

Deep Active Learning Using Scanning Laser Ophthalmoscopy Image Embedding Vectors to Optimize Training Data Selections



Yifan Zhang^{1,2}, Christy K. Sheehy PhD¹, Min Zhang PhD³, Ethan A. Rossi PhD³, Ari Green MD⁴, Joe Xing PhD¹



1. C. Light Technologies, Inc., San Jose and Boston, United States; 2. University of California Berkeley, Berkeley, United States; 3. University of Pittsburgh, Pittsburgh, United States; 4. University of California San Francisco, San Francisco, United States

INTRODUCTION

Deep learning (DL) has made breakthroughs in many fields in recent years. However, the success of DL requires constant re-training with large amounts of new data which are often costly, causing data paucity problems in the medical space. Active learning (AL), on the other hand, aims to reduce the amount of training data while still retaining similar performance through the exploration of the added value of data. In this study, we have designed an AL system to explore a faster learning curve in terms of training Artificial Intelligence (AI) models which minimize the need for data. The model is designed for object detection, using Deep Neural Network (DNN), on scanning laser ophthalmoscopy (SLO) retinal images. Image embedding vectors are then proactively used to select the most informative data.

METHODS

To ensure minimum selection bias, we established a baseline of k-fold cross validation frameworks to measure model performance; where 809 annotated images were randomly selected from our database and divided into multiple regions of interest with minimum selection bias. We first ran our DNN model using all images, then extracted the embedding vectors from the last to the second layer before it spit out prediction results. The extracted embedding vectors were then used to classify input images. We then applied a dimensionality reduction technique via PCA (Principal Component Analysis) to plot each image as a dot where x is its first eigenvalue and y is its second eigenvalue. After running a K-means clustering algorithm, we noticed there exist two distinct image clusters shown below (Fig. 1).

