COMP9517 Group Project Report

Yuanyao Qiu, Jia Li, Liwen Hu, Xinyue Chen, Ziyang Geng

I. Introduction

The segmentation of medical images by computer vision technology can assist medical personnel to diagnose diseases and reduce the burden of medical experts. The two parts of this project are respectively the segmentation of microaneurysms, hemorrhages, hard exudates and soft exudates in diabetic retinopathy (DR) and the segmentation of blood vessels in retinal images. Compared with traditional image-based processing methods, deep learning usually performs better. U-net is a commonly used technology in medical image processing. In the two parts, we used the U-net structure to process the image.

The first part of the project used the publicly available Indian diabetic retinopathy image data set (IDRiD), a clinically collected data set containing microaneurysms, hemorrhages, hard exudates, and soft exudates in retinopathy. The first 54 images are using as training set, 27 images as test set. In the given training set, there are 54 images with MA lesion, 54 images with EX lesion, 53 images with HE lesion and only 26 images with SE. The second part used the public data set DRIVE, which compared the segmentation of blood vessels in the retina, 20 images of which were used as training set and 20 images as test set. The image in the dataset contains a lot of noise that affects segmentation, such as bright spots, shadows, and other lesions. The image's angle, brightness and contrast are also different, which will affect the segmentation result.

II. Literature Review

In the literature review, we totally find 7 methods.

A. Active contours

Active contour model, also called snakes, is used for delineating an object outline from a possibly noisy 2D image.

A snake is an energy minimizing, deformable spline influenced by constraint and image forces that pull it towards object contours and internal forces that resist deformation. This technique performances well in noise images, but requires main object be salient in images.

B. Level sets

Level-set methods (LSM) can perform numerical computations involving curves and surfaces on a fixed Cartesian grid without having to parameterize these objects. Also, the level-set method makes it very easy to follow shapes that change topology.

C. Graph-based

Graph-based merging firstly segment image into regions, recursively split the whole image into pieces based on region statistics. Then recursively merge regions together in a hierarchical fashion. Define a predicate for measuring the evidence for a boundary between two regions using a graph-based representation of the image. Finally combine splitting and merging sequentially. This method performs well while images could be separate into some connected regions.

D. Mean Shift

The mean shift algorithm is a nonparametric clustering technique which does not require prior knowledge of the number of clusters, and does not constrain the shape of the clusters. Mean shift clustering aims to discover "blobs" in a smooth density of samples. It is a centroid-based algorithm, which works by updating candidates for centroids to be the mean of the points within a given region. These candidates are then filtered in a post-processing stage to eliminate near-duplicates to form the final set of centroids. This method work better in object-tracing in video as it can converge to real object instantly, but perform bad if objects are rotated or overlap.

E. Normalized cuts

Normalize Cut is a method extends Min-Cut, which divide by some measure of the size of the vertex set (eg vol A = the sum of the degrees of the vertices in all A, meaning the sum of the weights of all points in A to all points in the graph), that is, NormalizeCut(A, B) = Cut(A, B) / volA + Cut(A, B) / volB. It is clear from the formula that NormalizeCut can find the minimum weight of the points between different subsets.

F. Binary MRF

MRF includes Markov and random field. Markov indicates a random process, and the value of the current state is only related to the value of the previous state. It can be understood as the neighboring state: set the random process {Xn, ∀nT}, n∈T and i0, i1, ..., in+1∈I satisfy: P{Xn+1=in+1|X0=i0, ..., Xn=in}=P{Xn+ 1=in+1|Xn=in} Then {Xn, n∈T} is Markov with the process, that is, the conditional probability determines whether it is a Markov random field.

G. U-NET

U-Net is an improved CNN, which has good performance for biomedical image segmentation. And it is also the reason we finally choose this model in both task1 and task2. U-Net stems from FCN. But for biomedical images which in large size, FCN cannot learn details well. U-Net consists of a contracting path and an expansive path, which gives it the u-shaped architecture. The contracting path is a typical convolutional network that consists of repeated application of convolutions, each followed by a rectified linear unit (ReLU) and a max pooling operation. During the contraction, the spatial information is reduced while feature information is increased. The expansive pathway combines the feature and spatial information through a sequence of up-convolutions and concatenations with high-resolution features from the contracting path.

