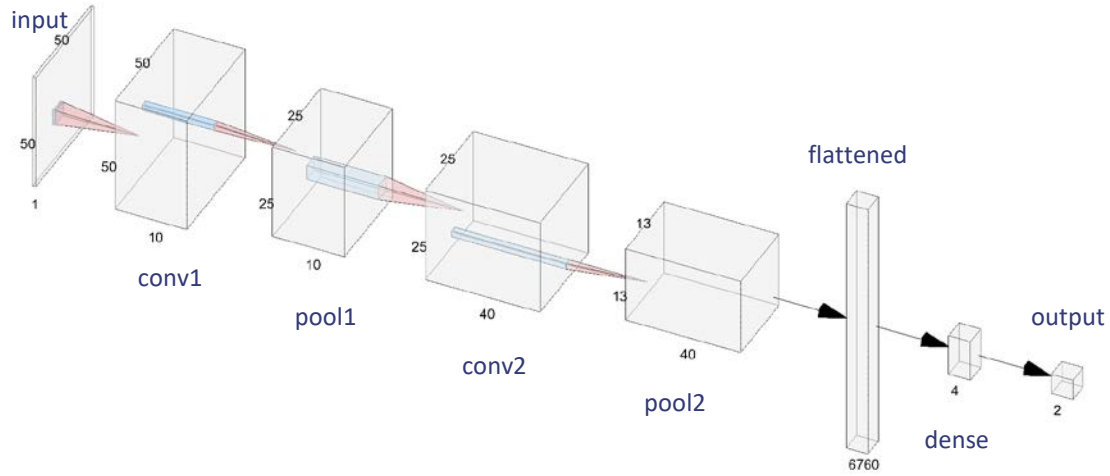

Figure 2. An example convolutional network diagram. The square face of the gray cuboids represent pixels in the image and processed virtual pixels for internal layers. Their third dimension represents the number of separate filters applied across those pixels. Blue cuboids within represent the size of the convolutional filter. Finally, the results are flattened and classified by a dense layer. In most diagrams the flattening layer is omitted.

These networks have a variety of parameters, referred to as hyperparameters that greatly influence the success of the training process. The most prominent of these are the decisions about network shape (e.g the size of convolutional filters or dense layers), the functions used for calculating neuron activations (the activation function) and weight modifications (the loss function), and how much data are passed through the network between weight updates. Selecting correct values is critical to creating a network that learns a generalizable classification function rather than failing to classify or creating a classifier that is highly tuned to the training data rather than the general traits of data in the domain, known as overfitting.

Overfitting can be recognized by observing the results on a validation set kept separately and not used for training. This set serves as a stand-in for novel data that a trained network would be exposed to in the wild and when the results of the loss function (representing error in classification) on the validation data diverges from the same function on the training data, the network is

overfitting. This can be corrected by making a smaller network, but Oceanit also applied batch normalization and dropout to help combat the effects.

2.5 Oceanit's Machine Learning Pipeline

Oceanit developed a data processing pipeline to classify lesions from raw collected imagery using its RetinaView software, custom Python code, and Google's TensorFlow machine learning framework. Early image processing and labeling were performed using previously developed tools from RetinaView to create labeled imagery with metadata about the imagery's source experiments. This metadata, imagery, and labeling, in the form of lesion mask images, were then ingested into Oceanit's custom Python code and used to generate labeled positive (lesion) and negative (non-lesion) examples with their associated metadata. From there the data were loaded into memory to be selected and manipulated to create datasets for different experiments before passing into a convolutional neural network for high-performance training on a GPU. Finally, Oceanit developed a variety of functions in order to analyze the training results in order to tune the network's hyperparameters.

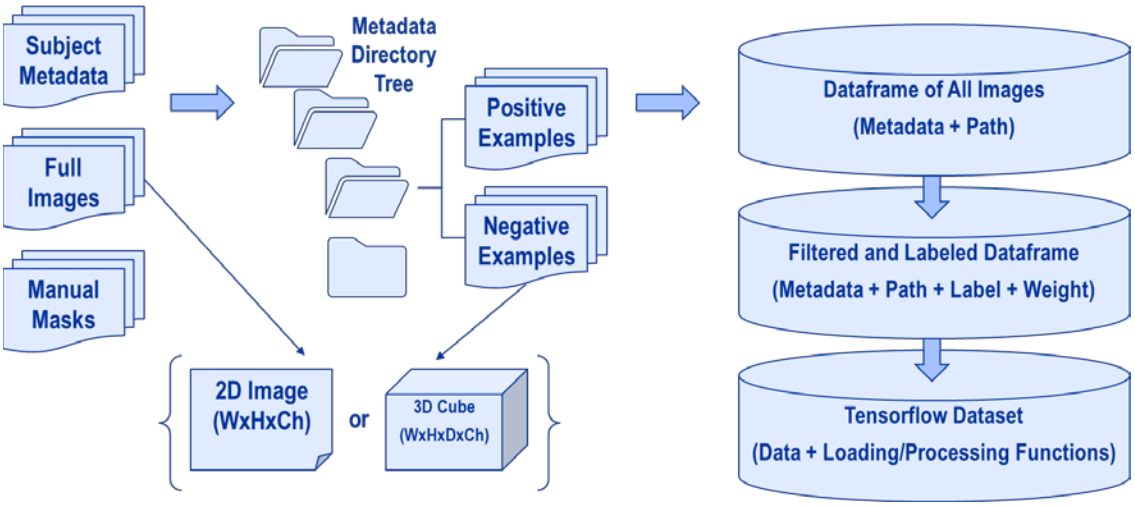

Figure 3. Overall structure of Oceanit's data processing pipeline. The resulting TensorFlow dataset is optimized for massive parallel data processing by Oceanit's deep convolutional neural network.

3 MACHINE LEARNING ON THE FUNDUS IMAGERY

3.1 Data Preparation

The BSELL fundus images were taken with a Topcon TCR-50x fundus camera, modified by attaching a Cannon EOS Rebel camera for digital photography. The images were stored in JPEG format with 3456 x 5184 pixels. They come in three resolutions, but we considered only the ones with the narrowest field of view. Multiple images were taken during each data collection session.

We used RetinaView to register every fundus image in BSELL. We began by registering sets of images taken on a particular day. We then averaged these sets to produce a clean image with reduced noise. Finally, we marked regions of interest on every averaged image that was taken on day 1 after laser exposure. We chose this day as it tended to show the lesions most prominently. All the ROIs were marked by the same person, who did not have medical training.

Next, we performed background subtraction on each raw (unaveraged) image to increase contrast and compensate for variations in exposure and illumination across the field of view. We smoothed the image with a 200 pixel wide gaussian and then subtracted this from the original. We then multiplied the difference by 6, as this was the largest value that would not result in saturation. We saved each background subtracted image in a lossless format with 8 bits per color channel. Note that this procedure does not presuppose knowledge about the locations of the laser lesions.

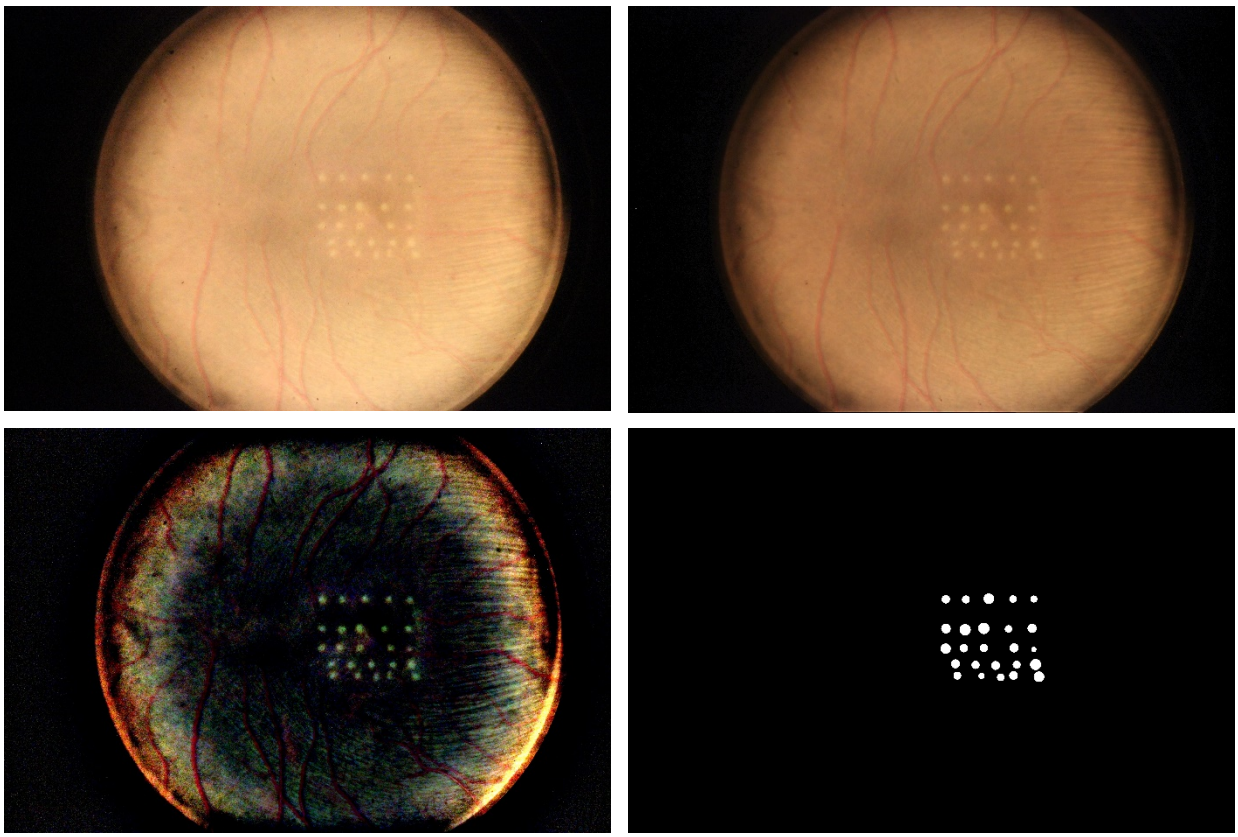

Figure 4. A raw fundus image from BSELL (top left), an averaged image (top right), a background subtracted image (bottom left), and a lesion mask (bottom right). These images are from the left eye (OS) of BSELL subject 28080 on day 1 after laser exposure.

