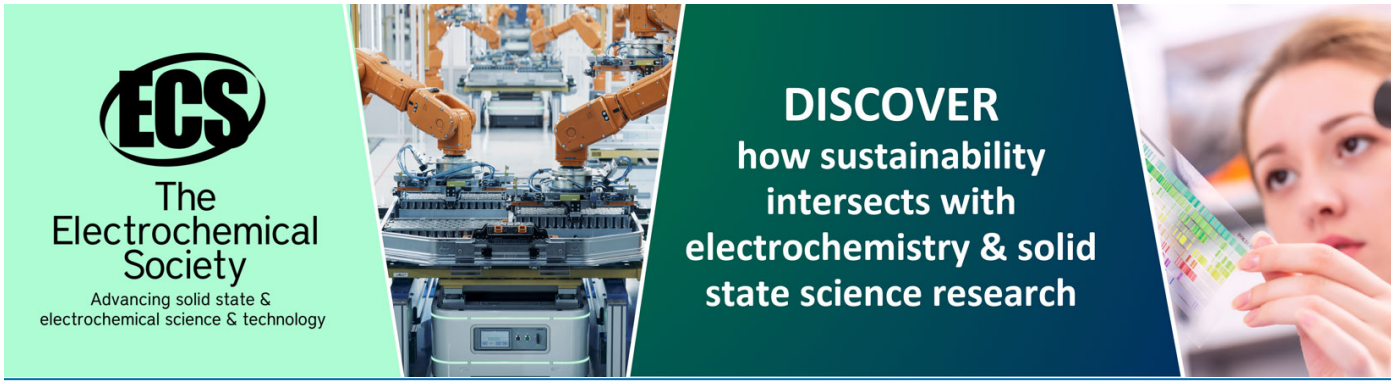
A Comparative Study on Deep Networks for Glaucoma Classification



A Comparative Study on Deep Networks for Glaucoma Classification

Zifan Ying[1,5], **Zhichong Wang**[2,6], **Hongbo Zhang**[3,7,9], **Rongxuan Zhang**[4,8]

[1]ZJU-UIUC Institute, ZJUI, Zhejiang University, Hangzhou, China
[2]School of Mechanical Engineering and Electronic Information, China University of Geosciences, Wuhan, China
[3]Department of Computer Science Technology, BNU-HKBU United International College, Zhuhai, China
[4]Department of Electrical Engineering, University of California Davis, Davis, California, the USA

[5]zifan.21@intl.zju.edu.cn
[6]zhichongwang38@gmail.com
[7]q030026199@mail.uic.edu.cn
[8]zrxzhang@ucdavis.edu
[9]corresponding author

Abstract. The purpose of this study is to classify glaucoma and non-glaucoma images from REFUGE dataset of fundus images. Due to the imbalance of dataset, we did data augmentation and preprocessing for dataset first (including feature extraction and enhancement). We then tested the performance of some deep convolutional neural networks as baselines, including ResNet, GoogLeNet, and VGGNet. Later we introduced self-attention layer into our CNN model and tried a method based on cup-to-disc ratio. Compared to the unprocessed dataset, the processed (data augmentation and feature enhancement) dataset gave a better performace. And self-attention model also improved performance beyond original CNN. Finally our method base on the cup-to-disc ratio was way better than the CNN models above.

Keywords: Glaucoma Classification, Fundus Images, Medical Imaging, Deep Learning

1. Introduction

Glaucoma is a neurodegenerative disorder that causes gradual damage to the optic nerve and retinal nerve fibers, leading to visual impairment or loss [1]. The early symptoms of glaucoma are often difficult to detect, making early diagnosis critical to prevent irreversible vision loss and blindness [2]. A confirmed glaucoma diagnosis currently requires multiple clinical examinations, including measuring intraocular pressure using a tonometer (a device used to measure intraocular pressure), inspecting the optic nerve head's integrity using optical coherence tomography (a non-invasive imaging technique that provides high-resolution images of the optic nerve head and retinal nerve fiber layers), and measuring the patient's visual field. Color fundus photography (CFP) is a retinal imaging method that captures detailed images of the retina's blood vessels and structures. This technique offers a cost-effective and non-invasive opportunity for screening glaucoma in at-risk populations [3, 4]. By analyzing the images obtained from CFP, clinicians can detect early signs of glaucoma and monitor disease progression, allowing for early intervention and treatment. In summary, glaucoma is a progressive neurodegenerative disorder that can cause visual impairment or loss. Early detection and treatment are crucial for preventing irreversible

vision loss and blindness. While diagnosing glaucoma currently requires multiple clinical examinations, CFP is a promising screening tool that can aid in the early detection and monitoring of the disease [3].

In clinical practice, identification of glaucoma from a retinal image typically involves image analysis and processing. In glaucoma patients, the optic disc commonly exhibits distortion and damage, characterized by defective, detached or thinning nerve fiber layer surrounding the optic disc, and a change in the optic disc shape from circular to oval or elongated in different directions. Moreover, an abnormal disc color, such as grayish or greenish-brown, is often observed in glaucoma patients.

Nerve fiber layer thickness analysis is a widely used method for detecting glaucoma, involving the examination of nerve fiber layer thickness and density around the optic disc. The optic disc, a round or oval structure observed in fundus images, serves as an indicator of eye health [3]. Optic disc morphology analysis is another technique used to diagnose glaucoma, involving the assessment of morphological parameters, such as optic disc diameter and aspect ratio, to evaluate optic disc deformation and damage. In addition, optic disc color analysis is utilized to detect abnormal color changes, which may indicate the presence of eye diseases, such as glaucoma [5]. Normally, the color of the optic disc is light red or yellow, whereas in patients with glaucoma, it typically appears grayish or greenish-brown due to optic nerve head and surrounding nerve fiber layer damage. In summary, nerve fiber layer thickness analysis, optic disc morphology analysis, and optic disc color analysis are common diagnostic methods for glaucoma. Through the analysis of optic disc thickness, density, morphology, and color, clinicians can determine the status of eye health and diagnose glaucoma.

Early detection and treatment are crucial for preventing irreversible vision loss and blindness; however, identifying early symptoms of glaucoma and diagnosing the condition involves time-consuming and costly clinical examinations [6]. Automated classification of glaucoma, which utilizes machine learning algorithms and computer vision techniques to analyze retinal images for signs of glaucoma, provides a promising solution to these challenges. This approach offers several advantages, including faster and more cost-effective screening for glaucoma, reduced subjectivity of diagnosis, and the potential to address the shortage of trained ophthalmologists, particularly in developing countries. Automated classification of glaucoma has the potential to revolutionize the screening and diagnosis of this condition by providing a faster, more objective, and cost-effective method for detecting glaucoma, which can ultimately improve patient outcomes and reduce the burden on healthcare systems.

2. Related Work

2.1. Overview of Glaucoma classification.

Medical image classification is a challenging task that requires the use of image processing, pattern recognition, and classification methods. The goal is to achieve high accuracy and identifying the affected parts of the human body by the disease. [4, 7].

Gomez et al. [8] used deep CNN to detect glaucoma. The authors analyzed the effect of the model architecture, size of data and training methods. Three data sets in total were utilized in evaluating the performance of five different model architectures, and VGG19 applied with transfer learning strategies proved to be the best option, reaching an AUC of 0.94, sensitivity of 87.01%, and specificity of 89.01%.