III. Methods

According to performance and evaluation of these methods, U-Net methods perform better than other methods.

The U-Net structure is very suitable for medical image segmentation because it can combine the underlying information which is helpful for high precision and the high-level information which is used to extract complex features. Therefore, we finally applied U-Net model to implement both task1 and task2.

A. Task 1: Retinal Lesions Segmentation

1) Pre-processing

a) Enhancing brightness:

Extract the RGB color channels of the original image, obtain the brightness of the image according to the following formula, and then adjust the brightness of all pixels to the same level.

$$brightness = \sqrt{0.241 * r^2 + 0.691 * g^2 + 0.068 * b^2}$$

b) Cropping images

Make 512×512 sub-images and sub-masks. The size of the initial image is (4288,2848), so first padding the image with 0 to (4608,3072), then split into 54 x (512,512).

c) Data augumentation

Transform images by flip and rotation.

2) Model Architecture

We used the U-Net structure to complete this task. The U-Net structure is very suitable for medical image segmentation because it can combine the underlying information which is helpful for high precision and the high-level information which is used to extract complex features. And the sample size required is smaller than others.

We modified the U-Net structure so that the up-sampling layer has the same number of feature maps and corresponding initial layers. We do this because the initial layers are as important as the up-sampling layers, so the number of them should be correspond. Then, we set U-Net is max-pooled 5 times for all four lesions (microaneurysms, hemorrhages, hard exudates and soft exudates) segment. We used inverse pixel shuffling to convert image from 512 × 512 × 3 to 512 × 512 × 64 as the input for U-Net. After 5 times max-pooled, concatenate and up-sample, when U-Net finish, convert image form 512 × 512 × 64 to 512 × 512 × 4.

The architecture of our U-Net model for task 1 is shown as table 1.

TABLE I. TASK 1 U-NET ARCHITECTURE

Details of model architecture for segmentation of Hard Exudates, Soft Exudates, Microaneurysm and Haemorrhage. Concatenation is described with brackets. up means up-scaling of feature maps by scale of 2.