We registered the averaged images from different days. This allowed us to transform the ROIs to every other fundus image for the same eye. We created two sets of masks for each image. The first set masked all the lesions and the second masked only a single lesion in each image. In the end we had 380 images with 8,978 labeled lesions.

3.1.1 Thumbnail Generation

The processed Fundus images were broken into 150x150 pixel thumbnail images to provide positive (lesion) and negative (non-lesion) examples that were independent of the injury's location within the entire eye. These thumbnails, along with their prediction metadata, were then processed by the network during training. When applied to a novel image for testing, the trained network scanned along the image by creating equal sized clippings while shifting horizontally or vertically between each clipping by 50 pixels. The 50 pixel stride ensures all lesions will have at least one clipping where the entire lesion is present.

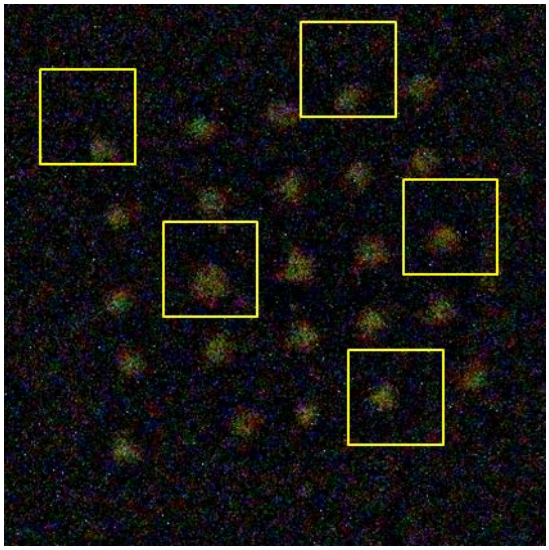

Figure 5. Examples of randomly shifted thumbnail bounds used to create off-center thumbnails around lesions. 5 examples are used to reduce visual clutter in the example, but all lesions would have thumbnails created if possible.

In order to allow for the unknown alignment of a lesion within a clipping on a novel image, the training thumbnails were created with a random horizontal and vertical shift such that the thumbnail still contained the entire marked area plus a 5 pixel padding, but was not always in the same location. These random shifts, where possible, were created to avoid overlapping any other labeled lesions. If a random shift that did not overlap other lesions could not be created after 100 tries, a centered box was returned. Negative examples were created by generating random bounding boxes that did not overlap with marked lesion areas.

For the Fundus imagery, 6 positive examples were created for each lesion and 150 negative examples are created per image. As each image contains 25 lesions, this generated roughly equal amounts of both positive and negative examples, though in some images there were lesions that could not be labeled as well as lesions that could not fit entirely within a 150x150 pixel region and thus did not have positive examples created.

In total 57,162 examples were created, with 47.5% being positive examples. 66% of the data represented photothermal injuries, with the rest being photomechanical. Figure 6 shows the distribution of generated positive examples by lesion age for each injury type.

The data were split between training and validation datasets at a 9 to 1 ratio. To prevent evaluation on training data, the validation data were generated such that an individual lesion would only occur in one or the other dataset. Finally, before processing each thumbnail was resized from 150x150 pixels to 50x50 pixels to reduce unnecessary dimensionality from the higher precision imagery in the network.

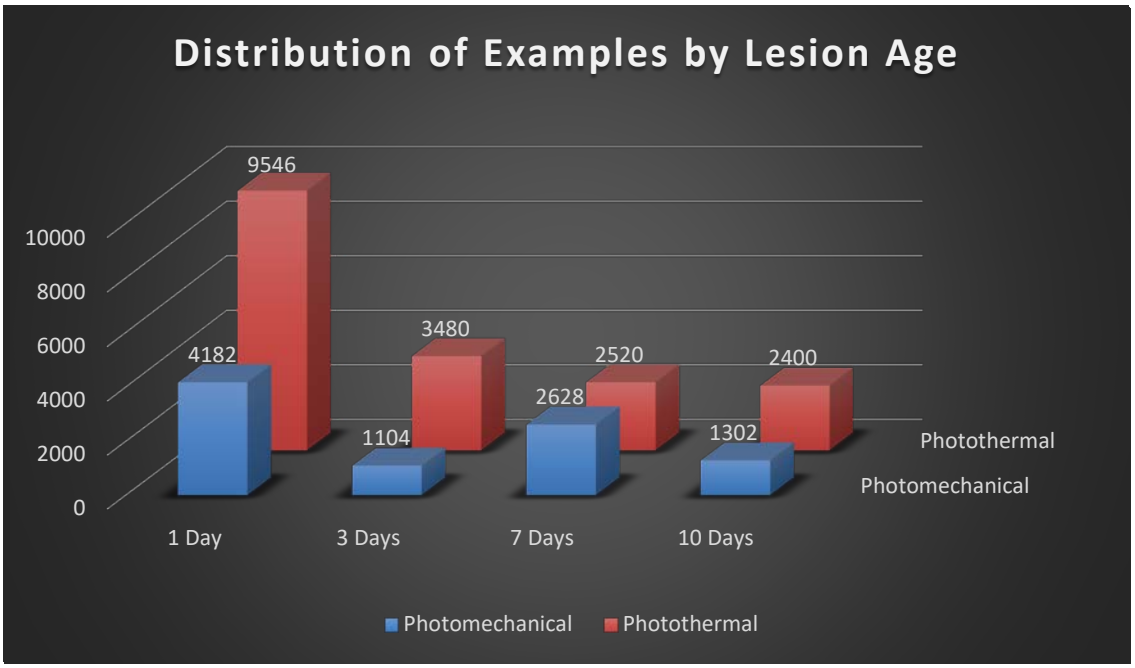

Figure 6. Distribution of Fundus dataset example thumbnails by age and injury type.

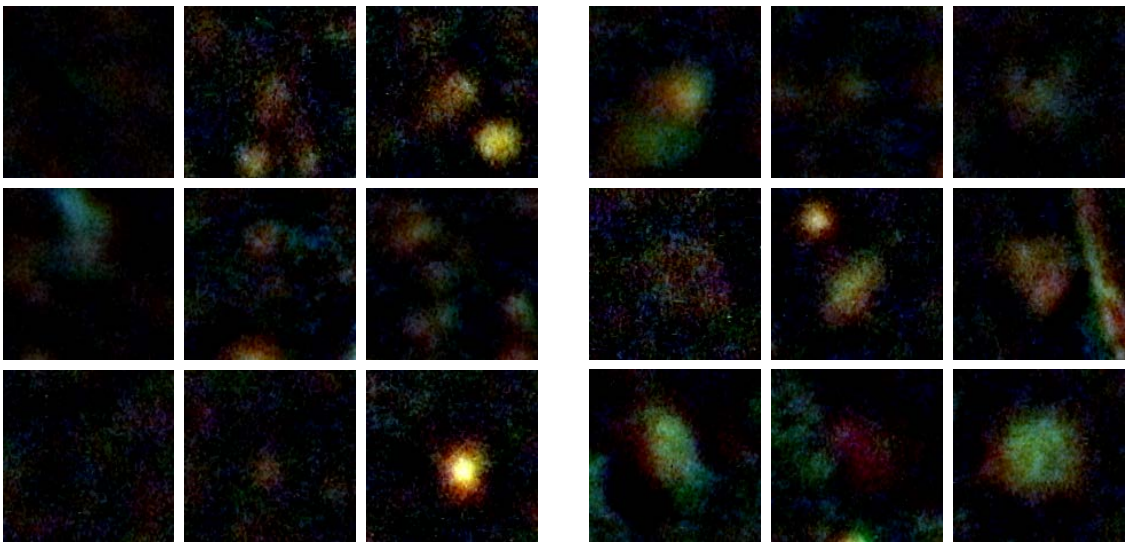

Figure 7. A sample of non-lesion (left) and lesion (right) thumbnails.

