AUTOMATIC RETINAL VESSEL SEGMENTATION/DETECTION FROM IMAGES

Kong Lingdong

School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore

Email: KONG0129@e.ntu.edu.sg, **Matri. Number:** G1902089A

ABSTRACT:

This project inherits the knowledge learned from the graduate-level course EE7403 Image Analysis & Pattern Recognition and desires to conduct segmentation and detection tasks automatically for retinal vessel images. The report consists of the following two parts (subprojects): image processing (Part I) and machine learning (Part II). In Part I, four image processing techniques are utilized to segment/detect retinal vessels from images, i.e., the baseline method given by Prof. Jiang [1,2], the modified baseline method, the Otsu's method [3], and the ensemble method. In Part II, a supervised learning-based model named VesselNet is used to segment retinal blood vessels, which combines the structure of U-Net [4] and the idea of the dense block in DenseNet [5] together. Experimental results illustrate the excellent performance of deep convolutional neural networks for segmenting retinal vessels from images in comparison with the traditional images processing methods. However, due to learning characteristics, this supervised learning-based method requires ground-truth labels annotated manually (e.g., the DRIVE dataset [7] is used to train the VesselNet model in this project), which is not easy to obtain in real-life applications. What is more, a lot of computational resources are needed to train a deep neural network, while image processing methods can achieve a much faster processing speed.

KEYWORDS: Segmentation; Thresholding; Deep Convolutional Neural Networks; Dense Layer.

I. RETINAL VESSEL SEGMENTATION

Retinal vessel segmentation/detection has a wide range of application scenarios, which vary from diagnosis, cardiovascular evaluation, and screening to treatment. Manual segmentation of retinal blood vessels is a tedious task that also requires high-professional skills. Therefore, a large number of techniques and algorithms have been proposed relating to segment/detect retinal blood vessels automatically. In this project, we focus on the extraction of blood vessels from two-dimensional, colored, and non-dilated retinal images. The segmentation is conducted based on two approaches: image processing (which is directly applied to the

objective image), and machine learning (which is trained first using training examples to generate high-dimensional features and then test the objective image).

A. Segmentation by Image Processing

In Part I, the retinal vessels are to be segmented using image processing. There are a lot of ideas to handle this problem by image-driven techniques, such as the edge-based approaches, the region-based approaches, automatic thresholding, etc. Automatically locating the accurate vascular pattern is very important in implementations.

The original retinal images are in RGB color whose blue channel is almost empty, whereas the red channel usually is saturated, and the green channel gives good representations of retinal features. What is more, in the inverted green channel, the blood vessels appear lighter than the background, and that is why researchers have used this channel for vessel enhancement. After this, we can apply the thresholding techniques to create binary masks for retinal vessel segmentation. The retinal vessels are marked by the masking procedure, which assigns one to all of the pixels belong to retinal vessels and zero to non-vessel pixels. In this project, four thresholding methods are used, i.e., the baseline method given by Prof. Jiang [1,2], the modified baseline method, the Otsu's method [3], and the ensemble method. We will illustrate the details as well as the segmentation results of these methods in Section II.

B. Segmentation by Machine Learning

In supervised learning-based methods, the rule of vessel extraction is learned by the algorithm on the basis of a training set of manually processed and segmented reference images, often termed as the gold standard. An ophthalmologist marks the vascular structure in these ground-truth images. By using machine learning techniques, the classification criteria are determined by the ground-truth data based on given features. Therefore, the prerequisite is the availability of the ground-truth data (labels), which may not easy to obtain in real-life applications. As supervised learning-based methods are designed based on pre-classified samples, their performances are usually better than that of the unsupervised ones.

