COMP9517 Computer Vision Project Report T3 2019

Zedong Li, Hanyu Liang, Zhiwei Wang, Xingsong Wang, Jiayi Cui

***Abstract*— Diabetic retinopathy, known as non-proliferative diabetic retinopathy, is a major feature of the early stage, including intraretinal hemorrhage and fluid inflow into the retina. In the next stage, diabetic retinopathy includes the growth of new abnormal blood vessels on the surface of retina. For diagnosing diabetic retina, the blood vessel and some features of DR detection and segmentation play an important role in DR detection and classification or grading. There are 54 original retinal images in JPG format from IDRiD database are used to evaluate the proposed algorithm.**

I. INTRODUCTION

THIS report is to identify Diabetic Retinopathy(DR) in fundus images by computer vision and artificial intelligence. According to the World Health Organization, the number of DR patients worldwide will increase to 366 million by 2030. [1] As a result, this project is to predict and prevent the Diabetic Retinopathy used by computer vision.

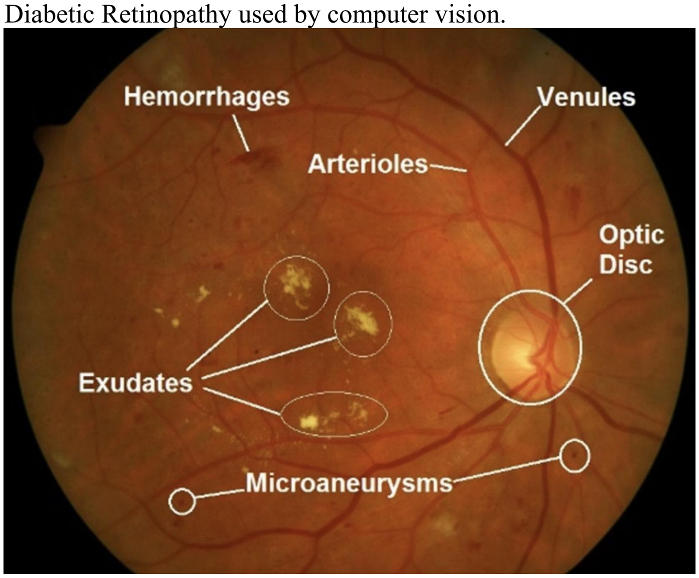


Fig. 1: The High-Resolution Fundus (HRF) Image

Diabetic retinopathy, known as non-proliferative diabetic retinopathy, is a major feature of the early stage, including intraretinal hemorrhage and fluid inflow into the retina. In the next stage, diabetic retinopathy includes the growth of new abnormal blood vessels on the surface of retina.

The fig. 1 is the original digital color retinal image, it is shown that the structure of the retinal. It consists of, as marked in figure 1, hemorrhages, arterioles, venules, exudates, optic disc and microaneurysms. Microaneurysms were described as circular, small, red dots while hemorrhages had dot-like blot or flame. An exudate is yellow and any fluid that filters from the circulatory system into lesions or areas of inflammation.

In this project, the purpose is to finish segment abnormal lesions associated with Diabetic Retinopathy (task 1) and Segment retinal blood vessels (task 2).

In this report, we briefly review a few articles which selected in the course of this project, and then the next section will present the proposed method for the two tasks, what difficulties we have encountered in these methods and how to solve them. In the next section, it will talk about the environment that this project needs to use and how to debug it. Results and evaluation are presented in “Results” section. At last, “Conclusion and future work” section is devoted to summing up what we should be going on in the area of computer version and concluding remarks.

II. LITERATURE SURVEY

Almotiri[2] used automatic detection, localization and extraction of three retinal anatomical structures to design a generic system, which is for using a hybrid of fuzzy set theory and morphological operations.

Popli [3]proposed a new method to detect the fundus images which can augment the data and provide the small patches of the image with single large image. The method of Unet proposed in this article is indeed an effective method, which provides a good idea for the completion of our project.

Carmona et al. [4] presented a novel way which genetic algorithm is used to approximate the optic nerve disk in elliptical image. This method is very useful in preprocessing.

A number of other useful and neoteric algorithms can be found in the literature or the newest learning periodical that used Unet and FCM algorithm. According to an article in IEEE Transactions on Medical Imaging [5], the research team found that doctors will also encounter problems in our project when they accurately segment the dirt and tumor. And they came up with a solution, UNET, for the precise segmentation of the liver and the tumor.