3.2 Detecting Lesions

Machine learning model training primarily focused on the detection of lesions, though other limited efforts attempted to predict the age of a detected lesion. A framework for training artificial neural networks to predict different metadata labels was created to support these tasks. Each network was composed of 9 layers (including input and output layers) and trained for 100,000 epochs using a mini-batch training method with batches of 32 examples trained simultaneously.

To maximize the value of the dataset and maintain comparability between later experiments in other imaging modalities, separate networks were trained to recognize photothermal and photomechanical lesions. Earlier tests trained on examples of both injury types simultaneously and showed similar results as the final split experiments for the Fundus imagery.

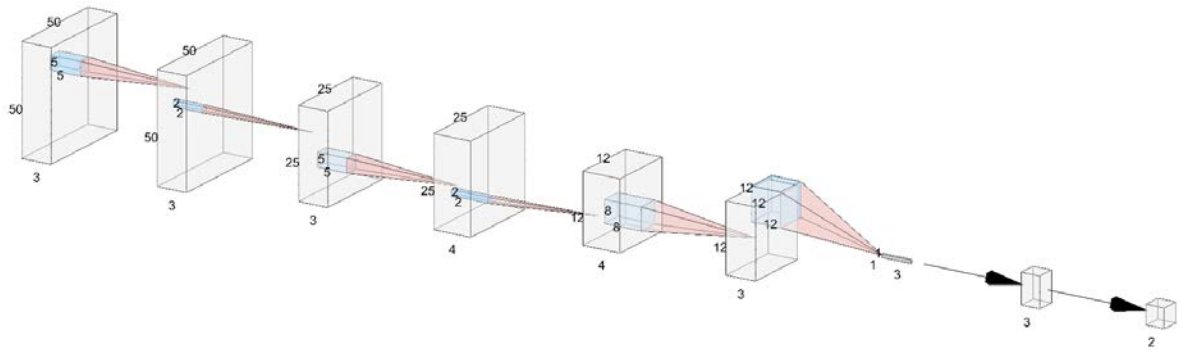

Figure 8. The network consisted of an input layer (layer 1), convolution filter layers (layers 2, 4, 6), max pooling layers (layers 3, 5, 7), fully connected layer (layer 8), and a one-hot class output layer (9).

3.2.1 Model Training Results

Models trained on examples of photothermal lesions could consistently reach 96% accuracy when tested against the validation dataset. After reaching their lowest point, the training showed little rise in loss, indicating a stable model. Figure 9 shows the validation accuracy and validation loss throughout training for several training runs on photothermal data. Due to the stochastic nature of training, multiple runs were included to demonstrate the stability of the solution.

The models for photomechanical lesions showed a slight decrease in accuracy, and a greater sensitivity to network size parameters. Networks trained using 4 filters in the middle layers showed a propensity to overfit (i.e., learning features present in the training dataset but not in the validation dataset). This is generally the result of a network with too much flexibility to learn the generalized features and/or not enough data to encourage a broader generalization. It is expected that this difference may be due to the more varied lesions caused by photomechanical injuries and exacerbated by the smaller datasets available for photomechanical training (about 50% of the photothermal dataset size). To correct for this overfitting, the number of filters in the middle convolutional layers was reduced from 4 to 3. Figure 10 shows the validation results during training for the photomechanical models.

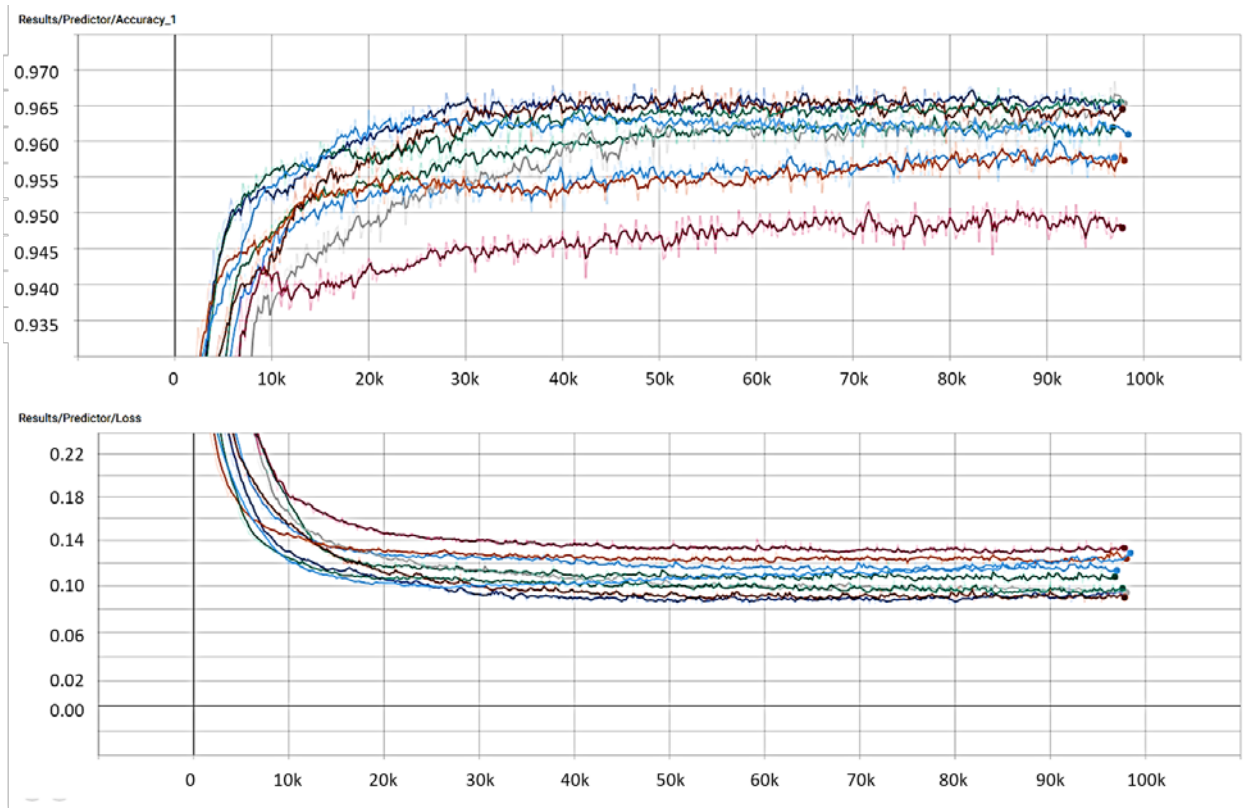

Figure 9. Validation accuracy (Y-axis, top) and validation loss (Y-axis, bottom) vs. training epoch (X-axis) for multiple training runs on photothermal images. Accuracy ranged from 94.7% to 96.6%, with most results clustered at 96%.

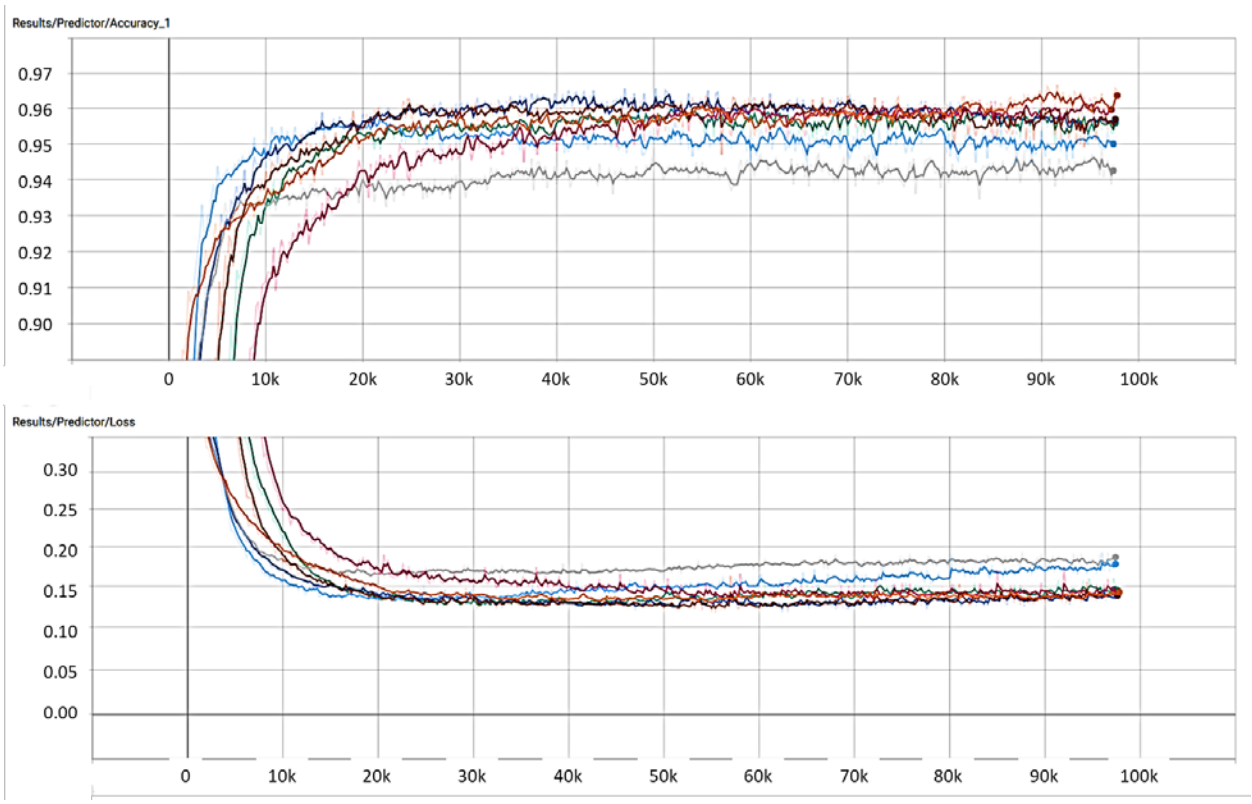

Figure 10. Validation accuracy (Y-axis, top) and validation loss (Y-axis, bottom) vs. training epoch (X-axis) for multiple training runs on photomechanical images. Accuracy ranged from 94.4% to 96.5%, with most results clustered at 95.5%. The blue and gray lines show moderate increase in validation loss over time, which suggest overfitting, and correspond to lower overall accuracy.