Issac et al. [9] presented a glaucoma detection method that utilizes CDR (cup-to-disc ratio), NRR (neuro-retinal rim area). An approach based on adaptive threshold that incorporates local statistical features of the fundus image was used in the algorithm segmenting optic cup and optic disc. SVM and ANN classifiers were used, reaching 94.11% accuracy, 100% sensitivity, and 90% specificity.

Another method proposed in [10] integrated both local and holistic features. It used a deformable shape model to locate ROI in retinal images and normalize with respect to the bounding box of optic disc. And features of ROI are extracted by CNN trained on more then 1 million images and SVM was used to detect glaucoma. An AUC of 0.8384 is reached.

Table 1 shows the benchmark of popular CNN models, including AlexNet [11], GoogleNet [12], VGGNet [13], etc., measured in AUC score.

Wang et al. [14] propose a framework that combines segmentation and image-based features for

Method	AUC		
	Local	Holistic	Holistic+Local
AlexNet	0.8184	0.8108	**0.8384**
GoogleNet	0.7149	0.5798	0.7187
VGG-16	0.7635	0.7479	0.7771
VGG-19	0.7529	0.7551	0.7785
Cheng et al. [5]	0.800		
Xu et al. [6]	0.823		
Manual	0.839		

Table 1: Benchmark of popular CNN models in AUC

glaucoma classification. Their work includes designing features to describe the segmented optic disc (OD) and optic cup (OC) regions and creating features based on Texture of Projection (ToP) [15] and color-based Bag of Words (BoW) [16], and classification performance demonstrated the effectiveness of these features. When using OD and OC features alone, the improvement in accuracy was 14.22% over CDR ratio based classification, while with ToP and BoW features alone, the improvement in accuracy was 10.92% over wavelet-based features.

2.2. Image processing methods

2.2.1. Sobel filter The Sobel filter [17] is a type of high-pass filter that captures the high-frequency components of the image, which are the edge details. The Sobel filter's kernels are

$$\begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} \text{ or } \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix},$$

and the Sobel filter derives the image from the x and y directions, respectively.

2.2.2. Histogram equalization Histogram equalization [17] is an image enhancement method that can improve the contrast of an image, making it clearer and easier to analyze. The method adjusts the gray distribution of the image by redistributing the gray values of the image pixels to make the distribution more uniform, thereby enhancing the image.

2.2.3. Histogram stretch Histogram stretching [17] is an image enhancement method that involves increasing the contrast of an image by stretching the pixel grayscale range. By remapping the gray levels of the image, the pixel value range of the original image is mapped to a wider range, allowing for a more even distribution of pixel values throughout the entire gray range. This process enhances the image by improving the contrast and making it easier to analyze.

2.2.4. Directly brighten the image As its name, this image processing method [17] has capacity to directly increase the brightness of the image. And the process of its realization are converting the color image from RGB format to HLS format, and adjusting the lightness (L) and saturation (S) parameters.

3. Methods

This section introduces the neural networks that we used in this whole project, including ResNet, GoogLeNet, and BoTNet.

3.1. ResNet

ResNet18 [18] is a deep convolutional neural network structure introduced by Microsoft Research Asia in 2015. It is a residual network designed to address the problems of gradient disappearance and overfitting during deep neural network training by incorporating residual blocks. ResNet's key innovation is the integration of "skip connections" in the network. In each residual block of the network, the input is directly added to the output, enabling the network to learn residual information more effectively and avoid the gradual loss of information with increasing depth. Fig. 1 illustrates the structure of the residual block.

ResNet18 is a smaller version of the ResNet architecture, with fewer layers and parameters than the original network. ResNet34 is another variant of the ResNet family of networks, characterized by their use of residual blocks. These blocks enable the network to learn a residual mapping instead of a direct mapping, allowing for the propagation of gradient information without being significantly diminished by the effects of vanishing gradients.

The ResNet34 architecture consists of 34 layers, with most of the layers being convolutional layers. In addition to the residual connections, ResNet34 also employs batch normalization and ReLU activation functions to improve training stability and accelerate convergence. ResNet34 has been shown to be highly effective in a wide range of image recognition tasks, achieving state-of-the-art performance on several benchmark datasets. Its success has also led to numerous follow-up studies and variations on its original architecture, contributing to the continued development of deep neural networks.

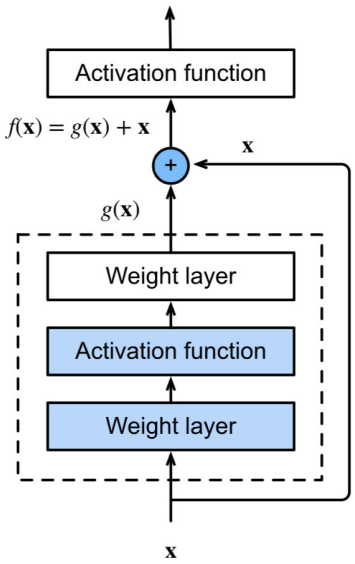


Figure 1: Residual block structure [19]

3.2. GoogLeNet

GoogLeNet is a deep convolutional neural network introduced by the Google team in 2014, which has a relatively low number of parameters compared to other deep neural networks. It incorporates the Inception Module, a module structure that combines convolutional and pooling kernels of varying scales to improve feature extraction efficiency. The overall structure of the GoogLeNet neural network is shown in Fig. 2, while the Inception Module structure is depicted in Fig. 3. These networks are widely recognized as classic deep neural networks for classification and were selected as a baseline for this project.

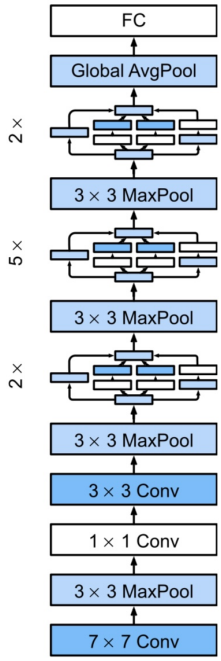


Figure 2: GoogLeNet network structure [19]

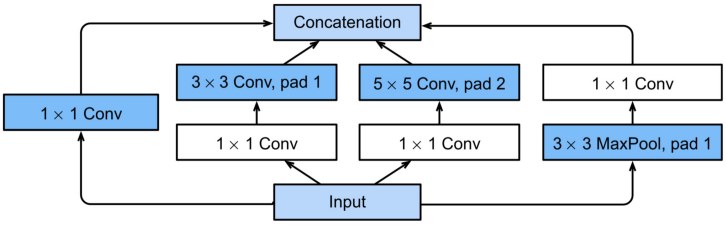


Figure 3: Inception block structure [19]

3.3. BoTNet

BoTNet is a recently proposed neural network architecture that uses self-attention for various tasks, such as object detection, image classification and segmentation [20]. BoTNet is designed to replace spatial convolutions with global self-attention in the last three bottleneck blocks (i.e. 1x1 Conv layer) of a ResNet, thereby improving feature representation and extraction efficiency. On the ImageNet benchmark, BoTNet achieves a top-1 accuracy of 84.7% [20].

Table. 2 shows the architecture of BoTNet-50 [20]: The only difference of BoTNet-50 and ResNet50 is that the 3×3 convolution layer in c5 is replaced with a MHSA (multi-head self-attention) layer proposed in the Transformer [21].