Neural networks have been extensively investigated for segmenting retinal features like vasculature, thus making classifications on the basis of statistical probabilities rather than objective reasoning [6]. These neural networks employ mathematical weights to decide the probability of input data belonging to a particular output. The weighting system can then be adjusted by training the network with data from known output typically with a feedback mechanism to allow retraining. In this project, we will use VesselNet to train the samples in the DRIVE dataset and use the trained features to test the objective image. VesselNet inherits most of the structure of U-Net [4], which is a powerful convolutional neural network proposed
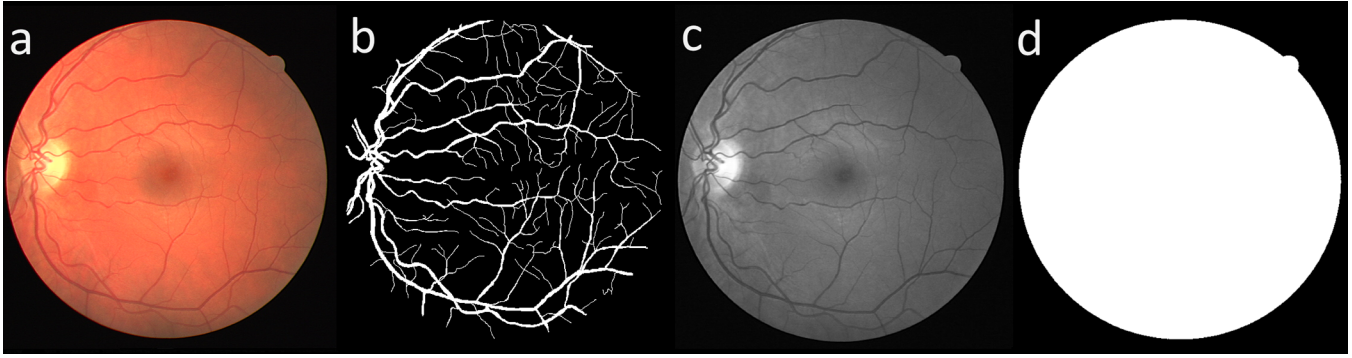
Figure. 1 Data sample of the DRIVE dataset

(a) Color fundus image. (b) Manual segmentation. (c) Green channel. (d) FOV mask.

by Ronneberger *et al.* in 2015. Due to the unique structure characteristics and the use of excessive data augmentation, U-Net achieves excellent performance in handling the medical image dataset with very little training data. We will discuss more details of the network architecture and training of U-Net in Section III.

C. Retinal Image Dataset

The training samples used in this project comes from the DRIVE dataset [7]. The DRIVE (stands for Digital Retinal Images for Vessel Extraction) is a publicly available database, consisting of a total of forty color fundus photographs (see Figure. 1). These photographs were obtained from a diabetic retinopathy screening program in Netherlands. The screening population consists of 453 subjects between 31 and 86 years of age. Each image has been JPEG compressed, which is a standard practice in screening programs. Of the forty images in the database, seven of them contain pathology, hemorrhages, and pigment epithelium changes. Each image is captured using 8 bits per color plane at 768 × 584 pixels. The field-of-view (FOV) of each image is circular with a diameter of 540 pixels.

The set of forty images in the DRIVE dataset was divided into a training set and a test set in which both of them are containing twenty images. Three observers manually segmented these images. The first observer segmented fourteen images of the training set while the second observer segmented the other six images. The test set was segmented twice, resulting in sets X and Y. The performance of the vessel segmentation algorithms can be measured on the test set using the features trained by the training set. In set X, the observer marked 577649 pixels as vessels and 3960494 pixels as background (with a proportion of 12.7% for vessels). In set Y, 556532 pixels are marked as vessels and 3981611 pixels as background (with a proportion of 12.3% for vessels).
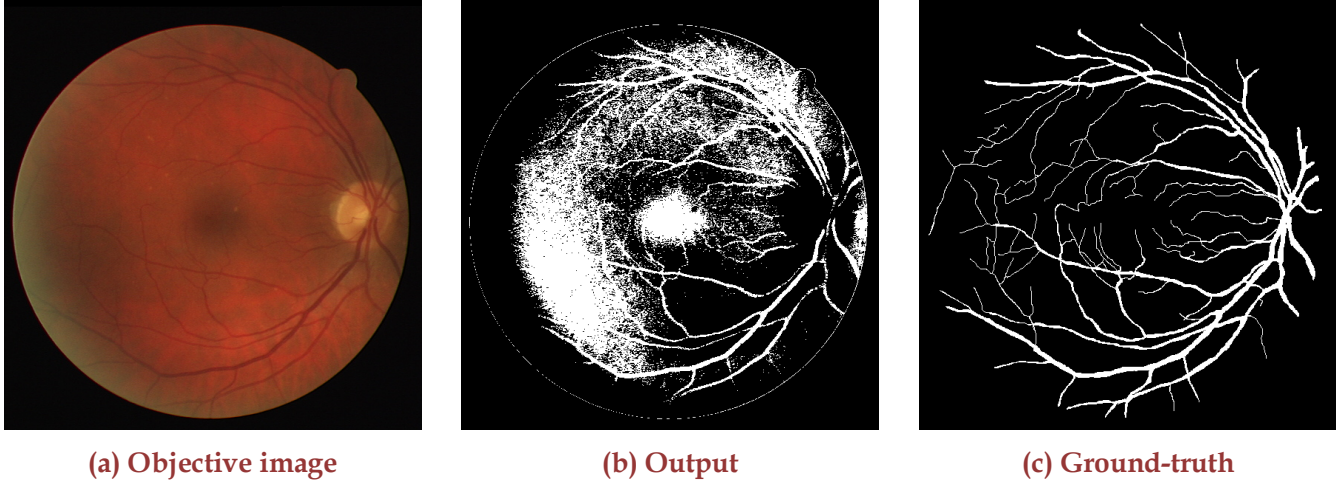
| (a) Objective image | (b) Output | (c) Ground-truth |