Block	**U-Net**	
	Operation	***Output size***
Input	fundus	(512,512,3)
conv 1	$\begin{Bmatrix} 3\times3\ \text{conv} \\ \text{batch}-\text{norm} \\ \text{ReLU} \end{Bmatrix}\times 2$	(512,512,64)
pool 1	2×2 max pool	(256,256,64)
conv 2	$\begin{Bmatrix} 3\times3\ \text{conv} \\ \text{batch}-\text{norm} \\ \text{ReLU} \end{Bmatrix}\times 2$	(256,256,128)
pool 2	2×2 max pool	(128,128,128)
conv 3	$\begin{Bmatrix} 3\times3\ \text{conv} \\ \text{batch}-\text{norm} \\ \text{ReLU} \end{Bmatrix}\times 2$	(128,128,256)
pool 3	2×2 max pool	(64,64,256)
conv 4	$\begin{Bmatrix} 3\times3\ \text{conv} \\ \text{batch}-\text{norm} \\ \text{ReLU} \end{Bmatrix}\times 2$	(64,64,512)
pool 4	2×2 max pool	(32,32,512)
conv 5	$\begin{Bmatrix} 3\times3\ \text{conv} \\ \text{batch}-\text{norm} \\ \text{ReLU} \end{Bmatrix}\times 2$	(32,32,1024)
pool5	2×2 max pool	(16,16,1024)
conv 6	$\begin{Bmatrix} 3\times3\ \text{conv} \\ \text{batch}-\text{norm} \\ \text{ReLU} \end{Bmatrix}\times 2$	(16,16,2048)
up 1	Upsample(conv 6)	(32,32,2048)
conv 7	$\begin{Bmatrix} 3\times3\ \text{conv} \\ \text{batch}-\text{norm} \\ \text{ReLU} \end{Bmatrix}\times 2$	(32,32,1024)
concat 1	[conv 7, conv 5]	(32,32,2048)
conv 8	$\begin{Bmatrix} 3\times3\ \text{conv} \\ \text{batch}-\text{norm} \\ \text{ReLU} \end{Bmatrix}\times 2$	(32,32,1024)
up 2	Upsample(conv 8)	(64,64,1024)
conv 9	$\begin{Bmatrix} 3\times3\ \text{conv} \\ \text{batch}-\text{norm} \\ \text{ReLU} \end{Bmatrix}\times 2$	(64,64,512)
concat 2	[conv 9, conv 4]	(64,64,1024)
conv 10	$\begin{Bmatrix} 3\times3\ \text{conv} \\ \text{batch}-\text{norm} \\ \text{ReLU} \end{Bmatrix}\times 2$	(64,64,512)
up 3	Upsample(conv 10)	(128,128,512)
conv 11	$\begin{Bmatrix} 3\times3\ \text{conv} \\ \text{batch}-\text{norm} \\ \text{ReLU} \end{Bmatrix}\times 2$	(128,128,256)
concat 3	[conv 11, conv 3]	(128,128,512)
conv 12	$\begin{Bmatrix} 3\times3\ \text{conv} \\ \text{batch}-\text{norm} \\ \text{ReLU} \end{Bmatrix}\times 2$	(128,128,256)
up 4	Upsample(conv 12)	(256,256,256)
conv 13	$\begin{Bmatrix} 3\times3\ \text{conv} \\ \text{batch}-\text{norm} \\ \text{ReLU} \end{Bmatrix}\times 2$	(256,256,128)
concat 2	[conv 13, conv 2]	(256,256,256)
conv 14	$\begin{Bmatrix} 3\times3\ \text{conv} \\ \text{batch}-\text{norm} \\ \text{ReLU} \end{Bmatrix}\times 2$	(256,256,128)
up 5	Upsample(conv 14)	(512,512,128)
conv 15	$\begin{Bmatrix} 3\times3\ \text{conv} \\ \text{batch}-\text{norm} \\ \text{ReLU} \end{Bmatrix}\times 2$	(512,512,64)
concat 1	[conv 15, conv 1]	(512,512,128)
conv 16	$\begin{Bmatrix} 3\times3\ \text{conv} \\ \text{batch}-\text{norm} \\ \text{ReLU} \end{Bmatrix}\times 2$	(512,512,64)
output	conv 1×1	(512,512,4)

3) Loss function

We used weighted BCE Loss as loss function which is given by

$$L = -\frac{1}{N}\sum_{i=0}^{n}[wy_{true}^{i}\, log(y_{pred}^{i}) + (1-y_{true}^{i}) log\,(1-y_{pred}^{i})$$

where n denotes the number of images in a batch, y_{true}^{i} and y_{pred}^{i} represent true segmentation and predicted segmentation for i^{th} image.

4) Post-processing

Output all the predicted images, and ervery 54 predicted image is stitched together to an original resolution image. Then 4 channels are divided to four different type retinal lesions images.

B. Task 2: Blood Vessels Segmentation In Retinal Images

The key steps are as follows: First, the color fundus image data is subjected to data preprocessing. These include grayscale transformation, normalization, contrast-limited

adaptive histogram equalization (CLAHE) and gamma transformation. Considering that the training set has only 20 fundus images, random slices are used to obtain more data. The first 90% of the data is entered into the designed U-Net network for training, and the last 10% is used for verification. Finally, the segmentation results are compared and output, and the accuracy, recall, and AUC/ROC curves are used for evaluation.

1) Pre-processing

a) Clahe histogram equalization

The interpolation value is used, that is, the value of the mapping function of the four sub-blocks around the pixel is bilinearly interpolated to obtain the value of each pixel.

b) Data Enhancement

A set of 190,000 patches is obtained by randomly extracting 9,500 patches in each of the 20 DRIVE training images, and the picture is divided into small pictures of 48*48. Different tiles may contain the same portion of the original image, but no further data enhancements are performed.

2) Model Architecture

The architecture of our U-Net model for task 2 is shown as table 2.