Fig. 4 shows the structure of MHSA layer. The center-position structure was trained to distribute ”attention” on image space.

stage	output	ResNet-50	**BoTNet-50**
c1	512 × 512	7×7, 64, stride 2	7×7, 64, stride 2
c2	256 × 256	3×3 max pool, stride 2	3×3 max pool, stride 2
		$\begin{bmatrix} 1\times1, 64 \\ 3\times3, 64 \\ 1\times1, 256 \end{bmatrix} \times3$	$\begin{bmatrix} 1\times1, 64 \\ 3\times3, 64 \\ 1\times1, 256 \end{bmatrix} \times3$
c3	128 × 128	$\begin{bmatrix} 1\times1, 128 \\ 3\times3, 128 \\ 1\times1, 512 \end{bmatrix} \times4$	$\begin{bmatrix} 1\times1, 128 \\ 3\times3, 128 \\ 1\times1, 512 \end{bmatrix} \times4$
c4	64 × 64	$\begin{bmatrix} 1\times1, 256 \\ 3\times3, 256 \\ 1\times1, 1024 \end{bmatrix} \times6$	$\begin{bmatrix} 1\times1, 256 \\ 3\times3, 256 \\ 1\times1, 1024 \end{bmatrix} \times6$
c5	32 × 32	$\begin{bmatrix} 1\times1, 512 \\ 3\times3, 512 \\ 1\times1, 2048 \end{bmatrix} \times3$	$\begin{bmatrix} 1\times1, 512 \\ \texttt{MHSA}, 512 \\ 1\times1, 2048 \end{bmatrix} \times3$
# params.		**25.5**$\times10^6$	**20.8**$\times10^6$
M.Adds		**85.4**$\times10^9$	**102.98**$\times10^9$
TPU steptime		**786.5** ms	**1032.66** ms

Table 2: Architecture of BoTNet-50 (BoT50) [20]

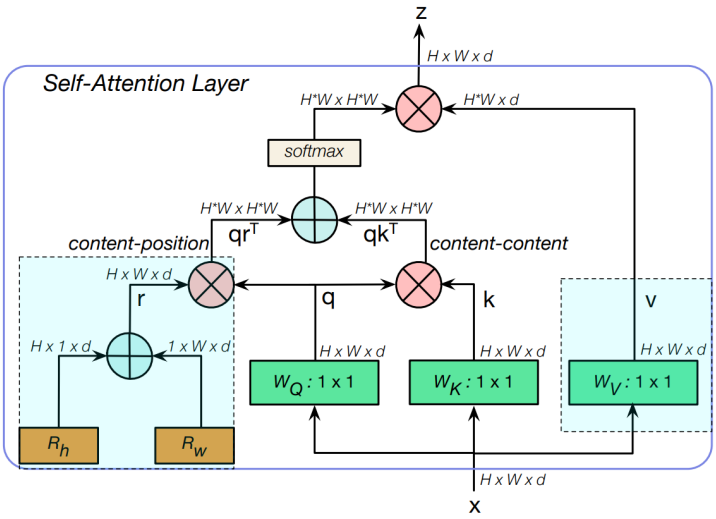


Figure 4: MHSA layer [20]

4. Numerical Experiments

4.1. Data and Evaluation Metrics

4.1.1. Dataset The REFUGE dataset [22] comprises 2,000 retinal color fundus photographs (CFPs), collected from the Zhongshan Ophthalmic Center at Sun Yat-Sen University in China. The images were taken by experienced ophthalmologists and technicians, using high-quality equipment in a darkroom environment. The CFPs are stored in JPG format, with each color channel consisting of 8 bits. The images were captured with the main focus on the optic disc (OD) region, the macular area, or the midpoint between the OD and macula (both visible), similar to a clinical setting. Left and right eye images were included if they met the necessary quality requirements. This dataset provides a valuable

resource for training and evaluating computer vision algorithms for glaucoma detection and other related tasks.

4.1.2. Metrics Due to the imbalanced data, an accuracy of 90% can be obtained by giving constant glaucoma prediction, so we employed other metrics to assess classification model.

- Cohen's κ coefficient
$$\kappa \equiv \frac{p_o - p_e}{1 - p_e} = 1 - \frac{1 - p_o}{1 - p_e}$$
with p_o being the relative observed agreement and p_e the hypothetical probability of chance agreement.
- F1 score F1 score is defined by
$$F_1 = \frac{2}{\text{recall}^{-1} + \text{precision}^{-1}}$$
where
$$\text{precision} = \frac{TP}{TP + FP}$$
$$\text{recall} = \frac{TP}{TP + FN}$$
- AUC (Area under the ROC Curve)
With different threshold, a classifier gives some resulting (FPR, TPR) points in the ROC space, where
$$TPR = \frac{TP}{TP + FN}$$
$$FPR = \frac{FP}{FP + TN}$$
And the AUC refers to the area under this ROC curve.

These metrics can handle the case where the data is unbalanced.

4.2. Experiment Set-up

4.2.1. Data Augmentation Due to the unbalance between Glaucoma and Non-Glaucoma (360 images against 40 images), we mainly employed data augmentation as a solution and the main augmentation done by applying rotation on the Non-Glaucoma images. The angle of the rotation is Gaussian distributed with $std = 20°$.

Fig. 5 shows some data samples from the preprocessed dataset. Specifically, in the baseline section, we employed 400 images from the training set of the Refuge dataset for training and 400 images from its validation set for testing.

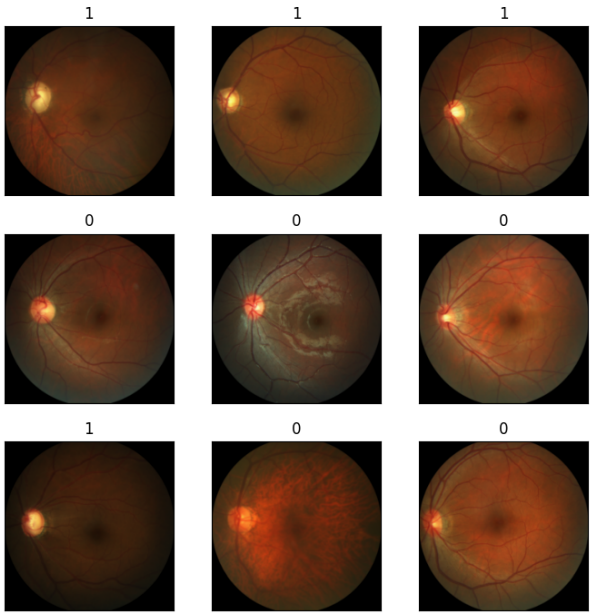


Figure 5: Data samples (with augmentation) from REFUGE [22] training set (0 = Non-Glaucoma, 1 = Glaucoma)

