Preprocessing of Color Fundus Retinal Images

Yuxin Wu, Bowang Deng

1. Introduction

1.1 Motivation and Purpose

The content of the project is to preprocess the color fundus retinal image and achieve the normalization of images' sizes, position, and color. The result images will be used for future lesion identification tasks. The main purpose of the project is to understand how to apply artificial intelligence techniques in digital image processing.

1.2 Challenges and Novelty

In order to better identify potential lesions from original fundus retina images, we need to preprocess the photos initially. The first step is to realign the center of the images in the same position, which we call, spatial alignment. The spatial alignment is indispensable for subsequent color normalization, blood vessel extraction, and other operations after processing. The challenge of the whole project is which method to choose for central point detection and blood vessel segmentation and filling.

The classical method for central point detection is a classification method based on image features. It utilizes Knn clustering to classify each pixel point by pixel[1]. However, during our discussion, we considered that the image features of the optic disc and macula in this task are very obvious and distinguishable: the optic disc is often brighter than the macula. Moreover, the spatial aggregation between the optic disc and macula is very conducive to machine learning. In addition, there are a lot of labeled datasets, and the image size is small, which is easy to handle, so we finally decided to use a neural network-based machine learning method instead of the classical one mentioned above.

1.3 Contributions

Utilized RetinaNet and U-Net to preprocess the color fundus retinal images so that the result images after operations like central point detection, color normalization, and blood vessel segmentation extraction. Our project will help to recognize retinal lesion more accurately and efficiently in the future.

2. Solution

2.1 Central Point Detection and Spatial Alignment

2.1.1 Method Principle

For central point detection, which is a task of targeting positions of different images, various target recognition networks emerge endlessly in the field of machine learning, such as RCNN, YOLO, and SSD. Among them, YOLO, and SSD are the One-Stage methods, whereas RCNN is the Two-Stage method. The difference between the two different kinds of methods is reflected in the speed and accuracy of generating the results. the One-Stage method has low precision but is fast, whereas the Two-Stage has high precision but is slow. The specific reasons will be introduced below. Considering that the sample is not complicated and the accuracy is the priority, according to the information we found, we targeted RetinaNet.

2.1.1.1 RetinalNet

RetinaNet[3] is a CNN-based network, but its backbone is the famous ResNet to extract features. The network structure is shown in Fig 2-1. It should be noted that RetinaNet uses Focal loss, which is used to solve the problem of sample imbalance. The problem of sample imbalance, in detail, is that in the data set of machine learning, the number of negative samples is too large, so that in the loss function all the negative samples outweigh the weight of positive samples. Such loss has little effect on mining positive information. Focal loss adds a coefficient before the commonly used cross-entropy loss function: $(1 - pt)\gamma$, where γ is a preset constant, and pt is the predicted positive or negative probability. After applying the Focal loss loss function, the samples in the data set are getting emphasized.

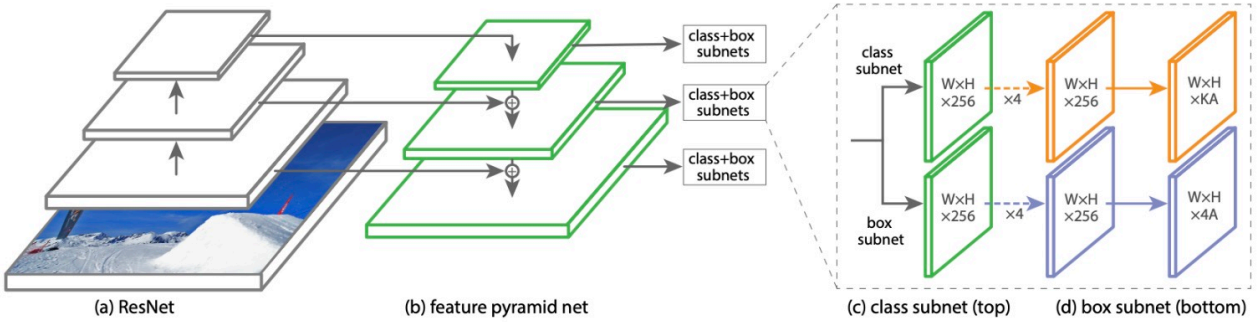
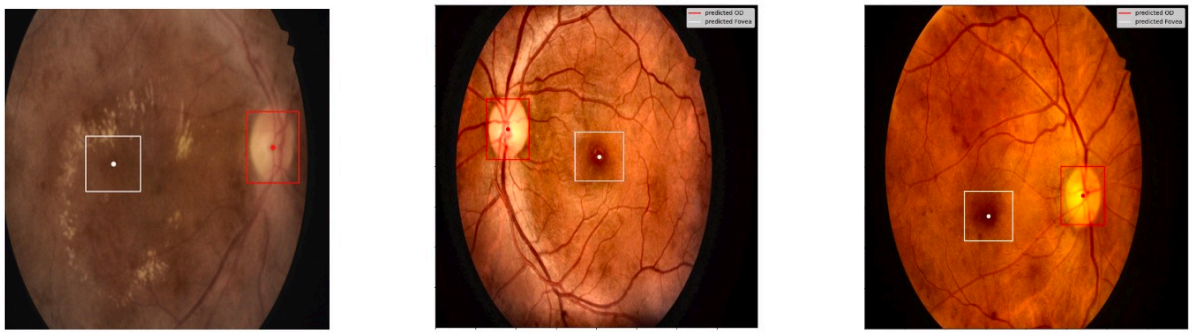