3.2.2 Confidence Thresholding to Reduce False Positives

While roughly equivalent class populations aid in training a discriminator, in application it would be expected that far more patches without lesions are tested than patches with lesions. It is thus desirable for a technician to be able to control the frequency of false-positive declarations (non-lesions detected as lesions) while avoiding false negatives (undetected lesions). To accomplish this, a measure of confidence for predictions was added to the processing.

$$conf_{positive} = \frac{V_{positive} - \min(V)}{V_{positive} + V_{negative} - 2 \cdot \min(V)} \quad (1)$$

Equation 1. Confidence estimation equation. V represents the output value for a class. V can be positive or negative with no fixed bounds. The lowest V in a batch is subtracted to create positive results which are then compared proportionally to

create a value between 0.5 and 1.0. As each batch will contain some easily identified negative patch, min(V) represents a relatively stable baseline.

The confidence metric can then be compared to a supplied threshold to reduce detection rate and reduce false detections. Since this confidence threshold can be applied after prediction, a user can actively modify it after processing to view predictions at different confidence levels.

Examples of a network with a 96% accuracy applied to an entire image with no confidence threshold and with a minimum confidence threshold of 0.65 can be seen in Figure 11. Even with the 96% accuracy, as each image contains 7,119 clips (using a stride of 50 pixels) significant numbers of false positives are likely. These false positives are generally low confidence predictions however, so the thresholded result eliminates most while eliminating only the lowest confidence true positives.

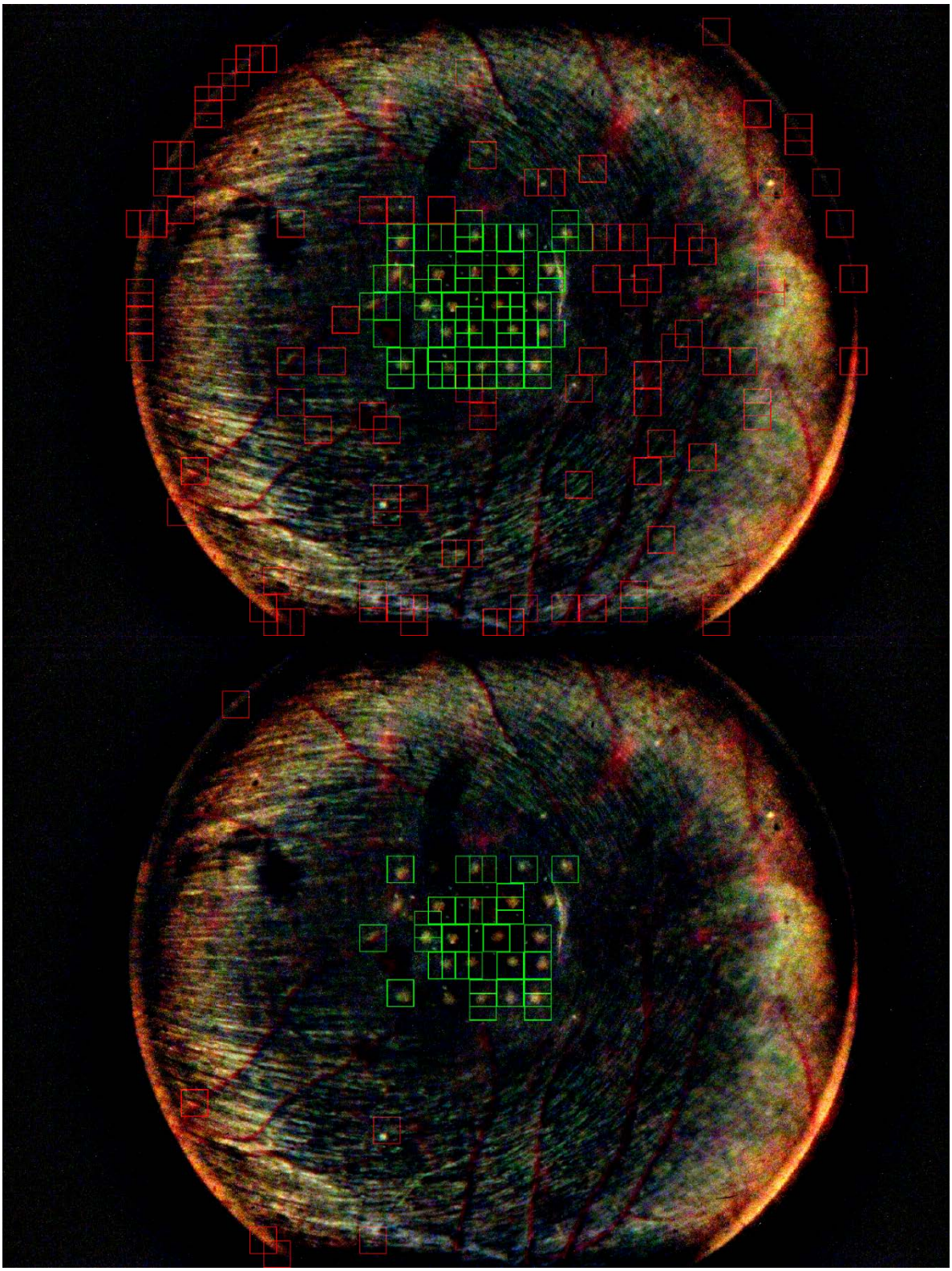

Figure 11. True positives (green) and false positives (red) with no confidence thresholding (top) and with a 0.65 threshold (bottom).

3.3 Predicting Lesion Age

A further experiment was performed to attempt to predict the age of a detected lesion. In this case, all training examples were positive (lesion) examples, and the network was instead trained to predict the age of a lesion as a discrete class. Two predictions were generated: predicting the exact number of days since injury (1, 3, 7, or 10) and predicting whether the lesion was new (1-3 days old) or old (7-10 days old). Due to the smaller dataset and more focused goal, only one convolutional layer and one max pooling layer were used.

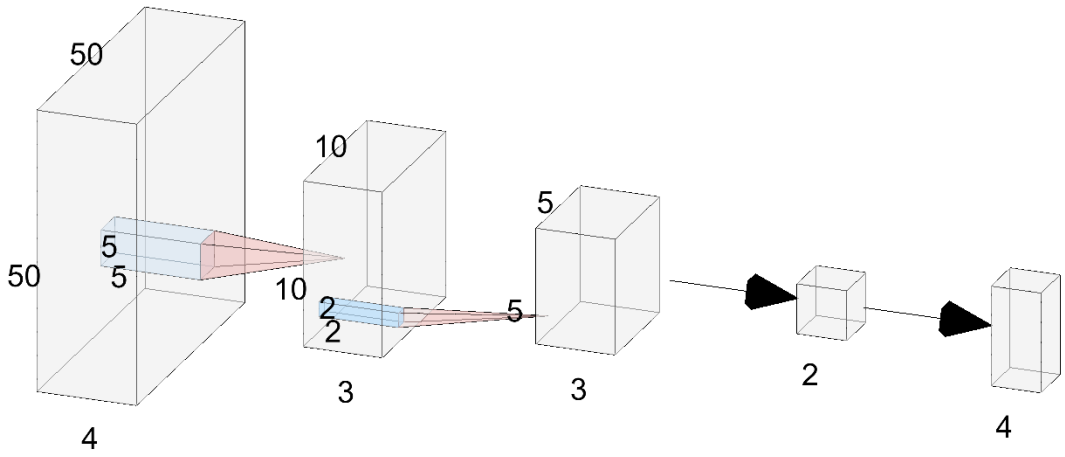

Figure 12. Network architecture for predicting the age of a lesion in Fundus imagery.