Some methods are based on the Contrast Limited Adaptive Histogram Equalization (CLAHE) which is capable of the enhancement of color retinal image. The purpose of this way is to reduce the noise effect in color retinal image.

III. METHODS

A. Task 1

1) U-Net

U-Net is no doubt one of the most successful methods in image segmentation, especially in medical image segmentation. The encoder (down-sampling)-decoder (up-sampling) structure and the jump connection used are a quite classical design method. At present, many novel design methods of the convolution neural network are available, but many still extend the core idea of U-Net, add new modules or

integrate into other design concepts. In this project, we have improved the traditional U-net to a certain extent.
The High-Resolution Fundus (HRF) Image

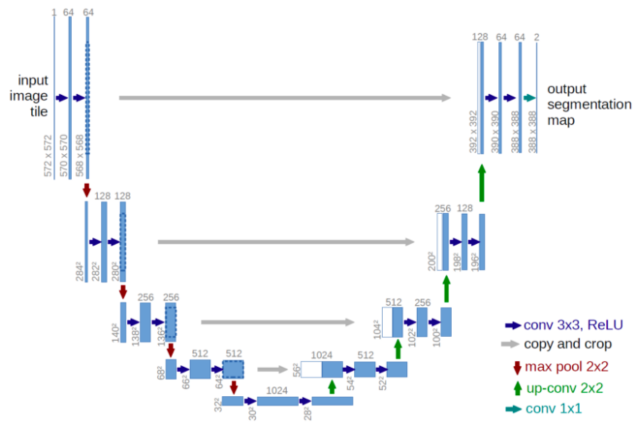


Fig. 2: Traditional U-net Constructure

The structure of U-Net is shown in the Fig.2. From the top to bottom, the left side(input) can be regarded as an encoder and the right side(output) as a decoder. The encoder has four sub-modules, each of which contains two convolution layers, followed by a lower sampling layer implemented by max pool. The resolution of the input image is 572x572, and the resolutions of the first and fifth modules are 572x572, 284x284, 140x140, 68x68 and 32x32, respectively. Because convolution uses valid mode, the resolution of the latter submodule here is equal to (the resolution of the previous submodule - 4) / 2. The decoder consists of four sub-modules, and the resolution increases in turn through the up-sampling operation until it is consistent with the resolution of the input image (because the convolution uses valid mode, the actual output is smaller than the input image). The network also uses a jump connection to connect the up-sampling results with the output of the sub-module with the same resolution in the encoder as the input of the next sub-module in the decoder.

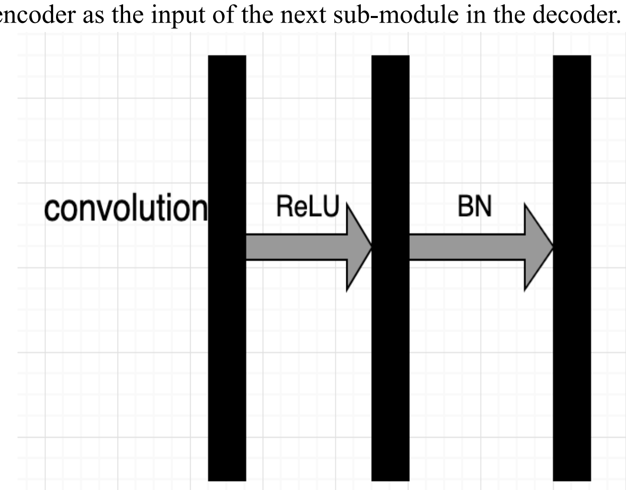
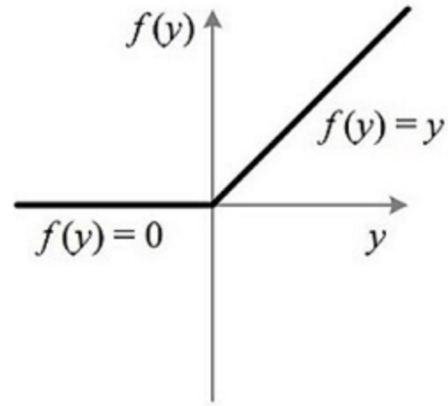