TABLE II. TASK 2 U-NET ARCHITECTURE

Block	**U-Net**	
	Operation	***Output size***
Input	fundus	(171000,1,48,48)
conv 1	$\begin{Bmatrix}3\times3\ \text{conv}\\ \text{ReLU}\end{Bmatrix}$ Dropout（0.2） $\begin{Bmatrix}3\times3\ \text{conv}\\ \text{ReLU}\end{Bmatrix}$	(171000,32,48,48)
pool 1	$2\times2\ maxpooling2D$	(171000,32,24,24)
conv 2	$\begin{Bmatrix}3\times3\ \text{conv}\\ \text{ReLU}\end{Bmatrix}$ Dropout（0.2） $\begin{Bmatrix}3\times3\ \text{conv}\\ \text{ReLU}\end{Bmatrix}$	(171000,64,24,24)
pool 2	$2\times2 maxpooling2D$	(171000,64,12,12)
conv 3	$\begin{Bmatrix}3\times3\ \text{conv}\\ \text{ReLU}\end{Bmatrix}$ Dropout（0.2） $\begin{Bmatrix}3\times3\ \text{conv}\\ \text{ReLU}\end{Bmatrix}$	(171000,128,12,12)
up 1	Upsampling2D(2,2) Concat(conv2, up1)	(171000,172,24,24)
conv 4	$\begin{Bmatrix}3\times3\ \text{conv}\\ \text{ReLU}\end{Bmatrix}$ Dropout（0.2） $\begin{Bmatrix}3\times3\ \text{conv}\\ \text{ReLU}\end{Bmatrix}$	(171000,64,24,24)
up 2	Upsampling2D(2,2) Concat(conv1, up2)	(171000,96,48,48)
conv 5	$\begin{Bmatrix}3\times3\ \text{conv}\\ \text{ReLU}\end{Bmatrix}$ Dropout（0.2） $\begin{Bmatrix}3\times3\ \text{conv}\\ \text{ReLU}\end{Bmatrix}$	(171000,32,48,48)
conv 6	$\begin{Bmatrix}1\times1\ \text{conv}\\ \text{ReLU}\end{Bmatrix}$ Reshape Permute	(171000,2,48,48) (None, 2304, 2)
conv 7	softmax	(None, 2304, 2)

3) Loss function

We used Cross Entropy Loss as loss function.

4) Dropout

Dropout refers to randomly letting the weights of some hidden layer nodes of the network not work when the model is trained. Those nodes that do not work can be temporarily considered not part of the network structure, but its weights are preserved (only temporarily not updated). A 0.2 dropout is used between two consecutive convolutional layers to prevent over-fitting due to fewer training samples.

5) Post-preprocessing

Output all the predicted images, and each predicted image is stitched together with the original image for visual comparison.

IV. EXPERMENTSKL SETUP

A. Dataset

1) Task 1

In task 1 we use the segmentation of retinal lesions associated with Diabetic Retinopathy in IDRiD (Indian Diabetic Retinopathy Image Dataset) datasets by Prasanna Porwal, Samiksha Pachade and Manesh Kokare. This dataset include 54 original retinal images in JPG as training image and 27 original retinal images in JPG as test images. After pre-processing, we get 54*54(crop)*6(transform) = 17496 512x512 images as train input. And 27*54(crop) = 1458 512x512 images as test input.

2) Task 2

In task 2 we use blood vessel segmentation in DRIVE (Digital Retinal Images for Vessel Extraction) datasets. This dataset include 40 original retinal images in TIF format, divided into 20 training images and 20 test images.

B. Software

1) Task 1

The programs we used in task 1 are Windows 10, Python 3.7, Pytorch 1.3.0, OpenCV 3.4.2 and PIL 5.3.0 in this task.

2) Task 2

The programs we used in task 2 are Windows 10, Python 3.7 and Pytorch 1.3.0 in this task.

C. Hardware

1) Task 1

The CPU of the computer we used in task1 to is Intel Core i7-9700K@3.6GHz, the GPU is Nvidia RTX 2080 8G and the memory is 16GB.