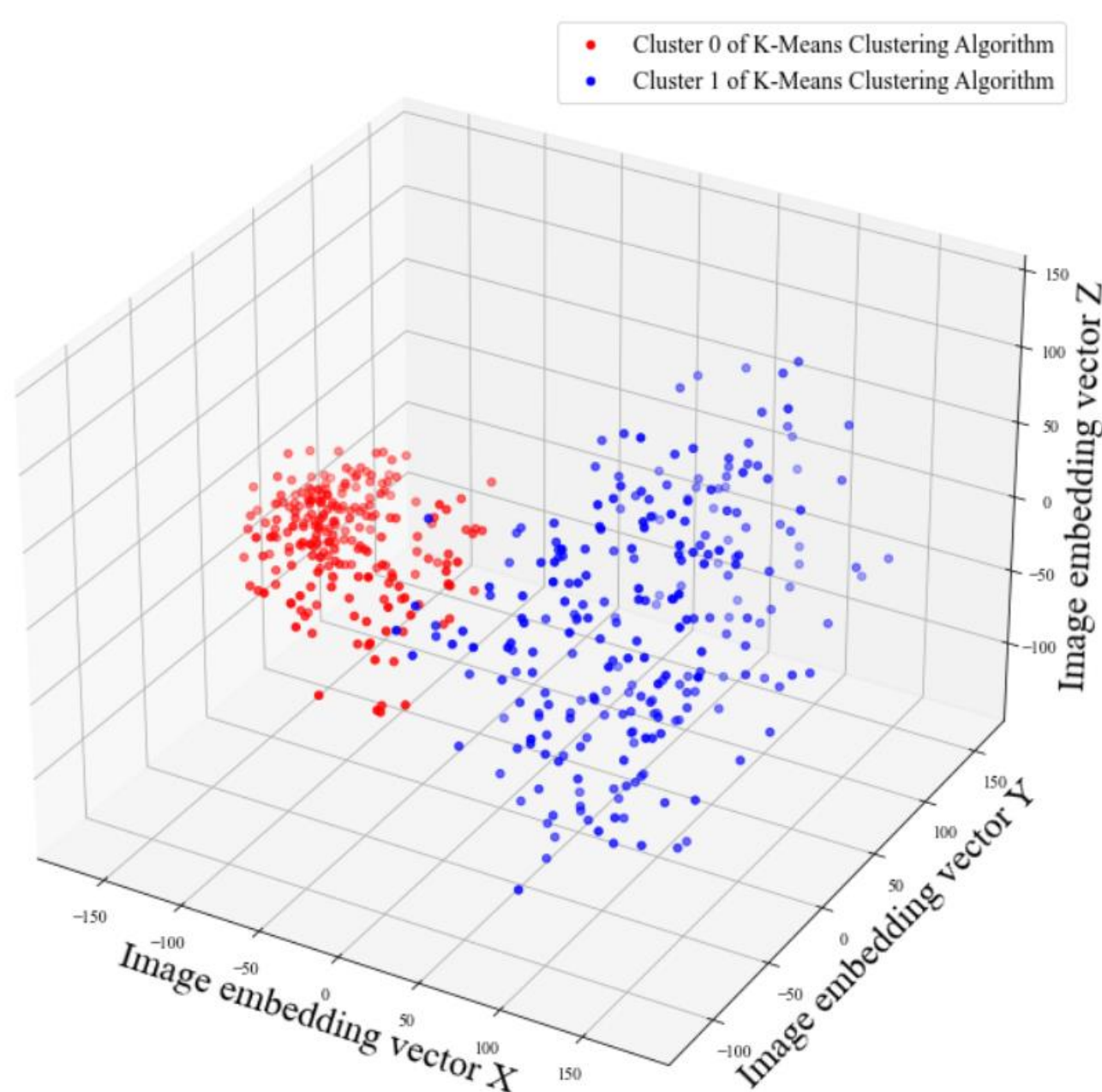


Figure 1. Two clusters of images after applying K-Means algorithm

Each point here represents an image with a vector length of 5016. The x, y, and z axes represent the three vectors with biggest eigenvalues, and they are used to visualize the two distinct clusters in 3-dimensional space.

THE MODEL

After applying CLAHE (Contrast Limited Adaptive Histogram Equalization), the retinal images were fed into the DL-based model, a convolutional neural network single-stage detector, that uses a ResNet-50 as a backbone.