Fig. 3 The difference between traditional convolution

But in this project, compared with the traditional Unet structure, we add some novel ideas to the input. Shown in Fig.3, it can be seen that compared with traditional convolution, our input appended a Rectified Linear Unit (ReLU) activation function and a BN layer. The purpose of the action function ReLU is to better mine the relevant features and fit the training data.



$$ReLU(x) = \begin{cases} x & if\, x > 0 \\ 0 & if\, x \leq 0 \end{cases}$$

Fig. 4 The ReLU Function

Compared with other activation functions, ReLU (Fig 4) has the following advantages: for linear functions, ReLU is more expressive, especially in deep networks; for nonlinear functions, ReLU has no Vanishing Gradient Problem because the gradient of nonnegative interval is constant, which keeps the convergence rate of the model in a stable state. here is a slight description of what is the gradient vanishing problem: when the gradient is less than 1, the error between the predicted value and the true value attenuates every propagating layer. if sigmoid is used as the activation function in the deep model, this phenomenon is particularly obvious and will lead to model convergence stagnation. Besides standard layers，BN layer is introduced for normalization and avoid gradient explosion.

2) FCM Method
step1:preprocess the image:use the clahe on green channel to enhance the contract of image.
step2:Use the FCM algorithm to process image.The algorithm is shows below:

$$J_m(U, v) = \sum_{k=1}^{N} \sum_{i=1}^{c} (U_i k)^m d_i k \tag{1}$$

In figure 5, based on center of cluster, The FCM separate image to some part.
The FCM is different with other cluster methods. It shows the probability of each pixel belonging to every clustering center respectively. After FCM，use Gaussian Bell function and cross correlation to smooth image.
step3: postprocess. Use the morphology filter to handle image. Opening and closing. Opening is the process of erosion and dilation. It is used to remove small element to make the image

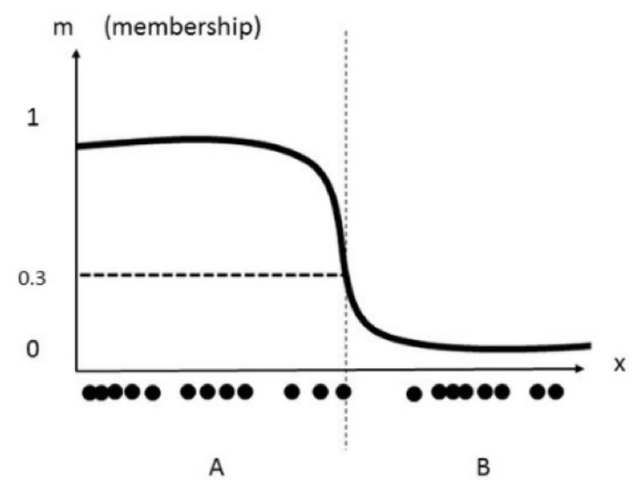


Fig. 5 FCM

more clear. Closing is the process of dilation and erosion. It is used to fill the hole to make the image more complete.

B.Task 2

1) Contrast Limited Adaptive Histogram Equalization(CLAHE)

CLAHE is a method to enhance image contrast. Its main idea is to transform the histogram distribution of an image into an approximate uniform distribution, so as to enhance the contrast of the image. Although histogram equalization is only the basic method in digital image processing (Digital Image Processing), it plays a very powerful role and is a very classical algorithm.

The basic principle of histogram equalization is described here. Suppose it has an image A, its histogram distribution $H_A(D)$. Using a monotone non-linear mapping $f: R \rightarrow R$ to change the image A to the image B, that is, to apply the transformation f to each pixel point in the image A, and the histogram of the image B is $H_B(D)$. The whole process can be described in accordance with the illustration of Figure 5: the lower right of the figure is the gray histogram distribution of the A-image (which is convenient for drawing, here is a continuous distribution), and the upper right of the figure is a monotone non-linear transformation function f, and the histogram distribution of the image B obtained in the upper left mode, as follows:

$$D_B + \Delta D_B = f(D_A + \Delta D_A) \, , \, D_B = f(D_A) \tag{2}$$

That is, it is to be understood that the function of[formula] is to change the gray scale of the pixel point in the A image to the D_A, then it is:

$$\int_{D_A}^{D_A+\Delta D_A} H_A(D) dD = \int_{D_B}^{D_B+\Delta D_B} H_B(D) dD \tag{3}$$