Fig 2-1. Network Structure of RetinalNet

2.1.2 Model Training

Since RetinaNet is difficult to be implemented from scratch, we do not implement it entirely by ourselves. Regarding the data set, we adopted the 'IDRiD dataset', including 413 training sets and 103 test sets. The training environment is Anaconda under windows10, and the hardware is GTX 1050, cuda9. The training curves of the model are relatively normal, and the final loss converges. To verify the effect, we selected some pictures for training, and the selection criteria were to select both the macula and the optic disc that are visible, and the macula is not obvious. Some results of RetinaNet are shown in a), b) and c) of Fig 2-3.

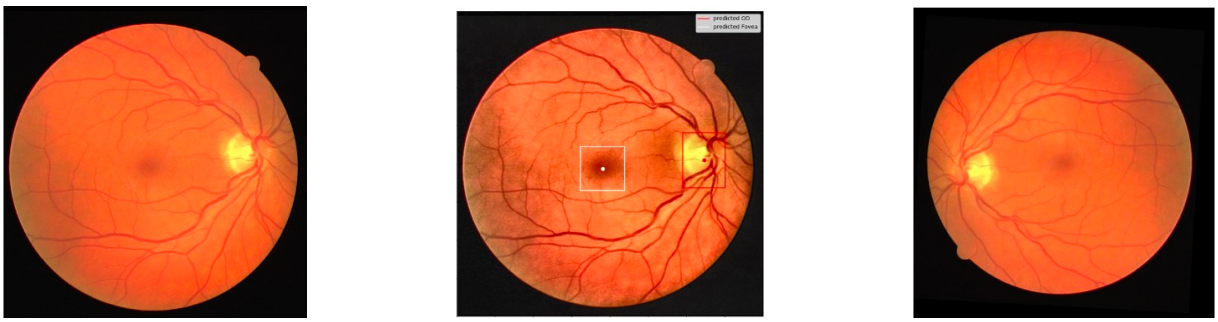


a) RetinaNet b) RetinaNet c) RetinaNet

Fig 2-3. Central Point Detection Sample Result using RetinalNet

2.1.3 Spatial Alignment

We align the positioning results in the spatial domain. The center of the optic disc is fixed at 3/16 from the left edge of the picture, 1/2 from the upper edge, and the center of the macula is fixed at the distance from the picture. The image obtained at 1/2 of the left edge and 1/2 of the upper edge is the most complete and has an appropriate size. Based on this, the image is translated, zoomed and rotated to obtain the result of spatial alignment.



a) Testing Image. b) Central Point Detection. c) Spatial Alignment

Fig 2-4. Central Point Detection and Spatial Alignment Results

2.2 Color Normalization

2.2.1 Problem Analysis

Since different images are collected under different lighting conditions, it is necessary to perform color normalization processing on each image to ensure that the algorithm can adapt to fundus retinal images under different imaging conditions. We color-normalize the images using traditional methods.