4.2.2. Feature extraction visualization we take a figure from glaucoma datasets as an example ”Fig.6”.

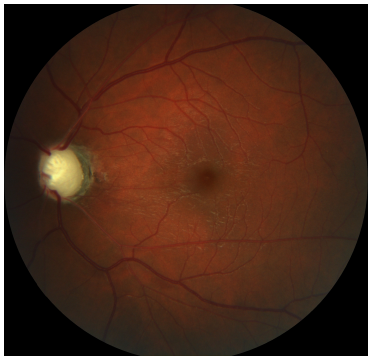
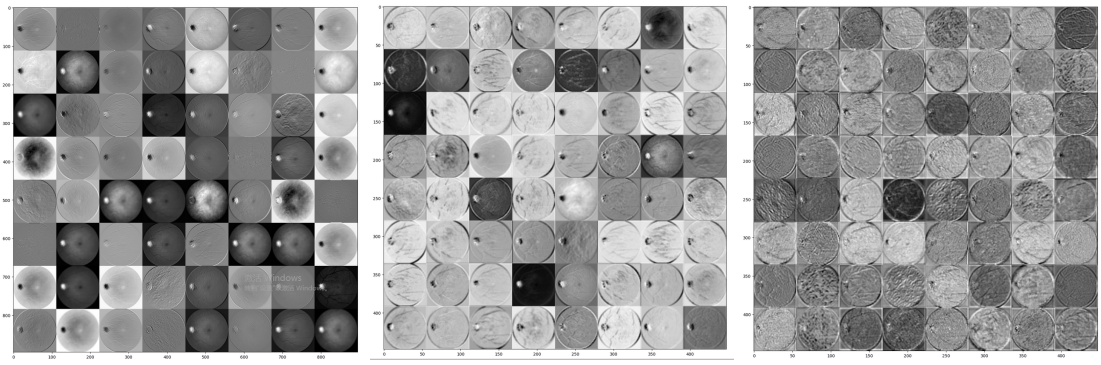


Figure 6: Example of a glaucoma image

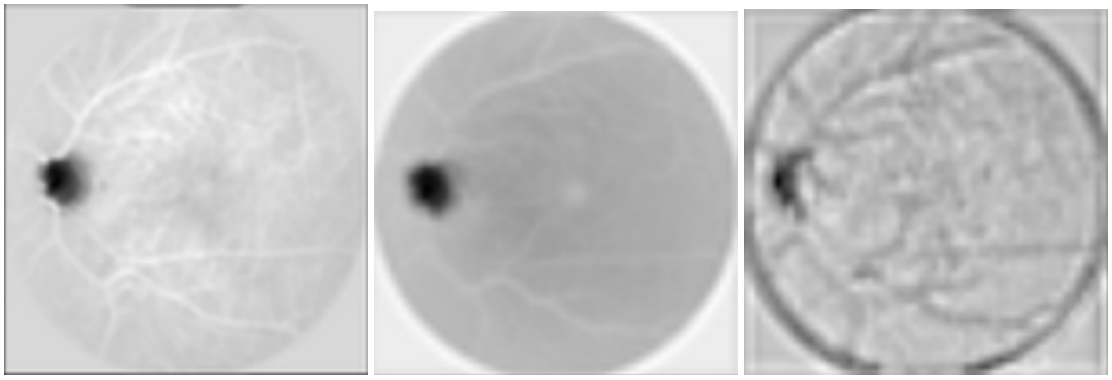
This study involved a visualization analysis of the first, second, and fifteenth convolutional layers in the ResNet34 architecture. Result shows in ”Fig.7” and a more detailed example shown in ”Fig.8”. These Results showed that the first convolutional layer, which is the output of the first layer of convolutional layers, had a clear outline with distinctive focus in each channel. The second layer, which is the output of the first layer of the second layer of convolutional layers, captured more high-level features that were capable of identifying local features, overall shape, color, and texture information of objects.

In our visualization analysis of the ResNet34 architecture, we observed that the high-level features of specific objects and overall shape were not immediately apparent. However, we did notice that color and texture information remained a key aspect of high-level feature extraction. Furthermore, as we moved to the deeper layers of the network, such as the fifteenth layer, we observed that the channels became more abstract, indicating that these features may be better suited for computer-based feature judgments rather than human interpretation.



(a) The first layer (b) The second layer (c) The fifth layer

Figure 7: Visualization of features in specific layers



(a) The sub-figure of first layer (b) The sub-figure of second layer (c) The sub-figure of fifth layer

Figure 8: An exemplar of visualization of features in specific layers

Notably, our analysis revealed that certain channels were effective at removing or fading other features, allowing for clearer identification of specific objects, such as the vessels in the optic disc area. Overall, our study demonstrated the utility of visualization in understanding the inner workings of deep neural networks and furthering their development.

4.3. GoogLeNet and ResNet

The section of our project is divided into two main components. The first component involves training the GoogLeNet and ResNet18 networks from scratch and analyzing the results. The second component entails utilizing the pre-trained ResNet34 model for transfer learning and feature extraction. We will present each of these components separately in the following report.

4.3.1. Train from scratch by using GoogLeNet and ResNet network model(the first phase) Upon completion of constructing the neural network and downloading the Refuge dataset, the initial training phase began. During the 60 epoch training period, test accuracy was closely monitored. However, we noted that the test accuracy often reached a value of 0.9, but occasionally dropped to 0.1, raising concerns about potential issues with the neural network. The results of training the GoogLeNet and ResNet18 networks are presented in Fig. 9, where the blue line represents the GoogLeNet network and the green line represents the ResNet18 network.

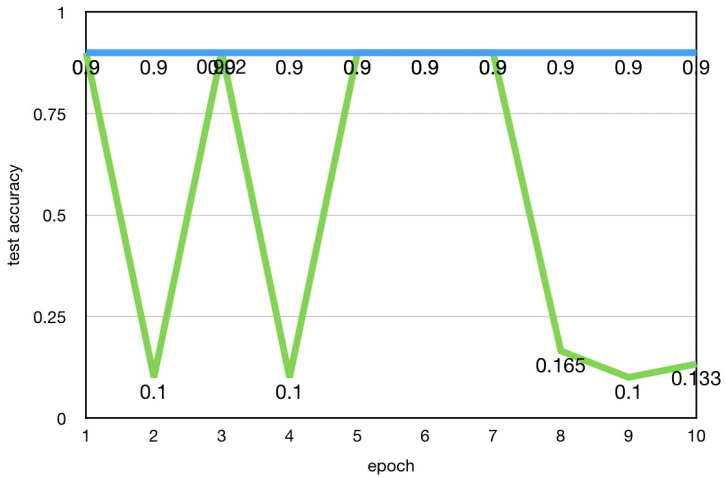


Figure 9: The results of training in the first phase(Last 10 epochs)(Blue line represents GoogLeNet network and green line represents ResNet18 network)

4.3.2. Down-sampling problem As discussed earlier, an issue was encountered during testing of the neural network, prompting further investigation into its root cause. It was observed that when a batch of test images was fed into the network for prediction, all the results obtained were 0, indicating non-glaucoma. This trend was consistent across all the batches of images tested. Following validation of the network with 400 images from the validation set, it was observed that all the images were non-glaucoma. Out of the 400 images, 360 were non-glaucoma and 40 were glaucoma. Therefore, if the network predicted all the images as non-glaucoma, the accuracy would be 0.9. Conversely, if it predicted all the images as glaucoma, the accuracy would be 0.1. It was thus concluded that both networks were incapable of accurately diagnosing glaucoma.

After identifying the potential root cause of the problem, we sought guidance from Professor Rittscher. The professor suggested that the issue may have been caused by the large-scale down-sampling resize of the images. Prior to inputting the training images into the network, we had down-sampled the original images of 2124 x 2056 resolution to a 256 x 256 resolution, leading to the loss of significant image feature information. This could have contributed to the problem we experienced. To address the issue, we decided to change the scaling-down size to 2048 x 2048. Additionally, we modified the last pooling layer of both the GoogLeNet and ResNet networks to an adaptive avg-pooling layer, allowing them to adapt to the input of this sized image. This ensured that the size of the output feature map was 1 x 1, and the length after flattening was still the number of channels.