Figure. 2 Segmentation by the baseline method

II.   PART I: IMAGE PROCESSING

In this section, we introduce four image-driven methods with thresholding operations for retinal vessel segmentation, i.e., the baseline method given by Prof. Jiang [1,2], the modified baseline method, the Otsu's method [3], and the ensemble method.

A.   The Baseline Method

The baseline method offers a series of thresholding operations to the green channel of the objective retinal vessel image. According to the project description, retinal vessel images `29_test.tif` is going to be segmented. The reason why we use the green channel is that the original retinal images are in RGB color whose blue channel is almost empty, whereas the red channel is saturated as shown in Figure. 2 (a). Therefore, the green channel gives good representations of retinal features.

Assume that the input objective retinal vessel image is denoted as $r(x, y)$. What the thresholding operation does is that by setting threshold T as follows:

$$r(x,y) = \begin{cases} 0, & \text{if } f(x,y) > T \\ 1, & \text{if } f(x,y) \leq T \end{cases}$$

where $f(x, y)$ denotes the gray-level value of the retinal vessel image. threshold T given by the following:

$$T = T\{x, y, f(x,y), p(x,y)\}$$

where $p(x, y)$ denotes the local neighbor's property of the retinal vessel image. Three thresholding strategies exist, i.e.,

4

- Global thresholding: threshold T depends on gray-level value $f(x, y)$ of the whole image alone.

- Local thresholding: threshold T depends on both gray-level value $f(x, y)$ and the local neighbor's property $p(x, y)$.

- Adaptive thresholding: threshold T depends on coordinates (x, y) of the retinal vessel image.

Without loss of generality, the baseline method uses the first thresholding strategy. The heuristic approach to get threshold T can be summarized as follows:

1. Choose T_0 as the initial estimate for threshold T;

2. Segment the retinal vessel image using T_0. Two groups of pixels will be produced, i.e., group G_1 (consisting of pixels with gray-level value $f(x, y) > T_0$), and group G_2 (consisting of pixels with gray-level value $f(x, y) \leq T_0$).

3. Compute the average gray-level values μ_1 and μ_2 for pixels in regions G_1 and G_2, respectively.

4. Compute a new threshold $T = 0.5(\mu_1 + \mu_2)$;

5. Repeat steps 2 to 4 until no noticeable change occurs or the difference in T in successive iterations is smaller than a predefined threshold value.

For the baseline method, three thresholding operations are used to divide the FOV (which is circular with a diameter of 540 pixels) and the background. This is reasonable because of the obvious difference between these two regions. After extracting the expecting region, three more thresholding operations are conducted, and this results in the final output, as shown in Figure. 2 (b). For the convenience of comparison, the ground-truth is illustrated in Figure. 2 (c). As we can see from these figures, the baseline method successfully segments the retinal vessels in the top and the bottom regions. However, some tiny retinal vessels in the left and the middle of the expecting region are not segmented. A lot of point-like noise particles are scattered in these areas. In general, the baseline method gives a simple and standard way for retinal vessel segmentation.