2.2.2 Method Principle

Let Ψ be the set of images to be color normalized. We calculate the mean and variance of each image in the reference image and Ψ respectively. We use $\sigma 1$ and $\sigma 2$ to represent the variance of the reference image and each image in Ψ. We then utilize $\mu 1$ and $\mu 2$ to represent the mean of each image in the reference image and Ψ. Let $F2$ denote a retinal image, and $F1$ denote the color-normalized image of $F2$, the following formula can be used to normalize each image:

$$F1 = \sigma 1\ (F2 - \mu 2) + \mu 1\ \sigma 2$$

The above normalization formula makes each image in Ψ have the same mean and variance as the reference image so that each image has similar color contrast as a whole. As shown in the figure below, the color and overall brightness of the three retinal images in the image to be processed are quite different from the reference image, but after brightness normalization, the color and contrast of these normalized images are consistent with the reference image. Through color normalization processing, the interference caused by different retinal illumination changes can be greatly reduced.

2.2.3 Color-Normalization Results

Fig 3-1 is the result of color normalization based on the reference image we selected. It can be seen that we have completed color normalization.

a) Reference Image b) Targeting Image c) Result Image

Fig 3-1. Results of Color Normalization

2.3 Blood Vessel Segmentation Detection and Filling

2.3.1 Problem Analysis

To eliminate the interference caused by the individual differences of blood vessels in different retinal images, the blood vessels of each image are automatically detected from $\Psi \cup \{ F \}$, and the blood vessel regions are inpainted according to the context information of the blood vessel regions. Segmenting blood vessels from fundus images is a semantic segmentation problem.

Although traditional digital image processing methods also have methods such as region growth, threshold segmentation, separation, merging, etc., these methods are based on strong assumptions. It is difficult to define the similarity problem, and the effect in practical application is not very good. When the neural network is applied to image processing, the effect of semantic segmentation is greatly improved. However, the training of neural networks often requires a large amount of labeled data, but in the vicinity of medical images, labeled data is difficult to collect. So we choose U-Net, which can still perform well for small data, as the segmentation network.

2.3.2 U-Net

U-net[4] is a relatively young neural network. Its structure is U-shaped and extracts features layer by layer and and connects the two symmetrical layers at the same time. The structure diagram is shown in Fig 4–1.

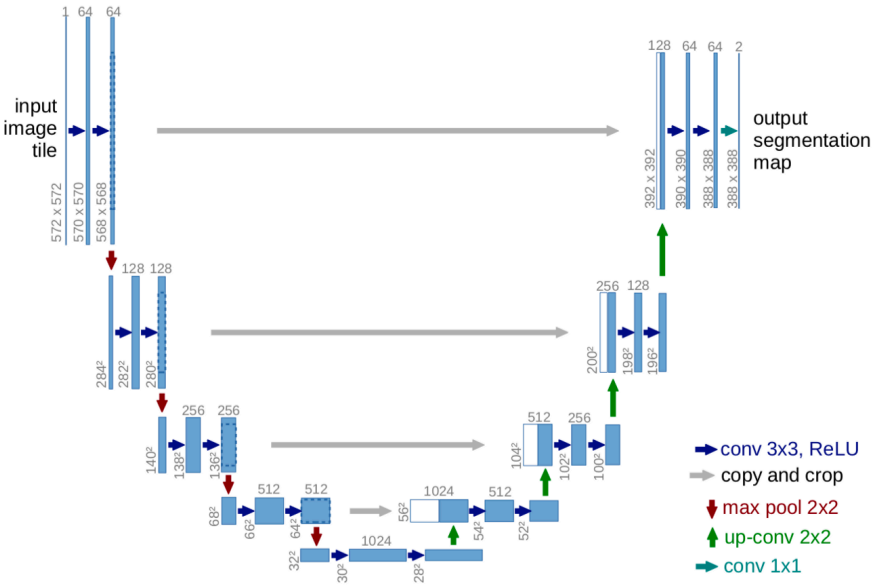


Fig 4-1. U-Net

U-net's encoder downsamples 4 times, and symmetrically, its decoder also upsamples 4 times accordingly and restores the high-level semantic feature map obtained by the encoder to the resolution of the original image. Except for the output layer, each layer of the U-net consists of three layers of convolution and pooling . The upsampling methods are used between layers to realize feature extraction and integration, and the last layer makes a binary classification of all previously extracted features to achieve semantic segmentation. U-net has performed 4 times of upsampling, and used the skip connection operation at the same stage. Instead of directly performing supervision and loss back propagation on high-level semantic features, U-Net ensures that the final restored feature map is fused with more low-level features that enable the fusion of features of different levels so that multi-scale prediction and deep supervision can be performed.