4.3.3. Train from scratch by using GoogLeNet and ResNet network model(the second phase) After resolving the issue in the first phase, we proceeded to the second phase of training. In this phase, both networks were trained for 30 epochs. The test accuracy of the GoogLeNet network was no longer fixed at 0.9 or 0.1 in the last few epochs, indicating that the network had overcome the issue. However, the test accuracy of the ResNet network was still fixed at 0.9 or 0.1, indicating that it needed further training. It can be inferred that preserving the details of the original image as much as possible can aid in resolving this issue. The results of the second phase of training for the GoogLeNet neural network and ResNet18 neural network are shown in Fig. 10, where the blue line represents the GoogLeNet network, and the green line represents the ResNet18 network.

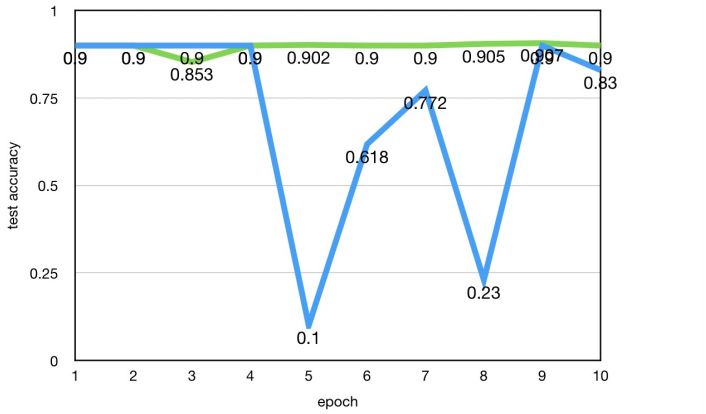


Figure 10: The results of training in the second phase(Last 10 epochs)(Blue line represents GoogLeNet network and green line represents ResNet18 network)

4.3.4. ResNet34 [18] Fig. 11 shows the performance of ResNet34 [18]

We got $F1 = 0.35$ on test set, with Learning rate=1e-5, Weight decay=1e-3, Epochs=20 and Adam optimizer [23].

As a result of the notable distinctions between medical images, such as fundus images, and more widely utilized image classification datasets, such as ImageNet [24], we chose not to employ pre-trained ResNet34 [18] models for our task. This decision deviated from common practices.

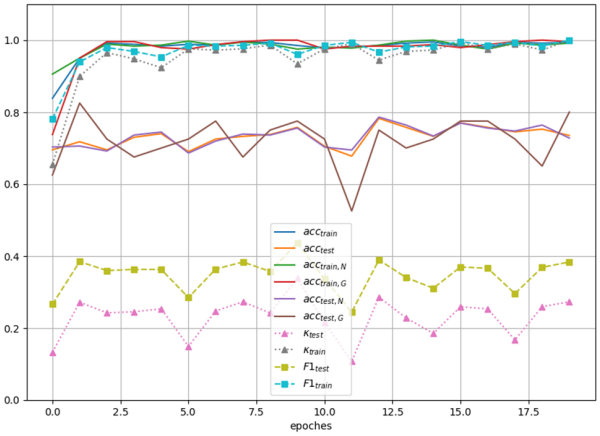


Figure 11: Baseline (ResNet34)

The experimental results demonstrated the effectiveness of data augmentation in addressing the issue of imbalanced datasets. Specifically, the F1 score and kappa coefficient obtained from training with augmented data were significantly higher and more consistent than those obtained from training without data augmentation, as illustrated in Fig. 11 and Fig. 12. These findings underscore the importance of data augmentation in improving the generalization performance of deep learning models trained on imbalanced datasets.

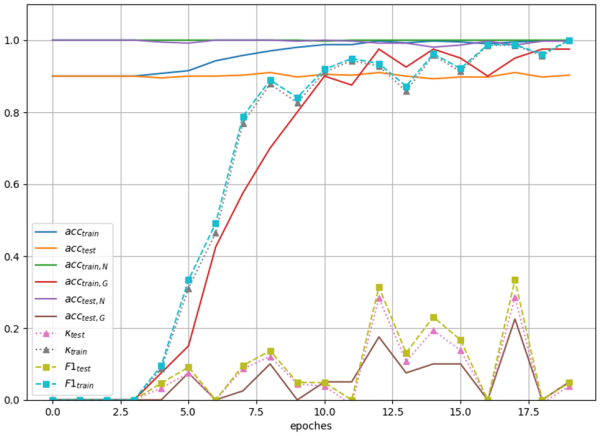


Figure 12: Training result without Data Augmentation

Due to the lack of a variable threshold in CNN ResNet or other CNN models, it is not possible to obtain a (near) continuous ROC curve. Therefore, the AUC metric cannot be applied in this case.

4.4. Self-attention model (BoTNet [20])

To address the issue of complexity in computation, we opted to customize BoTNet-34 based on ResNet34 with the BottleStack layer proposed in [20], as the original BoTNet-50 model based on ResNet50 was too large for this task. This approach also allowed us to directly compare the performance of BoTNet with plain ResNet.

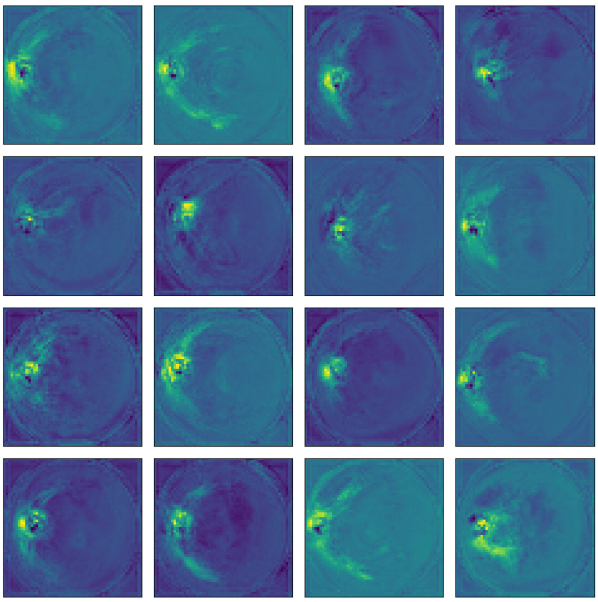


Figure 13: Visualization of feature map weight in BoTNet-34

Fig. 13 presents a visualization of the feature map in the trained BoTNet-34 model by retaining the dimension of the feature map size ($56 \times 56 = 3136$) and calculating the mean among all other dimensions. The visualization shows that the weight of the feature map of the Multi-Head Self-Attention (MHSA)

layer reached a peak around the optic disc and the connected blood vessels. This suggests that the self-attention layer extracted the important region of fundus images, indicating the potential usefulness of this layer in identifying glaucoma.