The above formula can be understood to be the same as the total number of pixel points in the corresponding interval. In order to achieve a histogram equalization, there are special:

$$\int_0^{D_A} H_A(D) dD = \int_0^{D_B} H_B(D) dD \tag{4}$$

Because the target is a uniform distribution of the histogram, then the ideal $H_B(D) = \frac{A_0}{L}$, A_0 is the number of pixel points, L is the gray-scale depth, and 256 is usually taken. it can be obtained as follows:

$$\int_0^{D_A} H_A(D) dD = \frac{A_0 D_B}{L} = \frac{A_0 f(D_A)}{L} \tag{5}$$

So, $f$can be found, and the result is:

$$f(D_A) = \frac{L}{A_0} \int_0^{D_A} H_A(D) dD \tag{6}$$

The discrete form is:

$$f(D_A) = \frac{L}{A_0} \sum_{u=0}^{D_A} H_A(u) \tag{7}$$

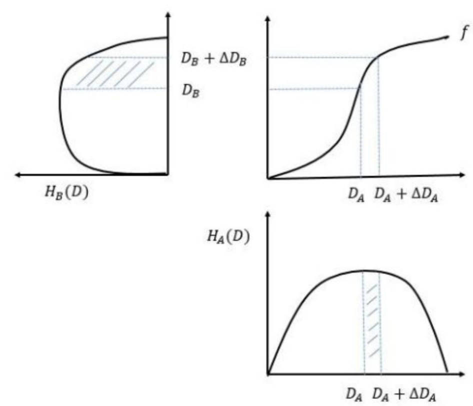


Fig. 6 Histogram Equalization Schematic Diagram

It can be seen from the above derivation that the mapping function f is closely related to CDF. Moreover, the derivation process is done on the continuous distribution, but in fact, the gray level of the image is usually 256, so it is a kind of continuous distribution approximate to the discrete case. If the gray level is high enough, the result will be a really uniform distribution histogram, because the gray level is only 256, so the histogram obtained in the actual situation is often not uniform distribution, but approximate uniform distribution. At the same time, if the gray histogram of image A changes dramatically and there are no pixel points in some gray intervals, which will cause the sharp change of CDF, then the gray level interval (x, x+1) may change violently, which leads to the great inhomogeneity of the histogram of the final B image. However, in the actual use process, the obtained image can be approximately evenly distributed has achieved a better contrast enhancement effect.

In this task, after extracting green channel, the image has a greater contrast. So the Contrast Limited Adaptive Histogram Equalization (CLAHE) is used to further increase contrast.

2) Morphological Methods

Morphological methods usually refer to a branch of biology that deals with the morphology and structure of animals and plants. The morphology of our image processing refers to mathematical morphology. The basic operations include binary corrosion and expansion, binary open and close operation, hit miss transformation, watershed transformation, gray value corrosion and expansion, grey value opening and closing operation, gray value morphological gradient and so on.

In this task, erosion, dilation, opening and closing could be used when processing the image. Erosion and Dilation are the fundamental morphological operation. The Erosion is to find the maximum value for the local, but the dilation is the opposite. An opening is defined as the erosion of both operations using the same structural elements, followed by dilation. Similarly, the closing is the opposite.

IV. Experimental Setup

A.Experimental environment:

python 3.6

B.Required libraries:

Task 1: module numpy, torch, cv2, matplotlib, sklearn, torchversion, keras.utils

Task 2: module cv2, numpy, os and csv

C.Experiment process:

1) Task 1:

In weight adjustment, here are two methods:

(1) Through the image showing, the contour observation of the image is observed, the weight and the size are adjusted, the disadvantage is that the time spent is relatively long, a large amount of time is needed. the weight needs to be increased when the profile of the image is not clear. At last, it could be found the suitable weight.

(2) From the recall, precision and other values of the image, it can infer the adjustment of weight. This method is easier than before, but it also needs to use experience and data to determine how to adjust the weight value. The smaller the value, the more positive weight increase.

2) Task 2:

Step 1: Extracting green channel and analysis the result image. It could be found that it has greater contrast. So Contrast Limited Adaptive Histogram Equalization (CLAHE) could be applied to further increase contrast. And the result is in Fig.6.

Step 2: As the second image of figure 2 shows, a filter is used to process the image, actually it is three times closing and opening.

Step 3: It can subtract image after Step2 from the first image and get faint traces of blood vessels with optic discs and other things removed.