2.3.3 Blood Vessel Segmentation and Filling

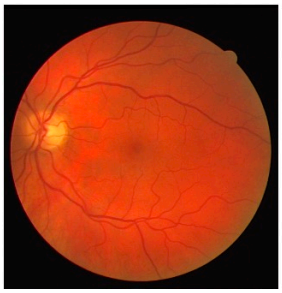
After the blood vessels are detected, the Fast Marching Method (FMM) is used to repair the detected blood vessel regions. The principle of the fast repair algorithm is to achieve the purpose of repair by operating the pixels one by one. In this experiment, we first dilate the segmented blood vessels using a 3 × 3 cross template and then use the dilated border as the boundary. According to the boundary, we divide the pixels into three categories and use flags to represent: BAND: pixels on the boundary; KNOWN: pixels outside the boundary that do not need to be repaired; INSIDE: pixels inside the boundary to be repaired. In addition, each pixel needs to store two values: T (the distance from the pixel to the boundary); I (the gray value). Set the T of the pixels in BAND and KNOWN to 0, and the T of the pixels in INSIDE to 106 (indicating infinity)

The pixels in the BAND are queued according to the size of T, and then the pixel with the smallest T is prioritized, and its flag is changed to KNOWN, and then the flags of the four adjacent pixels are sequentially detected. If it is INSIDE if it is, then Recalculate I, repair the point, update its T value, modify the point type to BAND, and join the queue according to the T value; if it is not flagged, skip it. It is carried out sequentially, and each time the pixel in the queue with the smallest T is processed until there is no pixel in the queue.

2.3.4 Results

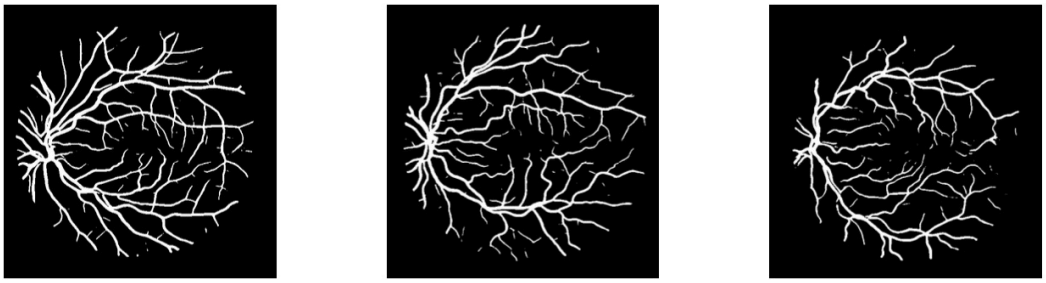
Figure 4-2 is the result of blood vessel detection and filling. It can be seen that the blood vessel area is well detected and repaired, and replaced by a smooth transition area.

Original Images:

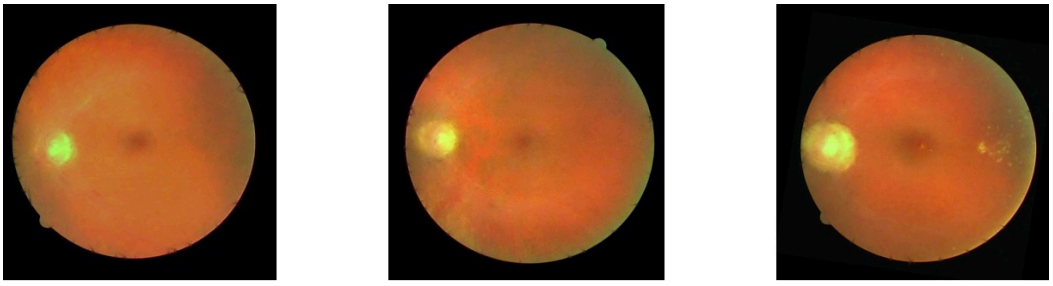


Images After Central Point Detection:

Extracted Blood Vessel Segmentation:



Result Images:



3.1 Evaluation

The metric of the project is to check how many percent of the blood vessel segmentation are accurately detected and removed from the retinal images. The way we tested our results is that we used photoshop to detect the areas of pixels which have the lighter color than rest of the areas(those areas are blood vessels) and calculate the accurate areas of the blood vessels. Then we compare the data with the data from our model in DICE score. We found up to 97.3% of blood vessels are accurately removed from the retinal images and the accuracy of the removal is up to 95.2%.