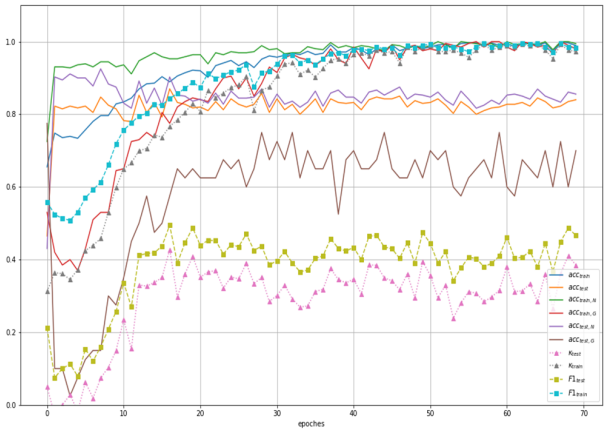


Figure 14: BoTNet-34 benchmark [20]

Fig. 14 presents the classification performance of BoTNet-34 on the REFUGE [22] dataset, with a particular emphasis on the F1 score. The model achieved an F1 score of approximately 0.45, indicating that the self-attention mechanism can enhance the classification performance on this task.

4.5. CDR related methods

The definition of glaucoma, as mentioned in [25], includes the enlargement of the optic cup as one of its symptoms. This can also be seen in the feature map weight of BoTNet-34. Therefore, the cup-to-disc ratio (CDR) is an important measure for assessing glaucoma. A higher CDR value indicates a greater likelihood of glaucoma.

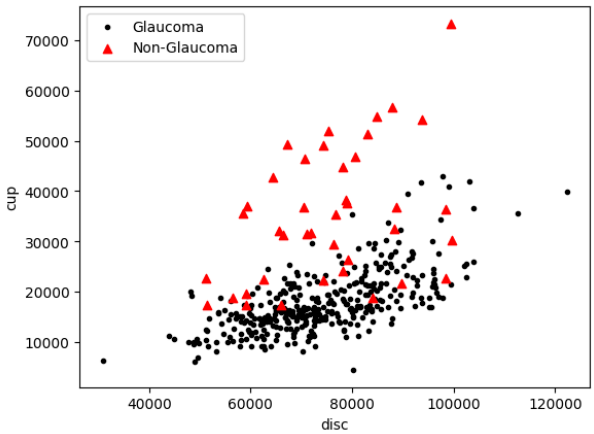


Figure 15: optic cup area against optic disc area in training set

Fig. 15 shows the area (number of pixels) of optic cup and optic disc in each image intraining set. It can be observed that the glaucoma samples (shown in red triangles) had a larger cup-to-disc ratio two classes had overlaped.

Our approach to classify glaucoma with CDR is to simply set a threshold.

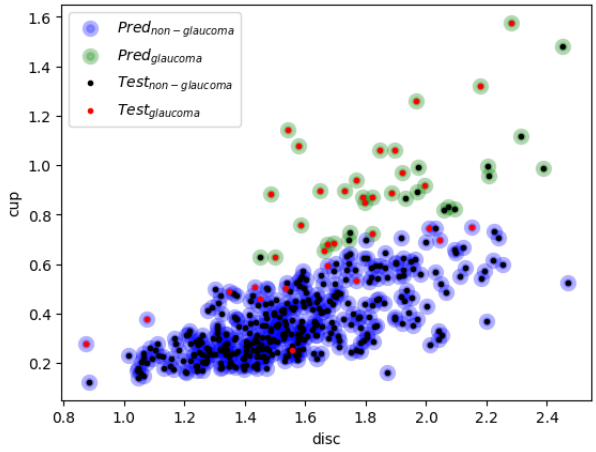


Figure 16: prediced result on test set

As Fig. 16 have shown, The result reached a $F1 = 0.64$ on test set.

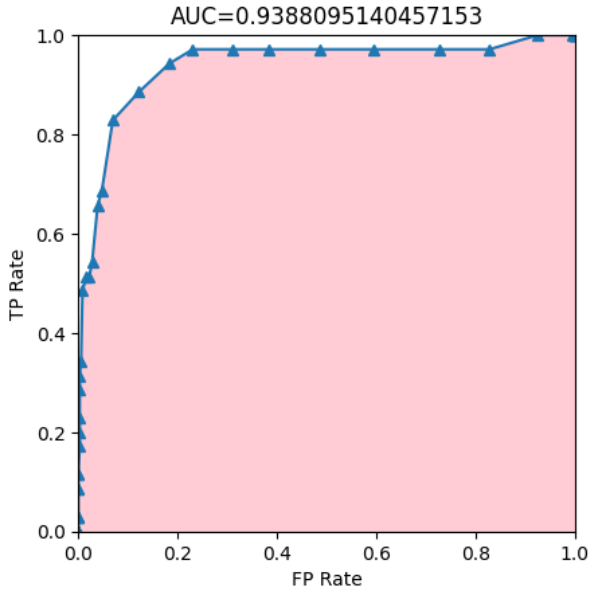


Figure 17: ROC curve and AUC of CDR threshold method

Fig. 17 showed a AUC=0.934 score in CDR threshold method. Its performance went beyond that of previous CNN models.

This CDR-based method is very straight-forward, for it has ignored the shape of optic disc and cup or even any information other than these 2 parts and only taken the ratio of their area into consideration, but the outcome was way better then our ResNet or BoTNet (the possibility that the hyperparameters of these models are not optimal is not excluded). This result inspired us to focus on optic disc and optic cup for the glaucoma classification task, because we can only get accurate area of optic discs and cups through a good segmentaion model.

4.6. Image Preprocessing

In the baseline section of this project, the neural network had difficulty in learning the ability to diagnose glaucoma. To address this, we reduced the down-sampling of the images during the initial input process, resulting in a significant increase in input image data and a decrease in the calculation speed of the entire network. However, this also led to a slower convergence speed of the neural network. In this experiment, our goal was to speed up the convergence of the network by reducing the number of epochs required for training to converge, since we could not change the time it takes to calculate each epoch. To achieve this, we explored preprocessing methods to manually enhance the features of glaucoma diagnosis in the image, which theoretically allows the entire network to focus on these features faster.

This experiment can be divided into two main parts: feature or detail enhancement, and overall improvement of the training and validation images.

4.6.1. Feature enhancement In the previous section, we discussed the method of diagnosing glaucoma through retinal fundus pictures, which focuses on the optic disc, optic cup, blood vessels outside the optic disc, and other typical features in the image. We observed that the expression of these features in the image is mainly through edges. Therefore, our first approach was to use the Sobel filter to derive the image and extract the edge information, then combine it with the original image, thereby enhancing the corresponding information in the image. An example of this approach is displayed in Fig. 18, where the left is the original image, and the right is the image processed by the Sobel filter. Visually, we can observe that the corresponding features in the latter image are significantly enhanced, thus achieving our idea.

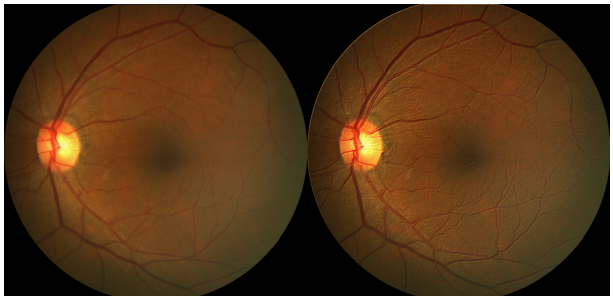


Figure 18: The original image and image processed by Sobel filter

4.6.2. Overall improvement of the image In the previous section, we discussed our approach to enhancing the image features using the Sobel filter. In this section, we focus on improving the overall quality of the training set images. We noticed that the quality of the images in the validation set was better than that of the training set. The validation set images had a brighter overall appearance and looked better in quality, while the training set images appeared darker and were of inferior quality. To address this issue, we attempted to improve the brightness and overall quality of the training set images to make them as consistent as possible with the validation set.