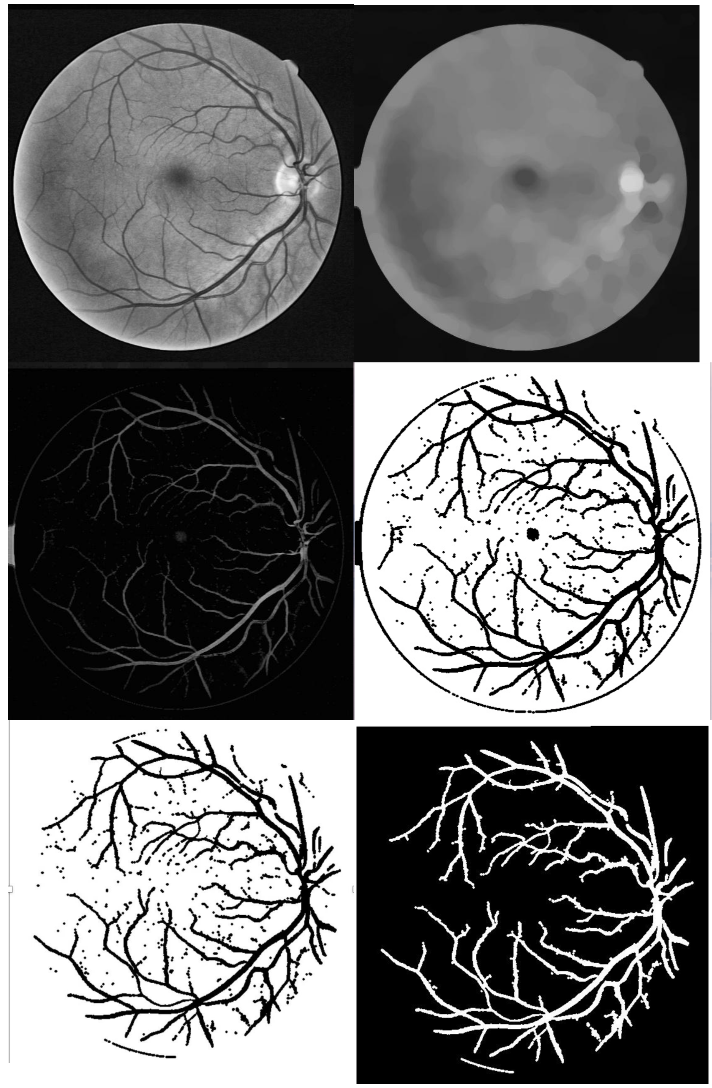


Fig. 7 Histogram Equalization Schematic Diagram

Step 4: Based on the last image, Binarization is used to obtain the processed result which shows as the fourth image in figure 7.

Step 5: The center of the image is dark, so there will be a circle in the center of the proceeded image. The functions which can find the counter are used to remove the circle.

Step 6: There still are some noise in the processed image. It can use Erosion to get a much clearer result. The final image is shown as the last image in figure 7.

Similarly, it should refer to the value of precision, recall and accuracy. But accuracy is not of value reference, because when the image is extremely dark, the accuracy is still so high. As a result, the value of precision and recall are the key of the success.

V. RESULTS

A.Task 1:

As the Fig.8 shown that, the prediction and recall can up to 70%. And from the image, the main features are also quite clear.

Fig. 8 Result of Soft_Exudates

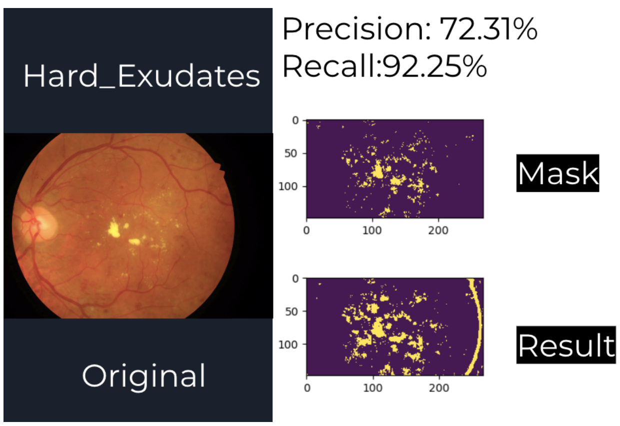


Fig. 9 Result of Hard_Exudates

The recall of fig.9 is more than 90% but its prediction is just approximately 72%.