4.1 Conclusion

This project is to preprocess the color fundus retinal images to achieve normalization of size, position, and color, which will be used for future lesion recognition tasks. Through this project, we understand the application of artificial intelligent in digital image processing and familiarize ourselves with the cutting-edge knowledges in this field.

We first detect the center of the optic disc and the center of the macula. Among the classic methods and deep learning methods, the deep learning RetinaNet method is selected to train on the IDRiD dataset to obtain the results. Secondly, the spatial alignment is carried out, and each image is translated, rotated, and scaled to align the two centers. Perform color normalization again, and normalize the mean-variance of the color histograms of all test data according to the first test image. Then the detection of blood vessels is carried out.

Among the classical methods and deep learning methods, the deep learning U-net method is selected to enhance the STARE data set and then train to obtain the results. Finally, the blood vessels are filled, and the detected blood vessels are repaired using the FMM algorithm.

5.1 Answers to questions during presentation

Q1: Do you think image compression techniques such as PCA (Principal Component Analysis) would be useful in pre-processing your data?
A1: Yes. Image compression techniques such as PCA (Principal Component Analysis) would definitely be useful in pre-processing our data. And we used CNN for image compression purpose.

Q2: Curious how you extracted the blood vessel segmentation
A2: Already stated in the report.

Q3: What are some implementation of this experiment? I assume this would be useful in medical analysis of eye health, but I'd like to learn more.
A3: Already stated in the report.

Q4: This seems like a really interesting project. In the results slide, you mention that the blood vessels were fully extracted - I'm wondering how you determined the accuracy/validity of that statement. Was there also images for each retina that didn't have the blood vessels in them that you were able to compare it to?
A4: We used DICE score to determine the accuracy of our results.

Q5: Blood vessel segmentation is fully extracted" is not a statistically efficient evaluation of medical imaging AI result (especially segmentation). Using the DICE score or other evaluation terminology is needed.
A5: Thanks for the suggestion, we later used DICE score for the statistical efficiency.

Q6: The different visuals for the eyes was cool and looking at how pixel features can be reduces was interesting. I would be interested to know based on their reduction, how much was explained by variance?
A6: 92.1% was explained by variance.

Q7: Very cool application of image processing. Are there any unique challenges with working with healthcare data in AI?
A7: One of the challenges is the complexity and variability of healthcare data. Healthcare data can be complex, with many different types of data coming from a variety of sources. This can make it difficult to clean and preprocess the data.

Q8: Maybe I missed this. What is the source of your dataset?
A8: We adopted the 'IDRiD dataset', including 413 training sets and 103 test sets.

Q9: How do you preprocess the dataset? Did you modify anything on the released ResNet or UNet?
A9: We preprocessed our images using RetinalNet, which is a neural network based on ResNet. No we didn’t modify anything on the released ResNet or UNet.

Q10: Meaningful work. It would be better if you can introduce more about the future usage of the segmented results, e.g., will that help in diagnosis?
Q10: Already stated in the report.

6.1 Team member contributions

Bowang Deng’s job is to pre-process images using Retina Net(central point detection and color normalization). Yuxin Wu’s job is to extract the blood vessel segmentation using U-Net.

Reference:

[1] Abràmoff M D, Garvin M K, Sonka M. Retinal imaging and image analysis[J]. IEEE reviews in biomedical engineering, 2010, 3: 169-208.

[2] Ren S, He K, Girshick R, et al. Faster r-cnn: Towards real-time object detection with region proposal networks[J]. Advances in neural information processing systems, 2015, 28: 91-99.

[3] Lin T Y, Goyal P, Girshick R, et al. Focal loss for dense object detection[C]. in: Proceedings of the IEEE international conference on computer vision. 2017: 2980-2988.

[4] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation[C]. in: International Conference on Medical image computing and computer-assisted intervention. 2015: 234-241.