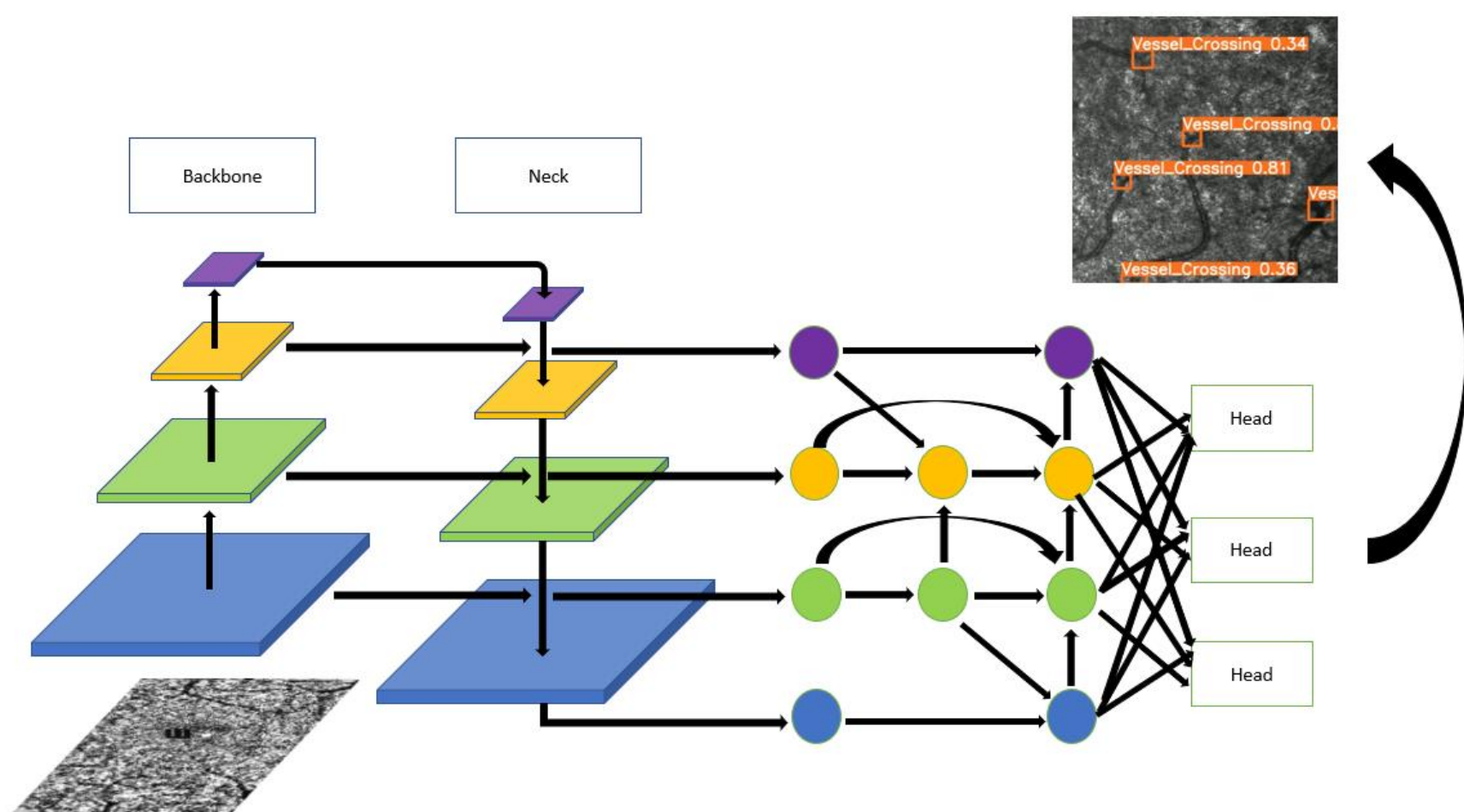


Figure 2. Architecture of Deep Neural Network

The DNN uses ResNet-50 as the backbone which pools image pixels to form features at different granularities. Bi-FPN is chosen as its neck to combine and mix the convolutional layer representations and feed into the multiple heads to generate predictions.

RESULTS

The trained model was used to predict vessel junctions on the SLO validation dataset. The results were benchmarked by mean average precision (mAP), precision, and recall through an in-house validation pipeline. Here we showcase the detection result on selected sample images as well as the metric differences between vessel junctions in two ranges of eccentricity levels from fovea. Through the study, we noted that the model's performance decreases as eccentricity, or distance from vessel junctions to the center of fovea, gets smaller. Other publications show similar eccentricity dependencies within the retina [Lombardo et al. 2013].