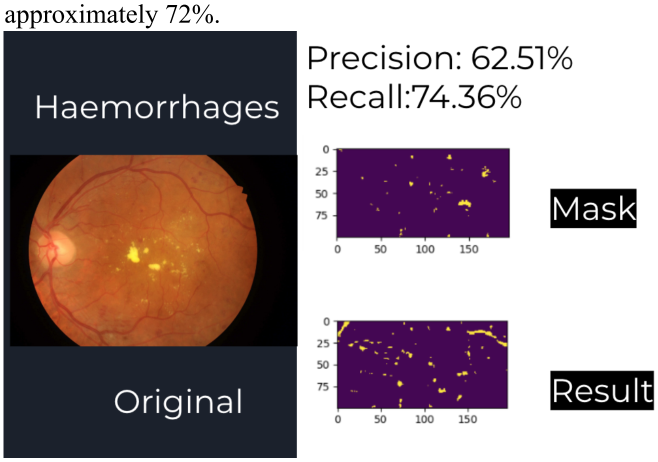


Fig. 10 Result of Haemorrhages

It can be found in Fig.10 that precision is smaller than before but recall rate take the same. The result of

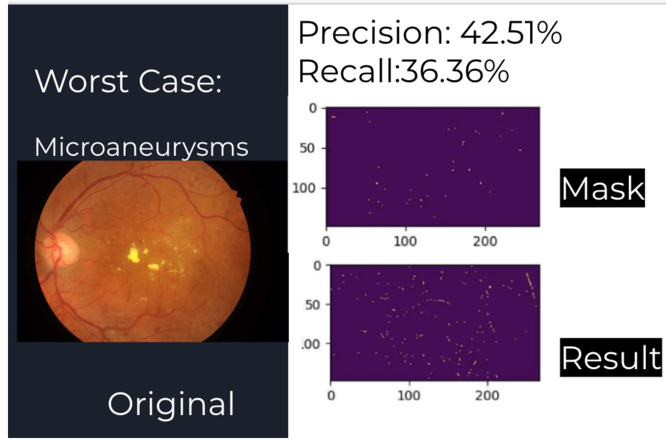


Fig. 11 Result of Microaneurysms

microaneurysms is shown in Fig.11. And the recall score less than 40% and precision score less than 45%.

In conclusion, it can get quite a good predicted result of soft and hard exudates because deep learning was applied for task 1. But there is a curve in the predicted image which lead to a lower precision. In other two symptoms, as the amount of positive examples is much less than negative. So it cannot learn features of them which lead to inaccurate results.

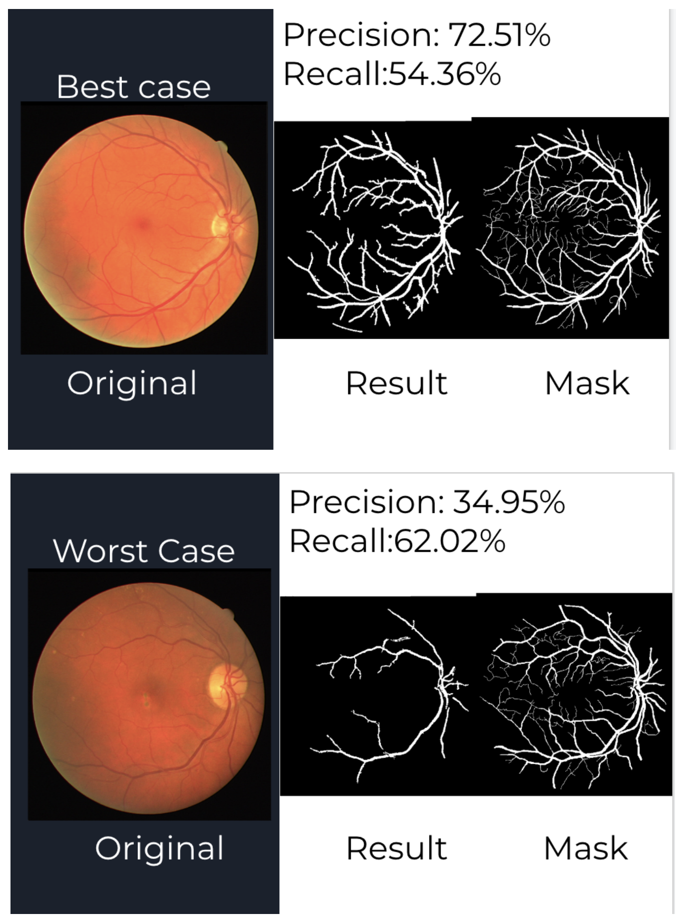


Fig. 12 Result of Task2

B.Task 2:

Two images in Fig.12 are both the result diagrams obtained by the traditional image processing method, rather than the use of deep learning.