2) Task 2

The CPU of the computer we used in task to is AMD 1700X@3.4GHz, the GPU is Nvidia RTX 2070 and the memory is 16GB.

D. Evaluation Method

In task 1 and task 2 ,we both use Recall rate,F1score and Precision rate as the evaluation methods. In task1,we also use IOU as the evaluation method, and we also use Accuracy, ROC and AUC as the evaluation methods in task 2 . All the evaluation methods are as follow.

1) Accuracy Rate

Predict the correct proportion of the correct (positive class negative).

$$\boldsymbol{Accuracy} = \frac{\mathbf{TP} + \mathbf{TN}}{\mathbf{TP} + \mathbf{TN} + \mathbf{FP} + \mathbf{FN}}$$

2) Precision Rate

Precision is the percentage of True Positives in all case prediction is true.

$$\boldsymbol{Precision} = \frac{\mathbf{TP}}{\mathbf{TP} + \mathbf{FP}}$$

3) Recall Rate

Correctly predicting positive proportions that are all positive.

$$\boldsymbol{Recall} = \frac{\mathbf{TP}}{\mathbf{TP} + \mathbf{FN}}$$

4) F1 score

The F1 score is the average of the precision and recall.

$$F1 = \frac{2\mathrm{TP}}{2TP + FP + FN}$$

5) IoU

Intersection over Union is an evaluation metric used to measure the accuracy of an object detector on a particular dataset.

$$IoU = \frac{area\ of\ overlap}{\text{area of union}}$$

6) ROC and AUC

The receiver operating characteristic curve is a comprehensive indicator reflecting the continuous variables of sensitivity and specificity. Each point on the ROC curve reflects the sensitivity to the same signal stimulus.

Abscissa: 1-Specificity, False positive rate (FPR, FPR = FP/(FP+TN)), the proportion of samples that are positive but negative is the proportion of all negative samples; ordinate: Sensitivity , True positive rate (TPR, TPR = TP / (TP + FN)), the proportion of samples predicted to be positive and actually positive accounts for all positive samples.

AUC (Area Under Curve) is defined as the area under the ROC curve. Obviously, the value of this area will not be greater than 1. Since the ROC curve is generally above the line y=x, the range of AUC is generally between 0.5 and 1. The AUC value is used as the evaluation criterion because many times the ROC curve does not clearly indicate which classifier works better, and as a numerical value, the classifier corresponding to the larger AUC is better.

V. RESULT

A. Task 1

Here is an original image and segment results for image IDRiD_71.

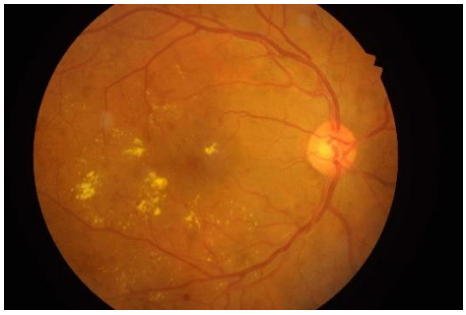

Fig. 1. Original image IDRiD_71.jpg

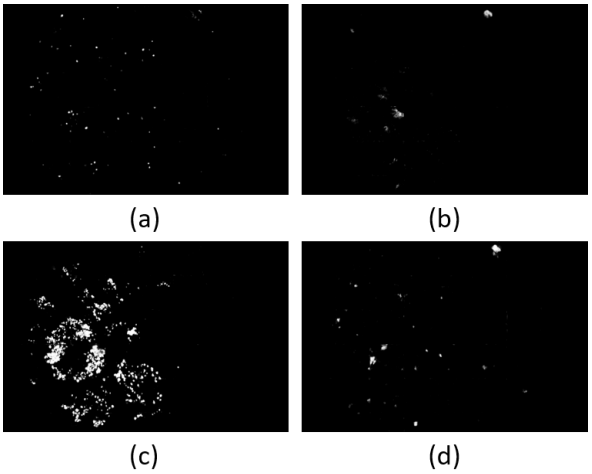

Fig. 2. The segment result of microaneurysm(a), soft exudate(b), hard exudate(c) and hemorrhage(d) for image IDRiD_71

In the given training set, there are 54 images with MA lesion, 54 images with EX lesion, 53 images with HE lesion and only 26 images with SE. Lesions only account for a small proportion of fundus images, and the number of positive cases and negative cases varies greatly between these images, which is a great challenge for processing data sets and evaluating results. Among all the data, the model segmentation EX performs best, and the HE is relatively low. Table 3. shows the results of the evaluation of the test.