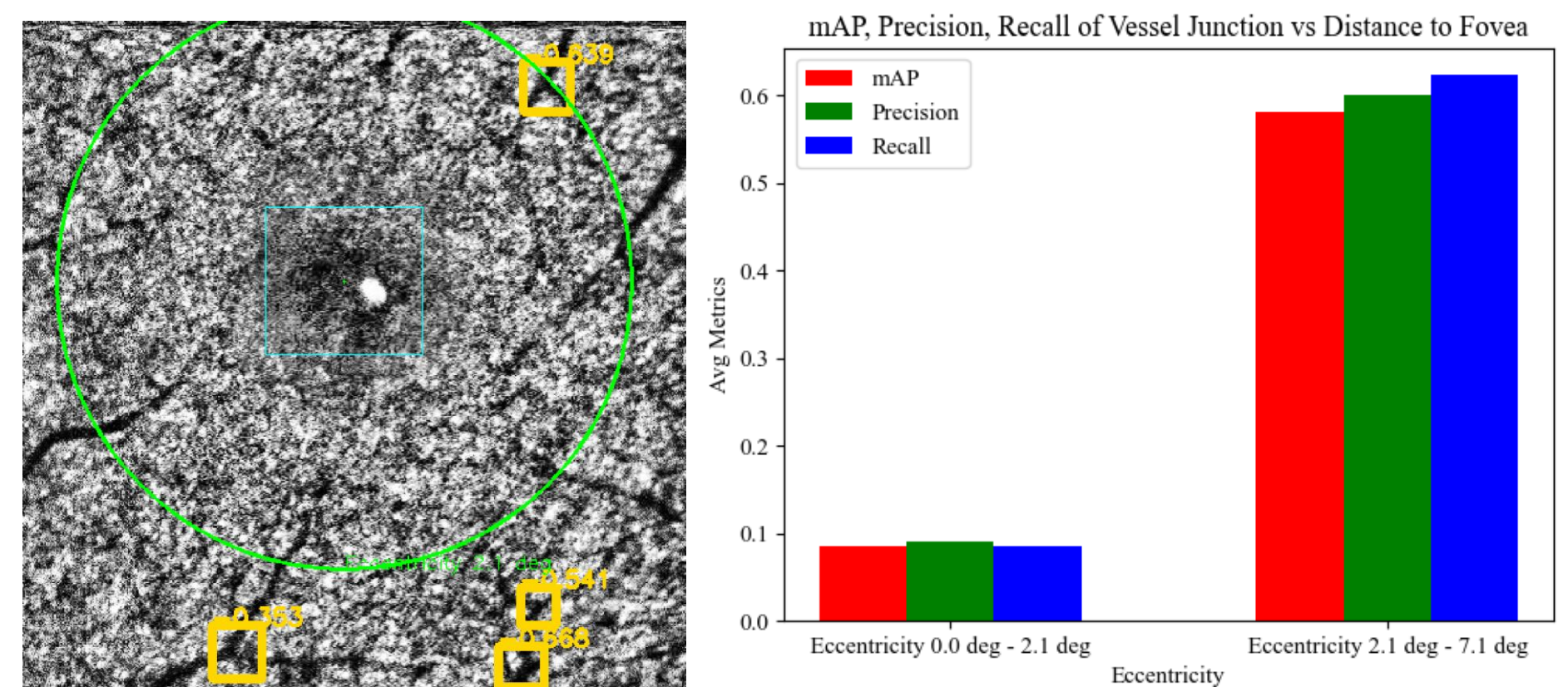


Figure 3. Study on model performances as a function of eccentricity

Vessel junctions with eccentricity between 0 and 2.1° are harder to detect than those with eccentricity degree larger than 2.1°. The mAP, precision, as well as recall on vessel junction detections with larger eccentricity degree can be six to seven times higher than those of detections with smaller eccentricities.