Lesion age prediction proved to be a difficult task, with a maximum accuracy of 86.0% for predicting whether a photothermal lesion was recent or over a week old, followed by the same type of prediction for photomechanical lesions (74.2%), prediction of exact day for photothermal lesions (69.0%), and prediction of exact day for photomechanical lesions (54.5%). This result is to be expected given the inherent variability in damage and healing but may provide some benefit in identifying the source of an injury.

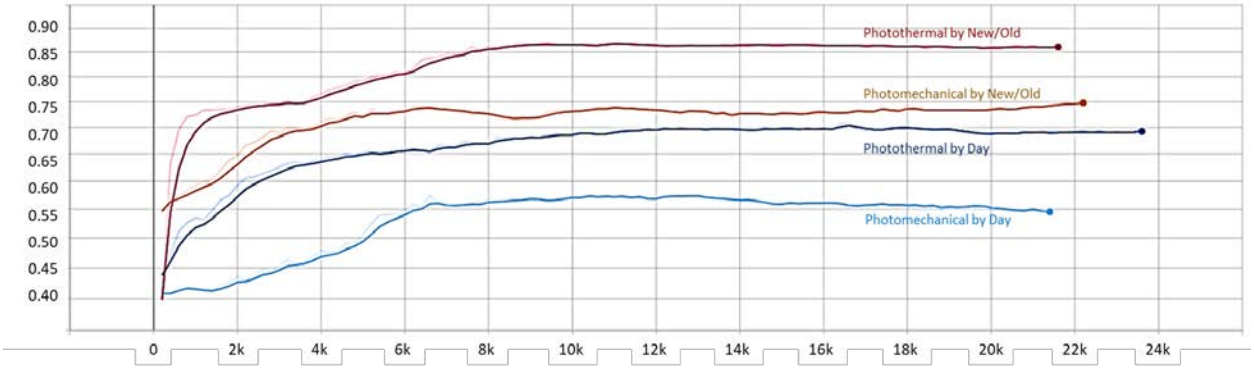

Figure 13. Validation accuracy (Y-axis) vs. epoch of training for networks predicting lesion age. Photothermal and photomechanical lesions were trained separately, and each were trained to predict the exact age of a lesion or to predict whether the injury new (1-3 days) or old (7-10 days).

4 MACHINE LEARNING ON THE HYPERSPECTRAL IMAGERY

4.1 Data Preparation

The BSELL hyperspectral data were acquired with a modified LSLO instrument from PSI Inc. The optics were replaced to image across a broad spectrum and the light source was replaced with a tunable one. The light source could operate in two modes. One scanned between 500 and 700 nm, and the other scanned from 760 to 840 nm, both in 5 nm steps. Several images were taken at each wavelength.

A single scan took a significant amount of time to complete, so it is common for there to be eye movement between images. To compensate we used VIPA to register the images in each scan using phase correlation. This is a Fourier transform based technique that is fully automated and robust to noise. We then averaged images with the same wavelength and assembled the averages into a data cube with two spatial dimensions and one spectral dimension.

From the data cube we created color images that mapped wavelengths to colors with a weighting meant to mimic the light response of human visual perception. We also created grayscale images averaged over the entire scan. We registered these combined images with the averaged fundus images that we described above. We made a subjective judgement as to whether to register the color or grayscale image, based on which showed greater contrast in the blood vessels and other features. Finally, we used these registrations to transform the ROIs labeled on the fundus images to the coordinate system of each data cube. In the end we had 55 hyperspectral cubes and 1,333 labeled lesions.

These full-eye hyperspectral cubes were then split into visible and infrared scans based on the wavelengths present. As some cubes included wavelengths not present in others in the same group, each cube was reduced to include only the wavelengths present throughout the entire group.

Extracting positive and negative thumbnails was achieved in a similar manner to the Fundus imagery, with lesions being randomly shifted within the final thumbnail (cube). For each lesion,

we attempted to create 4 different thumbnails that would not overlap neighboring lesions. In total, 889 photothermal lesion examples and 1,145 photomechanical lesion examples were available in the visible dataset, and 1,550 photothermal lesion examples, and 1,005 photomechanical lesion examples were available in the infrared dataset. Thumbnails were 30x30 pixels to match the lower resolution imagery.

Table 1. Hyperspectral lesion examples by injury type and spectrum represented.

	Visible (500-700 nm)	Infrared (760-840 nm)
Photothermal	889	1550
Photomechanical	1145	1005

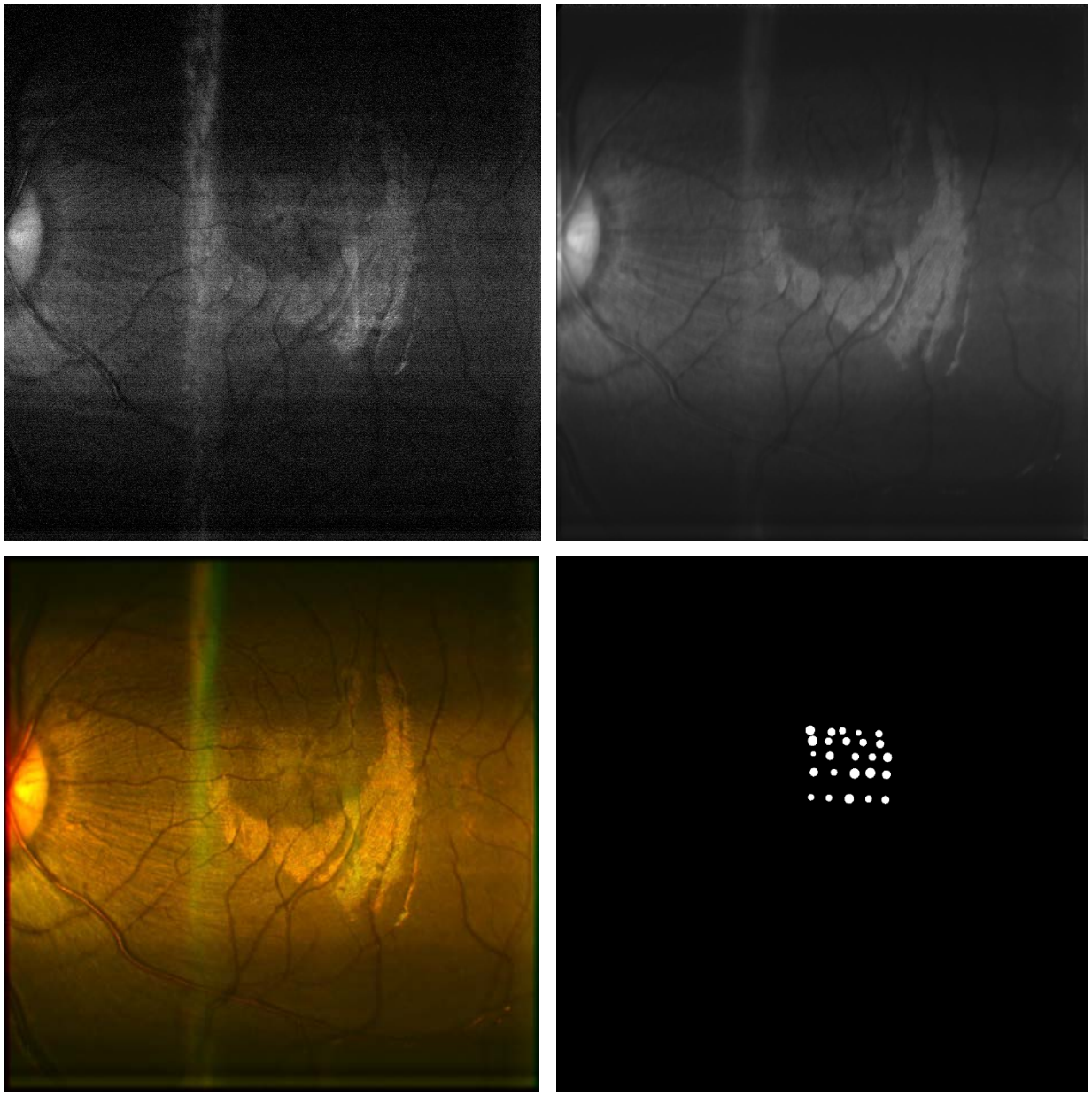

Figure 14. A single 515 nm frame from a hyperspectral scan (top left), an average over the entire scan (top right), a color image mapping wavelength to color (bottom left), and a lesion mask transformed from the fundus imagery (bottom right). These data are from the left (OS) eye of BSELL subject 28080 on day 1 after laser exposure.

4.2 Detecting Lesions

Initial experiments used both photothermal and photomechanical lesions to train a model for detecting lesions in a hyperspectral cube. These initial experiments yielded worse accuracy than

using the Fundus imagery. We expect that this is due to both issues with the data quality and the size of the feature space (i.e. the high number of channels). We therefore curtailed this work in order to concentrate available project time on OCT imagery. Our initial test accuracies ranged from 76.2% to 86.7%.