The first image in Fig.12 is the best result of task2, and its precision is 72% and recall is above 50%. Most of the blood vessels can be found, meanwhile a lot of little branches are lost. And in the worst case, it just got a little bit of the blood vessels. The reason is that this picture is dark overall, so the contrast is small. This lead to many blood vessels are cut off through the processing. So, the precision is very low, just 34%.

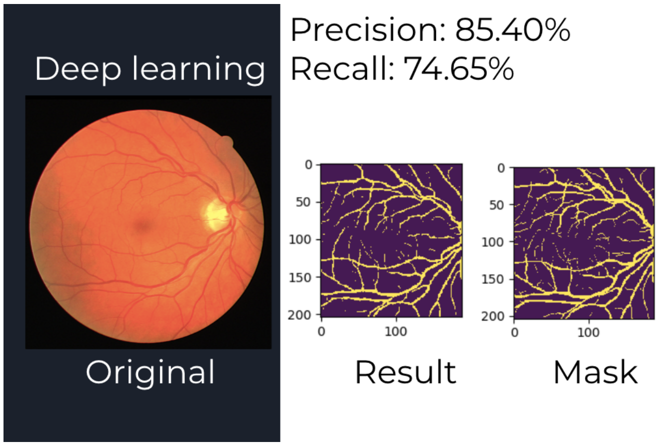


Fig. 13 Deep Learning Result of Task 2

Table 1: Evaluation Metric of task1

	HE	EX	MA	SE
Ave_Recall	0.73	0.92	0.36	0.70
Ave_ Precision	0.62	0.72	0.42	0.76

Table 2: Evaluation Metric of task2

	DL	Traditional
Ave_Recall	0.72	0.55
Ave_ Precision	0.82	0.67

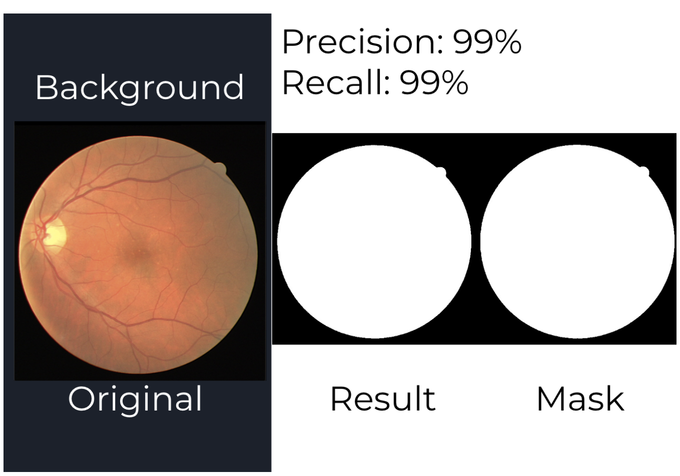


Fig. 14 result of background

Deep learning method is also be used to segment blood vessels, it is clear that its result is better than the traditional method showing in the Table 2. It could prevent the loss of small blood vessels. Deep learning method get greater average precision. Because it could learn the vessel features at multiple scales and levels so it could find the little branches which are not clear in the image.

In Fig.14, using Binarization to segment the background can get a result which is very close to the mask.

VI. Discusion

A.Limitations

1) Traditional method

As Fig.15 shows, there are many tiny branches are lost, some of them are disappear due to using erosion, and some are cut off through the processing as their small contrast. This leads to a lower accuracy.

2) Deep learning

In the U-Net model, the predicted results are inaccurate like Fig.10 shows. The reason is that the original images are impressed when they are used for training the model. It could decrease the run time. However, many details are lost through the compress , hence the model can not learn the features of symptoms well.

B.Solutions

1) Traditional method

Many methods could be used to find the little blood vessels. They can process the retinal images in which the branched have small contrast. Preventing the loss of little branches could improve the accuracy of results, but the extra processing may increase the run time. Furthermore, the deep learning method is also a good choice.