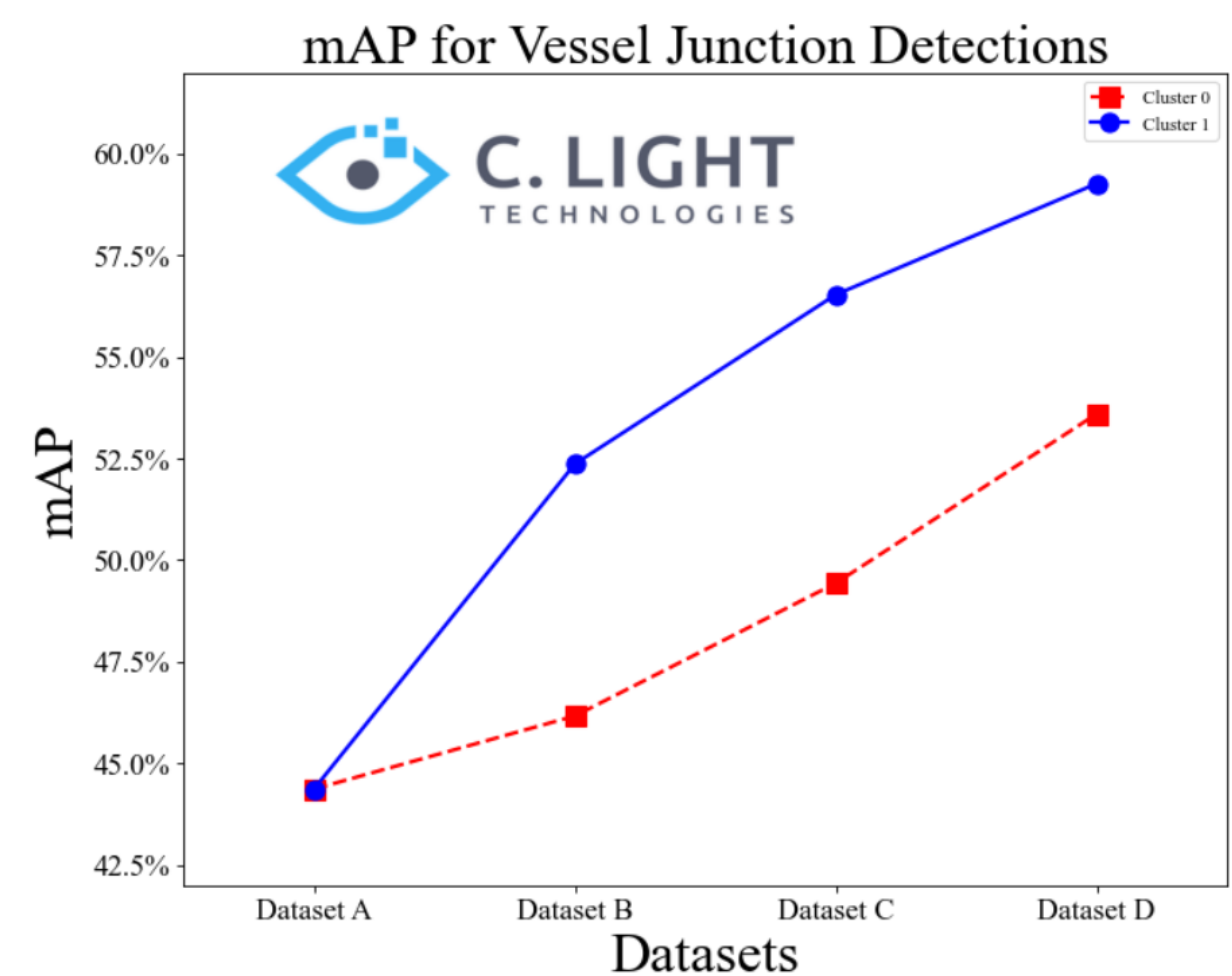


Figure 4. mAP of model trained with incrementally added images from each cluster

Dataset A is the baseline training set with no added images from either clusters. Dataset B refers to two data sets that each takes 70 randomly selected images from one of the clusters. Dataset C refers to two data sets that each takes 140 randomly selected images from one of the clusters. Dataset D refers to two data sets that each takes 210 from one of the clusters.

CONCLUSIONS

After running the DNN model iteratively for each dataset with added images from clusters, we observed that (1) the mAP increased as expected as we were adding more images to the training set and (2) there was a difference in learning curves for the two clusters indicated by their slopes. We observed one cluster tended to converge faster to the desired performance than the other by a factor of 2.3. Results from our study show a strong indication that AL can indeed accelerate the learning process of our object detection model, meaning a potential cost reduction for data labeling services and faster training speed for models.

REFERENCES

Lombardo, Marco, et al. "Eccentricity Dependent Changes of Density, Spacing and Packing Arrangement of Parafoveal Cones." *Ophthalmic and Physiological Optics*, vol. 33, no. 4, 2013, pp. 516–526., <https://doi.org/10.1111/opo.12053>.

This research was supported by grants from the National Institutes of Health R41 NS100222-01A1 (A.J.G. and C.K.S.).