TABLE III. TASK 1 RESULT

Lesions	Precision	Recall	F1	IoU
Microaneurysms	0.425	0.447	0.423	0.271
Hard Exudates	0.619	0.696	0.627	0.471
Soft Exudates	0.577	0.353	0.415	0.277
Haemorrhages	0.393	0.304	0.322	0.239

B. *Task 2*

Here is an original image and segment results for image 01_test.

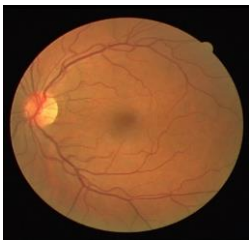

Fig. 4. Original 01_test.tif

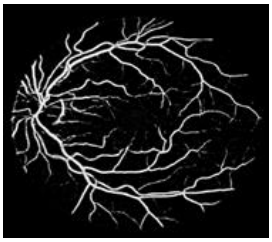

Fig. 5. The segment result

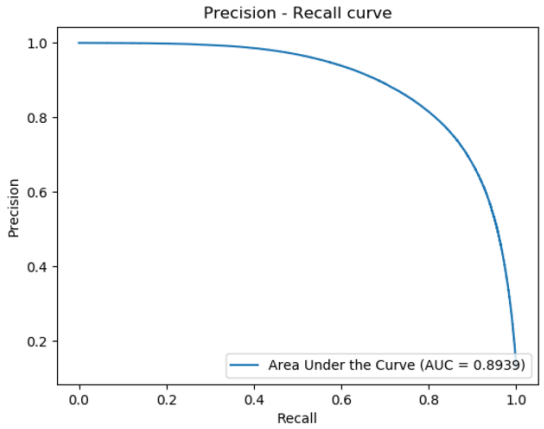

Fig. 6. Recall curve of task2

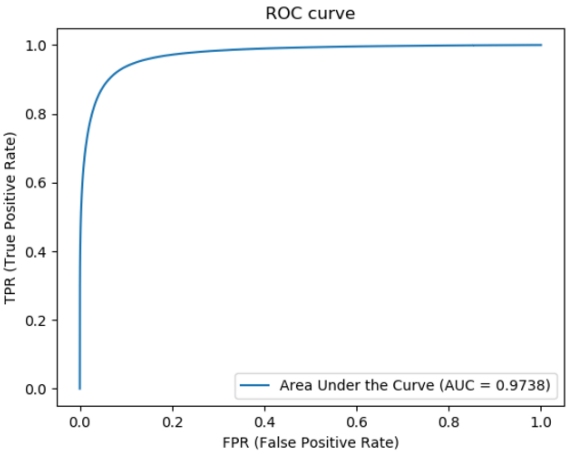

Fig. 7. Roc curve of task2

TABLE IV. RESULT OF TASK 2

Evaluation Methods	Value
Area under the ROC curve	0.974
Area under the Precsion-Recall curve	0.894
F1 scores	0.789
Accuracy	0.951
Recall	0.713
Precision	0.882

VI. DISCUSSITION AND CONCLUTION

In task1, we implement a general approach for automatic retinal lesion segmentation in color fundus image based on deep learning U-Net model . Our network achieves good performance in some cases. But, Our current defects in this task is a little overfitting due to training data is too little, it only contains 54 images with masks. And we think it could be extended by rotate. Also, adjust architecture and hyperparameters could be able to improve performance, like modify network layers or convolution kernels. Besides, we find some strategies such as adding shortcut layers to improve performance of U-net.

In task 2, our model implement a general framework by U-Net to detect and segment some key features of lesions, which able to help diagnosis of related diseases. Although some results are not idea, our solution in task 2 achieved good result in segment blood vessels in the retina. In addition, we can continue to optimize training procedure to improve performance of our solutions.