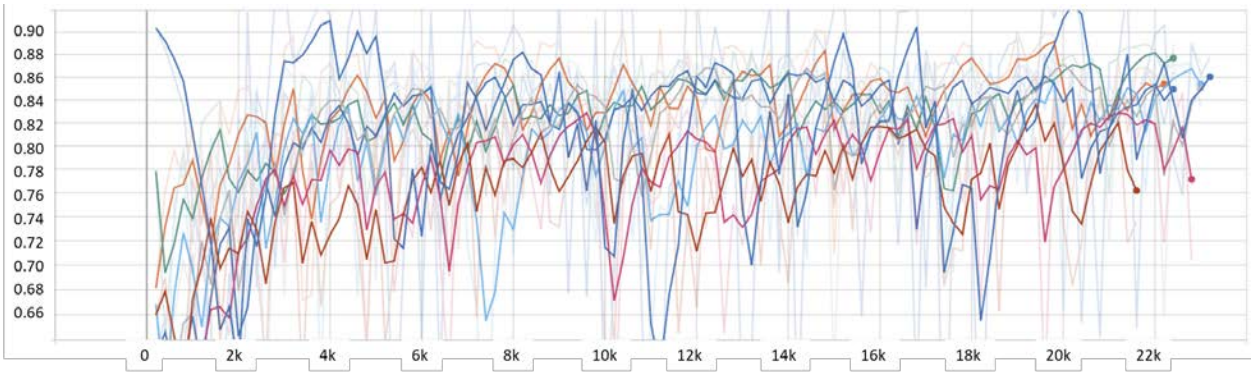

Figure 15. Validation accuracy (Y-axis) vs. training epoch (X-axis) for a series of models with different parameters. The significant instability in the training results likely stems from the large number of channels for the limited dataset and issues with the imagery.

5 MACHINE LEARNING ON THE OCT IMAGERY

5.1 Data Preparation

The BSELL OCT data were taken with a Heidelberg Spectralis SD-OCT-cSLO unit. Although BSELL includes near infrared and autoflourescence cSLO images taken with this instrument, we only examined the spectral domain OCT data volumes.

RetinaView can slice OCT data volumes along curved surfaces that parallel the ILM or RPE. For the shape of these surfaces we use the layer segmentation data extracted by the Heidelberg instrument and stored in the ".vol" data files. For our first attempt to label the lesions in the OCT scans, we extracted en face images of the blood vessels in each scan by shifting the ILM downward. We then registered these images with the corresponding averaged day 1 fundus images.

When we transformed the lesion ROIs marked on the fundus images to the coordinates of the OCT data, we discovered that the fundus-marked ROIs were consistently larger than the lesions in the OCT data. To compare them we extracted slices parallel to the RPE that passed through the lesions. We selected a layer as close to the RPE as possible that showed the lesions clearly. We then hand labeled the lesion ROIs directly on the OCT slices.

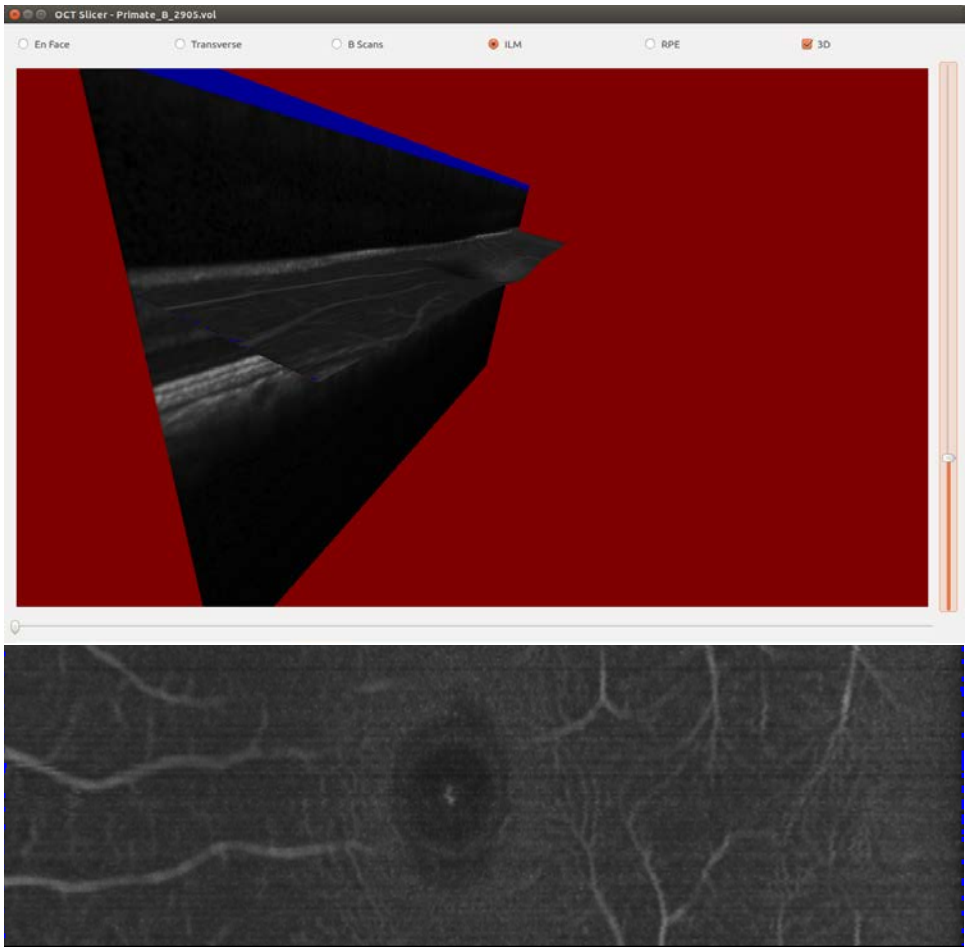

Figure 16. A 3D rendering of a curved surface slice through the blood vessels (top) and a 2D image of this slice. We used these slices to register the OCT and fundus imagery.

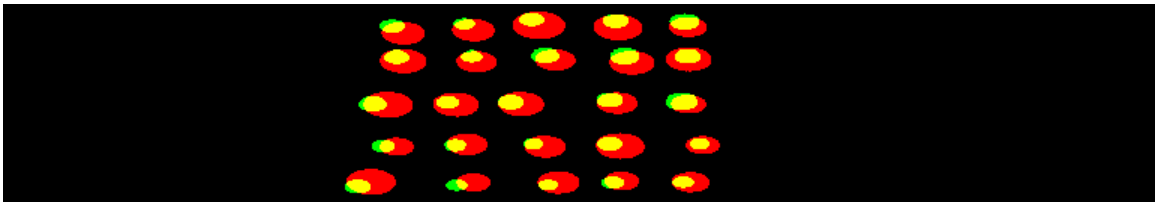

Figure 17. An example comparison of fundus-labeled lesions (red) and OCT-labeled lesions (green). Where the two overlap they appear yellow. We found that the Fundus-labeled ROIs were consistently larger than the OCT-labeled ROIs. Their positions generally coincide. The position discrepancies seen here might be due to transform errors. The ROIs are not circular because the OCT pixels are not square.

We therefore used the lesion ROIs labeled directly on the OCT data for the rest of our analysis. We extracted flat *en face* slices through the OCT data volumes and created individual lesion masks, as we did for the other modalities.

Oceanit then assembled the en face slices into 3-dimensional cuboids. As with other modalities, each lesion mask is randomly centered within a 25x25 pixel square in the *en face* dimensions. The depth slices used are the top 300 slices for all thumbnails. Figure 18 demonstrates the area used to create a thumbnail cuboid.

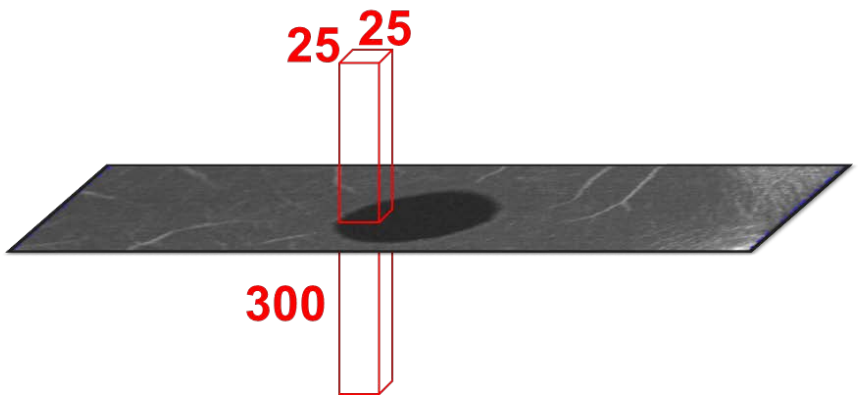

Figure 18. OCT thumbnails are 3-dimensional cuboids with a 25x25 pixel box randomly centered around a lesion and extending 300 slices from the front (top) to the back (bottom) of the eye.

5.2 Detecting Lesions

While the two previous modalities focused on 2-dimensional imagery (albeit many-channeled for the hyperspectral images), the B-Scan imagery presents 3-dimensional features that require a network structure that incorporates them into the learning process. To support this, Oceanit modified its neural network model to use convolutional layers that step through not just the width and height dimensions, but also the depth dimension. These convolutional layers thus have exponentially more hyperparameters, as both the filter dimensions, the stride, and the max pooling layers can be modified along the third axis.