Histogram equalization: Our initial approach was to use histogram equalization. However, we found that this method was not suitable for retinal fundus images due to their unique pixel value distribution. Specifically, retinal fundus images have a large number of pixel values clustered in a certain range, which can cause significant color distortion when applying histogram equalization. As shown in Fig. 19, the left is the original image, and the right is the image processed by histogram equalization, we can see that the image after processing has serious color distortion. Therefore, we decided not to use histogram equalization for image enhancement.

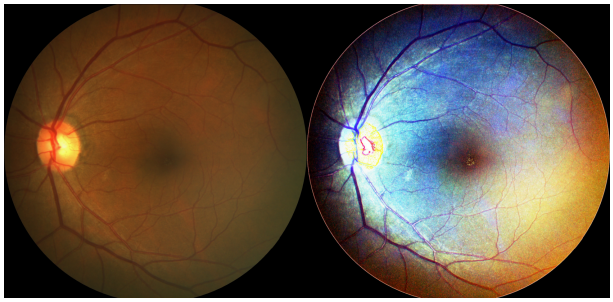


Figure 19: The original image and image processed by histogram equalization

Histogram stretch: The second approach we attempted was histogram stretching. However, we found that this method also led to color distortion, similar to the histogram equalization method. An example of the results is shown in Fig. 20 for the original image and the image processed by histogram stretch. Therefore, we abandoned this method as well.

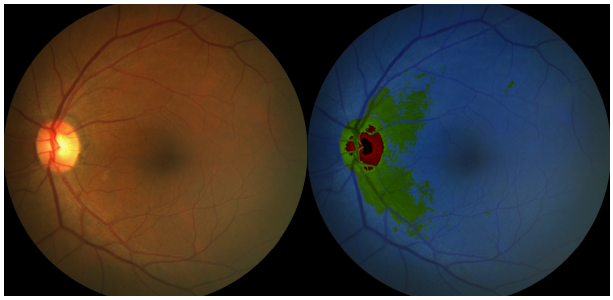


Figure 20: The original image and image processed by histogram stretch

Directly brighten the image: The final approach we attempted was to directly adjust the brightness of the image. The results obtained from this method of adjusting the brightness and saturation parameters were quite promising. Fig. 21 shows an example of this method, where the left is the original image, and the right is the image processed by directly adjusting the brightness. Consequently, we concluded that this image preprocessing method is suitable for our purpose.

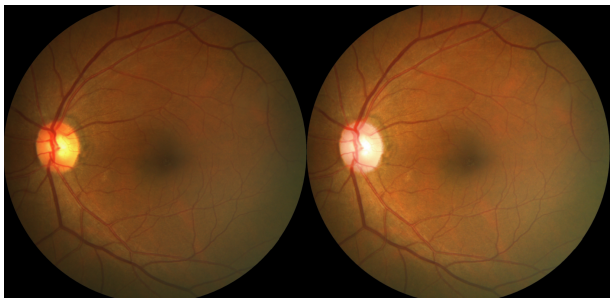


Figure 21: The original image and image processed by directly increasing the brightness

4.6.3. Result Feature enhancement: Moving on to the results of applying the image processing method to network training and testing, let's first discuss the application of the Sobel filter to the training data. We utilized this preprocessing method to train the GoogLeNet network in the baseline section. During the first 10 epochs of training, the test accuracy displayed values other than 0.9 and 0.1. In contrast, as

we learned in the previous report, training without using this method usually requires up to 30 epochs to obtain values other than 0.9 and 0.1 of test accuracy. Thus, the Sobel filter preprocessing method appears to have a positive effect. The first 10 epochs training result is shown in Fig. 22.

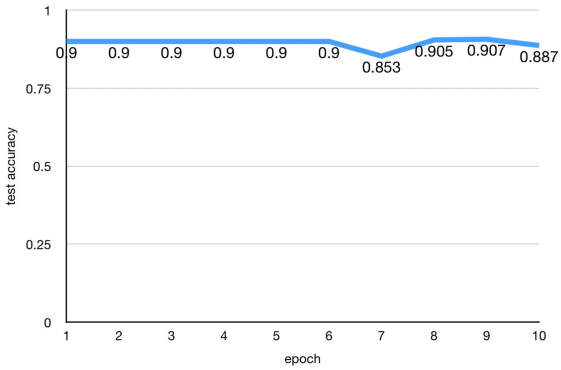


Figure 22: The first 10 epochs training result using Sobel filter

Overall improvement of the image: However, we did not observe any significant improvements in the overall method of enhancing the brightness of the training set images. In addition, we identified a potential issue, whereby increasing the brightness of the images may lead to overexposure of the optic disc and optic cup, making it more challenging to visually distinguish them. Fig. 23 is the example, the left is the original image in the training set, and the right is the image processed by directly increasing the brightness of the image. This is especially problematic because the ratio of the optic disc to the optic cup is a crucial factor in diagnosing glaucoma, and increasing the brightness of the training set images may result in the loss of relevant information.

Therefore to address this issue, we have altered our approach and decided to maintain the original brightness of the training set images while reducing the brightness of the images in the brighter validation set during testing. This would help to ensure that the optic disc and optic cup can be more easily distinguished visually. Fig. 24 provides an example of this method, where the left shows the original image in the validation set, and the right displays the image with reduced brightness. Interestingly, we have also observed that the typical features of glaucoma can be more easily distinguished visually with this method. Currently, we are using this approach during the accuracy testing phase, and we are still evaluating its effectiveness.

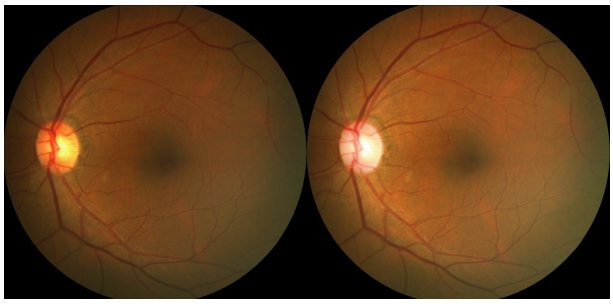


Figure 23: The original image in the training set and image processed by increasing brightness

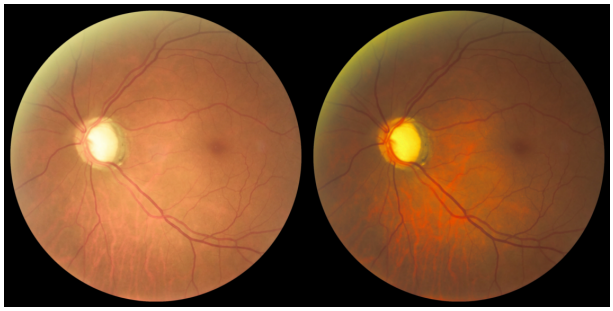


Figure 24: The original image in the validation set and image processed by reducing brightness

5. Conclusion

Plain GoogLeNet [26], ResNet18 and ResNet34 [18] did not perform well on REFUGE [22] dataset, and one cause might be the insufficient data for model to train. Some data augmentation methods could help to resolve the insufficiency and imbalance in dataset and therefore enhance the performance. And there is another possibility that the performance of these networks is indeed not enough for this task of glaucoma diagnosis.

BoTNet [20] introduced self-attention layer based on ResNet [18], which can extract the "important" region (optic cup and optic disc) in the fundus image, and got a better performance in calssification task.