Through this project, we learned how to use CV technology address real-world problem and design neural network architecture in deep learning tasks. Thanks to good lessons from professor and kindly guidance from tutors.

VII. CONTRUBUTION OF GROUP MEMBER

n this group project, we divide our group into to groups: writing group(Yuanyao Qiu, Jia Li, Xinyue Chen) and code group(Liwen Hu, Ziyang Geng).But when we do this project , we all work together to complete each part. The writing group also help the code group to coding and the code group also help the write group write the report.

The main contribution of our team members are follow:

Yuanyao Qiu: Writing group leader. Mainly responsible for managing the report, writing the report for task 2 and helping the code team complete the code and the slides of the corresponding part.

Jia Li: Mainly responsible for writing the report for task 1 and helping the code team complete the code and slides of the corresponding part.

Liwen Hu: Code group leader. Mainly responsible for managing the code tasks, writing the code for task 1 and helping the writing team complete the report and slides of the corresponding part.

Xinyue Chen: Mainly responsible for writing the introduction, literature review, discussion and conclusion , making slides and help the code team to complete the code for task1 and task2.

Ziyang Geng: Mainly responsible for writing the code for task 2 and helping the writing team complete the report and slides of the corresponding part.

REFERENCES

[1] Liang-Chieh Chen, George Papandreou, Iasonas Kokki- nos, Kevin Murphy, and Alan L Yuille, "Deeplab: Se- mantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs," *arXiv preprint arXiv:1606.00915*, 2016.

[2] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2015, pp. 234–241.

[3] Wenzhe Shi, Jose Caballero, Ferenc Huszár, Johannes Totz, Andrew P Aitken, Rob Bishop, Daniel Rueck- ert, and Zehan Wang, “Real-time single image and video super-resolution using an efficient sub-pixel con- volutional neural network,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 1874–1883.

[4] Diederik Kingma and Jimmy Ba, “Adam: A method for stochastic optimization,” in *Proceedings of the Interna- tional Conference on Learning Representations (ICLR)*, 2015.

[5] Salamat, Nadeem, Malik M. Saad Missen, and Aqsa Rashid. "Diabetic retinopathy techniques in retinal images: a review." Artificial intelligence in medicine (2018).

[6] Sinthanayothin, Chanjira, James F. Boyce, Tom H. Williamson, Helen L. Cook, Evelyn Mensah,Shantanu Lal, and David Usher. "Automated detection of diabetic retinopathy on digital fundus images." Diabetic medicine 19, no. 2 (2002): 105-112.

[7] Quellec, Gwenolé, Katia Charrière, Yassine Boudi, Béatrice Cochener, and Mathieu Lamard. "Deep image mining for diabetic retinopathy screening." Medical image analysis 39 (2017): 178-193.

[8] Porwal, Prasanna, Samiksha Pachade, Manesh Kokare, Girish Deshmukh, Jaemin Son, Woong Bae, Lihong Liu et al. "IDRiD: Diabetic Retinopathy–Segmentation and Grading Challenge." Medical Image Analysis (2019): 101561.

[9] Budai, Attila, Georg Michelson, and Joachim Hornegger. "Multiscale Blood Vessel Segmentation in Retinal Fundus Images." In Bildverarbeitung für die Medizin, pp. 261-265. 2010.

[10] Bilal, Sara, Fatin Munir, and Mostafa Karbasi. "BLOOD VESSELS SEGMENTATION BASED ON THREE RETINAL IMAGES DATASETS." (2006).

[11] Salamat, Nadeem, Malik M. Saad Missen, and Aqsa Rashid. "Diabetic retinopathy techniques in retinal images: a review." Artificial intelligence in medicine (2018).

[12] Fraz, Muhammad Moazam, Paolo Remagnino, Andreas Hoppe, Bunyarit Uyyanonvara, Alicja R.Rudnicka, Christopher G. Owen, and Sarah A. Barman. "An ensemble classification-based approach applied to retinal blood vessel segmentation." IEEE Transactions on Biomedical Engineering 59, no. 9 (2012): 2538-2548.