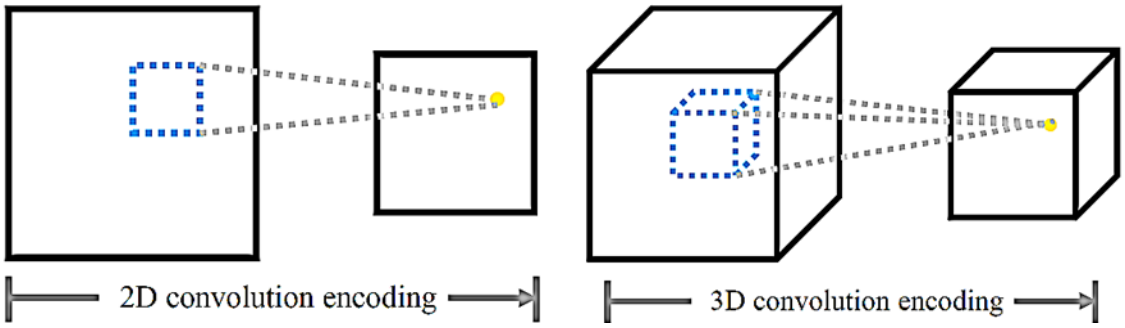

Figure 19. 2-dimensional (left) convolutional filters are repeated along the height and width dimensions, while 3-dimensional convolutional filters repeat along height, width, and depth. This increases the complexity of network hyperparameter tuning.

While photothermal and photomechanical lesions have more differences in OCT imagery (shown in Figure 20), the models were developed with largely similar hyperparameters, as their components occur in similar spaces with similar relationships. Each lesion can be roughly described to contain a stalk that is largely immobile in the two dimensions in the *en face* view and a head that may be shifted radially in those dimensions (thus can differ from depth slice to depth slice in one or both *en face* dimensions).

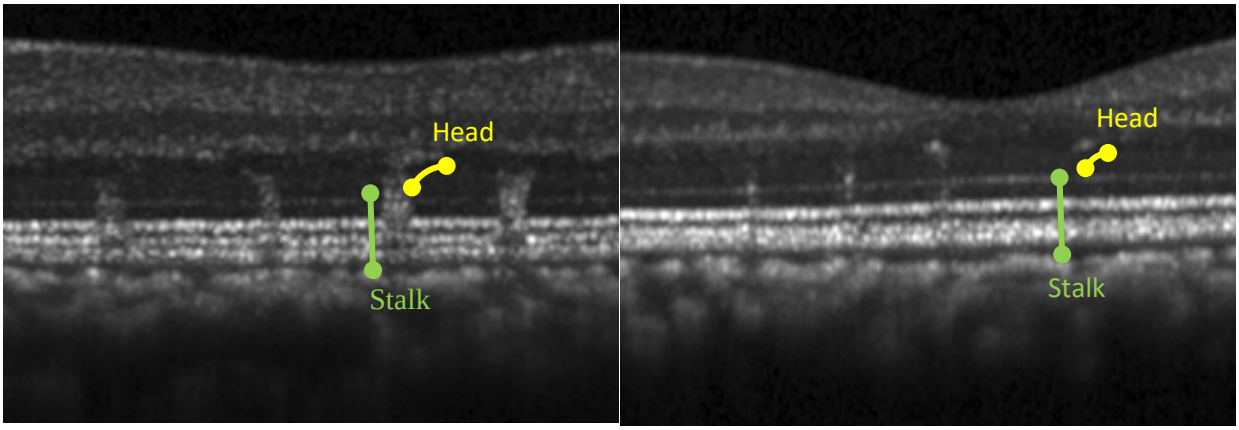

Figure 20. "Side view" comparison of photothermal (left) and photomechanical (right) lesions with references to each type's "stalk" and "head" feature.

The neural network models were designed to take advantage of this understanding of the primary component features of a lesion. Figure 21 shows a diagram of the 3D convolutional network. To express the third dimension a different network diagramming style was required.

- Convolutional layer 1 finds circular shapes in heads and columns
- Convolutional layer 2 connects shapes together in stacks and diagonal transitions
- Convolutional layer 3 connects columns and heads together

  - The photothermal network uses 8 filters here, while the photomechanical network uses 10.
- Convolutional layer 4 filters for matches in 25x25x80 patches
- The dense layers convert convolutional results into lesion/no-lesion classifications

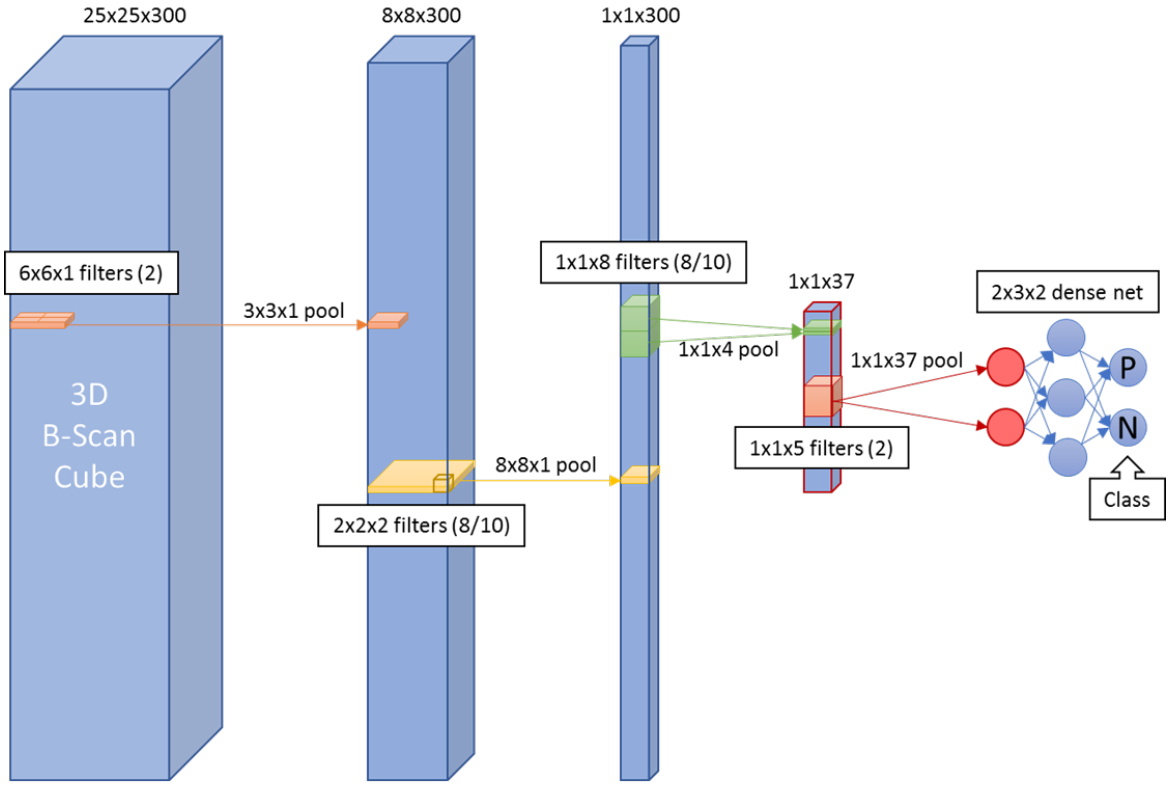

Figure 21. Network diagram for the 3D convolutional network. Height and width (the en face plane) is the top on the cuboids, with the vertical extent representing depth. The number of filters used for each layer is expressed in text labels instead of as a visible dimension. Convolutional filters are displayed in alternate colors in the cuboids, while pooling layers are between them.

5.2.1 Model Training Results

Like the Fundus imagery, models trained on examples of photothermal lesions could consistently reach 96% accuracy when tested against the validation dataset and produced very stable and reliable training, showing little overfitting. Figure 22 shows the validation results for models during training on the photothermal lesion OCT dataset.

Investigation of the misclassified instances in one of the runs discovered that two of the “false positives” (non-lesion examples that were classified as lesions) were actually lesions that had not

been labeled during the manual labeling effort. For those lesions the model had performed more accurately than the human labeler.

The models for photomechanical lesions also produced relatively reliable validation results, though two lower performing runs demonstrated some slight overfitting. Figure 23 shows the validation results for the photothermal models. Intriguingly, the OCT photomechanical models performed significantly worse than the models developed for photomechanical lesions in the Fundus imagery (87% vs. 95.5%). While there was significantly more training data for the Fundus imagery, which could contribute to the higher accuracy, this may also suggest that for photomechanical lesions, which do not have as much of a presence in the 1-channel 3D images, the color channels may be more useful for detection than the additional depth information. Intuitively one would expect that the OCT imagery provides more opportunity to distinguish between lesions and other aspects of the eye, and perhaps with more training data it may be possible to improve the OCT classification performance to match that of the Fundus imagery.