As for the image preprocessing, the application of Sobel filter shows a certain effect in accelerating the convergence of the network by manually enhance the feature used for glaucoma diagnosis in the image. Meanwhile, the application of reducing the brightness of image in Refuge dataset [22] to highlight the optic disc and cup to improve the test accuracy is still being tested.

Further steps showed that classification model based on CDR (cup-to-disc ratio) performed much better than ResNet and BoTNet above. This showed us a direction of using CDR information and extracting feature together might be a good way to classify glaucoma images. But so far we only used the given mask image of optic cups and discs to obtain CDR, so this also required us to make more efforts on the segmentation of optic cups and optic discs.

6. Future Works

Although we already have a dataset on glaucoma, there are still some limitations. Firstly, our dataset only focuses on the Chinese patient population [22]. According to information found on the internet, the most serious glaucoma cases are found in Africa and Europe. Therefore, it is significant for us to increase the population range to cover different countries and areas.

In the field of glaucoma screening, manually analyzing retinal images for changes in the size and shape of the optic disc and optic cup can be time-consuming and subjective, leading to potential errors in diagnosis. Edge detection algorithms offer a potential solution by allowing for automatic identification and measurement of these structures with improved efficiency and accuracy. The edge detection results can be used to calculate the cup-to-disc ratio (CDR), which is an important indicator of glaucoma. However, accuracy can be further improved by training the model on datasets of groups of people of different ages who have glaucoma, in order to capture similarities and differences among the different age groups. In addition to improving the diagnosis of glaucoma, edge detection can also identify other abnormalities in the retinal images, such as nerve fiber layer defects, which are also important indicators of glaucoma. By analyzing the edges of these abnormalities, ophthalmologists can make more informed decisions about the diagnosis and treatment of glaucoma, as well as monitor changes in the optic disc and optic cup over time.

Division of Labor

Zifan Ying (Frank)

- Abstract
- Related Work
- Metrics
- Data Augmentation
- BoTNet
- ResNet34 [18]
- Self-attention model (BoTNet [20])
- CDR related methods
- (part of) Conclusion

Zhichong Wang (Max)

- Image processing methods in related work
- GoogLeNet in methods
- ResNet in methods
- GoogLeNet and ResNet in numerical experiments
- Image preprocessing in numerical experiments
- (part of) Conclusion

Hongbo Zhang (Howard)

- Introduction 1
- Datasets REFUGE24.1.1
- Feature extraction visualization 4.2.2
- ResNet34 in method A Resnet
- References Integration

Rongxuan Zhang (Jerry)

- Current Limitation of data
- The Improvement for model
- Future Works

Acknowledgement
Zifan Ying, Zhichong wang, Hongbo Zhang and Rongxuan Zhang contributed equally to this work and should be considered to co-first authers.

References

[1] D. Križaj, "What is glaucoma?," *Webvision: The Organization of the Retina and Visual System [Internet]*, 2019. 1
[2] V. Mahalakshmi and S. Karthikeyan, "Clustering based optic disc and optic cup segmentation for glaucoma detection," *Int. j. adv. res. comput. commun. eng.*, vol. 2, no. 4, pp. 3756–3761, 2014. 1
[3] R. Thomas and R. S. Parikh, "How to assess a patient for glaucoma," *Community Eye Health*, vol. 19, no. 59, pp. 36, 2006. 1
[4] W. Tang et al., "A global and patch-wise contrastive loss for accurate automated exudate detection," *arXiv preprint arXiv:2302.11517*, 2023. 1, 2.1
[5] K. U. Bartz-Schmidt et al., "Quantitative morphologic and functional evaluation of the optic nerve head in chronic open-angle glaucoma," *Surv. Ophthalmol.*, vol. 44, pp. S41–S53, 1999. 1
[6] M. Claro et al., "Automatic glaucoma detection based on optic disc segmentation and texture feature extraction," *clei Electron.J.*, vol. 19, no. 2, pp. 5–5, 2016. 1
[7] E. Miranda, M. Aryuni, and E. Irwansyah, "A survey of medical image classification techniques," in *2016 international conference on information management and technology (ICIMTech)*. IEEE, 2016, pp. 56–61. 2.1
[8] J. J. Gómez-Valverde et al., "Automatic glaucoma classification using color fundus images based on convolutional neural networks and transfer learning," *Biomed. Opt. Express*, vol. 10, no. 2, pp. 892–913, 2019. 2.1
[9] A. Issac, M. P. Sarathi, and M. K. Dutta, "An adaptive threshold based image processing technique for improved glaucoma detection and classification," *Comput. Methods. Programs. Biomed.*, vol. 122, no. 2, pp. 229–244, 2015. 2.1
[10] A. Li, J. Cheng, D. W. K. Wong, and J. Liu, "Integrating holistic and local deep features for glaucoma classification," in *2016 38th annual international conference of the IEEE engineering in medicine and biology society (EMBC)*. IEEE, 2016, pp. 1328–1331. 2.1
[11] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, 2017. 2.1
[12] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1–9. 2.1
[13] K. Simonyan and Andrew Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014. 2.1
[14] S. Wang, Z. Su, L. Ying, X. Peng, S. Zhu, and F. Liang, "Ieee 13th international symposium on biomedical imaging (isbi)," *IEEE, Prague, Czech Republic*, vol. 2016, pp. 514–517, 2016. 2.1
[15] N. K. Medathati and J. Sivaswamy, "Local descriptor based on texture of projections," in *Proceedings of the Seventh Indian Conference on Computer Vision, Graphics and Image Processing*, 2010, pp. 398–404. 2.1
[16] G. Csurka, C. Dance, L. Fan, J. Willamowski, and C. Bray, "Visual categorization with bags of keypoints," in *Workshop on statistical learning in computer vision, ECCV*. Prague, 2004. 2.1
[17] Rafael C. Gonzalez and Richard E. Woods, *Digital image processing*, Prentice Hall, Upper Saddle River, N.J., 2008. 2.2.1, 2.2.2, 2.2.3, 2.2.4
[18] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778. 3.1, 4.3.4, 5, 6
[19] A. Zhang, Z. C. Lipton, M. Li, and A. J. Smola, "Dive into deep learning," 2023. 1, 2, 3
[20] A. Srinivas, T.Y. Lin, N. Parmar, J. Shlens, P. Abbeel, and A. Vaswani, "Bottleneck transformers for visual recognition," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 16519–16529. 3.3, 3.3, 2, 4, 4.4, 14, 5, 6
[21] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," *Adv. Neural. Inf. Process. Syst.*, vol. 30, 2017. 3.3
[22] J. I. Orlando et al., "Refuge challenge: A unified framework for evaluating automated methods for glaucoma assessment from fundus photographs," *Med. Image. Anal.*, vol. 59, pp. 101570, 2020. 4.1.1, 5, 4.4, 5, 6
[23] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014. 4.3.4
[24] J. Deng et al., "Imagenet: A large-scale hierarchical image database," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 248–255. 4.3.4
[25] R. J. Casson, G. Chidlow, J. P. Wood, J. G. Crowston, and I. Goldberg, "Definition of glaucoma: clinical and experimental concepts," *Clin. Experiment. Ophthalmol.*, vol. 40, no. 4, pp. 341–349, 2012. 4.5
[26] C. Szegedy et al., "Going deeper with convolutions," 2014. 5