2) Deep learning

The better way to avoid the loss of some details is using sliding window or more accurate patches on training. It could improve the predicted results though it may increase the runtime.

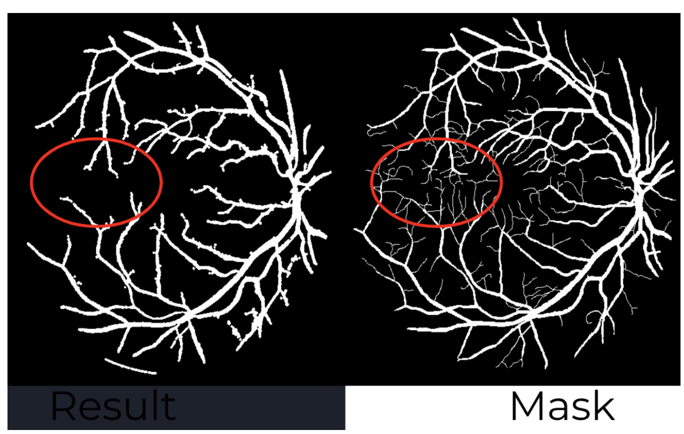


Fig. 15 The traditional method of segment

VII. CONCLUSION AND FUTURE WORK

To sum up, deep learning method is used in task. U-Net model is good to learn features and find those symptoms. But it still needs to apply for a better way to train our model. For instance, sliding window is a good way to deal with this issue.

In task2, traditional methods are used to segment the blood vessels. Due to the loss of some branches, the average precision and recall are lower than deep learning method. The reason is that the deep learning model can learn vessel features at multiple scales and levels. As a result, deep learning method can prevent the loss of branches to obtain more accurate results.

VIII. CONTRIBUTE OF GROUP MEMBERS

1.Hanyu Liang: make the powerpoint presentation and participate in demo and presentation. research the traditional method for task2. write the group report

2.Zhiwei Wang: make the powerpoint presentation and participate in demo and presentation. research the traditional method for task2. write the group report

3.Xingsong Wang: make the powerpoint presentation and participate in demo and presentation.research the U-net and fcm method for task1 and task2.write the group report

4.Zedong Li: make the powerpoint presentation and participate in demo and presentation.write the group report

5.Jiayi Cui: make the powerpoint presentation and participate in demo and presentation.write the group report

REFERENCES

[1] C. Bradley, Gamsu, D. and (1994), Guidelines for Encouraging Psychological Well‐being. Diabetic Medicine, 11: 510-516. doi:10.1111/j.1464-5491.1994.tb00316.x W.-K. Chen, Linear Networks and Systems. Belmont, CA, USA: Wadsworth, 1993, pp. 123–135.

[2] J. Almotiri, K. Elleithy and A. Elleithy, "A Multi-Anatomical Retinal Structure Segmentation System for Automatic Eye Screening Using Morphological Adaptive Fuzzy Thresholding," in IEEE Journal of Translational Engineering in Health and Medicine, vol. 6, pp. 1-23, 2018, Art no. 3800123.

[3] A. Popli, G. Jindal, G. Pillai, H. Khan, M. Agarwal and V. Yadav, Automated hard exudates segmentation in retinal images using patch based UNET, Indian Institute Of Technology Roorkee, 2017

[4] EJ Carmona, M Rincón, J García-Feijoo, JM Martínez de-la-Casa, Identification of the optic nerve head with genetic algorithms. Artif. Intell. Med. 43(3), 243–259 (2008)

[5] X. Li, H. Chen, X. Qi, Q. Dou, C. Fu and P. Heng, "H-DenseUNet: Hybrid Densely Connected UNet for Liver and Tumor Segmentation From CT Volumes," in IEEE Transactions on Medical Imaging, vol. 37, no. 12, pp. 2663-2674, Dec. 2018.

[6] A. W. Setiawan, T. R. Mengko, O. S. Santoso and A. B. Suksmono, "Color retinal image enhancement using CLAHE," International Conference on ICT for Smart Society, Jakarta, 2013, pp. 1-3.

[7] J. Almotiri, K. Elleithy and A. Elleithy, "A Multi-Anatomical Retinal Structure Segmentation System for Automatic Eye Screening Using Morphological Adaptive Fuzzy Thresholding," in IEEE Journal of Translational Engineering in Health and Medicine, vol. 6, pp. 1-23, 2018, Art no. 3800123.