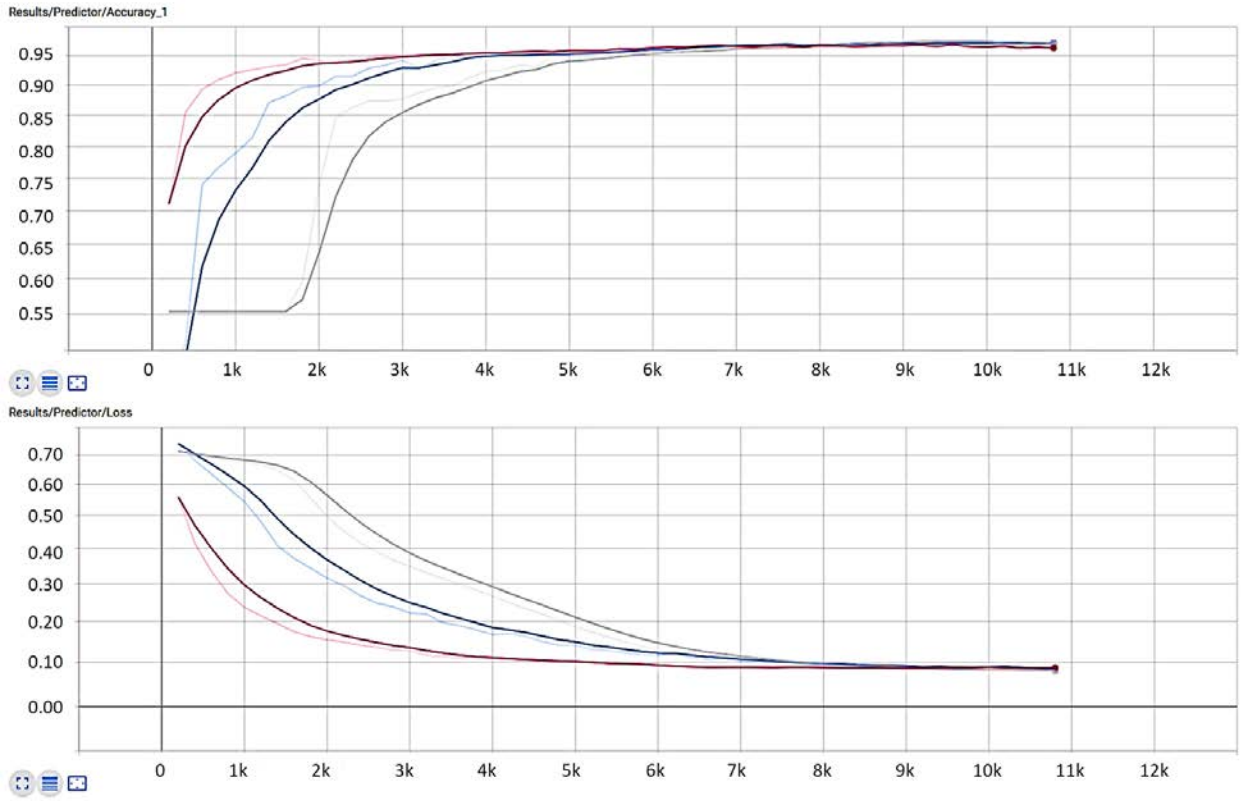

Figure 22. Validation accuracy (Y-axis, top) and validation loss (Y-axis, bottom) vs. training epoch (X-axis) for multiple training runs on photothermal OCT images. Accuracy ranged from 95.6% to 97.2%. Fewer runs were recorded due to the stability of the results.

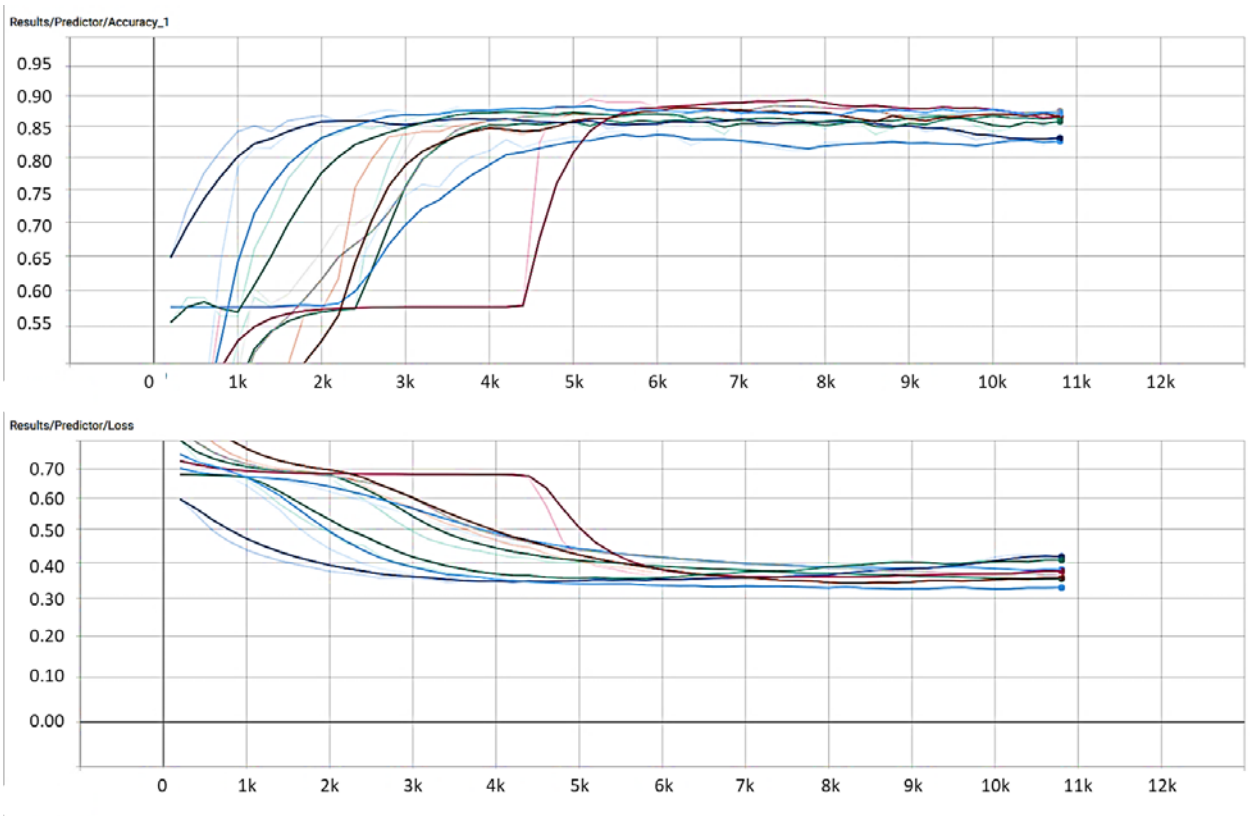

Figure 23. Validation accuracy (Y-axis, top) and validation loss (Y-axis, bottom) vs. training epoch (X-axis) for multiple training runs on photomechanical OCT images. Accuracy ranged from 82.8% to 87.7%. Most results clustered around 87%.

6 CONCLUSIONS

Oceanit has successfully demonstrated an ability to detect both photothermal (96% accuracy) and photomechanical (95.5% accuracy) lesions with high accuracy in Fundus imagery, a similarly high accuracy ability to detect photothermal lesions (96.5% accuracy) in OCT imagery, and a lower but significant ability to detect photomechanical lesions (87% accuracy) in OCT imagery. Due to flaws within the hyperspectral data, the experiments on hyperspectral imagery were inconclusive.

Differences between the classification capability of the Fundus and OCT imagery suggest that there may be some merit to producing a hybrid model that uses an ensemble of both types. While a definitive conclusion cannot be drawn, the weaker classification accuracy for photomechanical lesions in the single-channel 3-dimensional OCT imagery vs. the 3-channel 2-dimensional Fundus imagery suggest that there may be aspects of photomechanical lesions that are more easily recognized using color than depth information. Similarly, while characterizing the misclassified images was not one of the areas focused on, observation of those misclassifications by the experimenter suggests that OCT imagery has little bias against older lesions that can be difficult to detect in Fundus imagery, providing a potential to classify lesions even after they had experienced some healing.

Finally, the addition of a confidence metric to the network output significantly cut down on false positive results without significantly reducing true positives. This method of sensitivity tuning can be used in real time with a slider control to allow a clinician to investigate the confidence of predictions or manually select an appropriate sensitivity. With this sensitivity tuning capability, we believe there is promise for the lesion detection software to be provided to clinicians without expertise in laser lesion injuries as a first-line tool to recognize such injuries and refer patients to experts for further analysis.

7 REFERENCES

[1] "Federal Aviation Administration," 2017. [Online]. Available: https://www.faa.giv/about/initiatives/lasers/hazards.

[2] K. Schuster, R. Vincelette, A. Shingledecker, J. Bixler, B. Rockwell and W. Kinerk, "International Laser Safety Conference," in *Biomarkers and sensitive evaluation of retinal laser lesions*, 2015.

[3] Oceanit, "RetinaView," Oceanit, [Online]. Available: https://www.oceanit.com/products/retinaview.