B. The Modified Baseline Method

Inspired by the baseline method illustrated above and on the basis of its thresholding operations, some modifications are applied. Specifically, instead of using global directly, we use the adaptive strategy to find a more suitable threshold T. Meanwhile, more thresholding operations are conducted in order to find a more sophisticated value. As an example, instead of using just three thresholding operations for segmentation, ten operations are used to the modified baseline method. The segmentation result is shown in Figure. 3 (b). In comparison
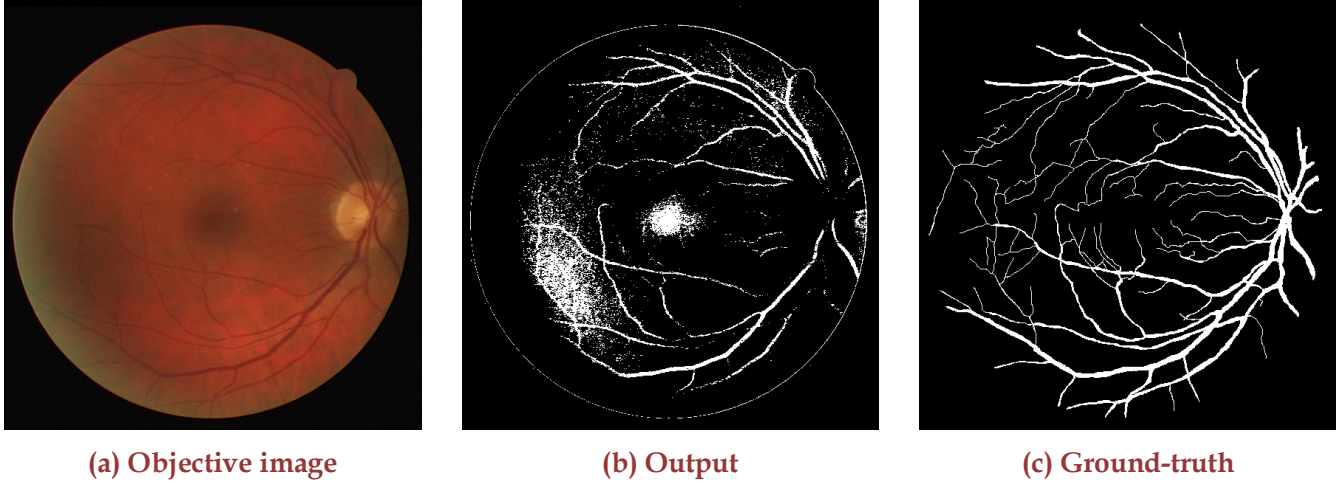
| (a) Objective image | (b) Output | (c) Ground-truth |

Figure. 3 Segmentation by the modified baseline method

with the segmentation result of the baseline method shown in Figure. 2 (b), the modified version suppresses more point-like noise particles in the retinal region and has more clear views of the vessels. However, there is still a distance between the proceed result and the ground-truth. More advanced techniques are required to get better segmentation results.

C. The Otsu's Method

Otsu's method, a classic automatic thresholding technique, is the basis for several automatic detection proposals. It determines optimal threshold values that maximize the between-class variance of the foreground and background. Studies demonstrate that the Otsu's method is effective for thresholding a histogram with bimodal or multimodal distribution. Otsu's method fails if the histogram is unimodal or almost unimodal [5]. Therefor, it provides acceptable results for thresholding general real-life images (e.g., the retinal vessel image). Real-life image histograms typically have multimodal distributions because the objects are usually large compared with the image size. This property enables Otsu's method to determine the optimal thresholds from visual inspection systems generally have unimodal distributions because of the uniformity in luminance and size of the retinal vessel regions, which prevents Otsu's method from identifying the desired threshold values.

The basic procedures of the Otsu's method can be summarized as follows. Let $I = f(x, y)$ be a digital image, where f is an intensity function. The value of $f(x, y)$ is the gray value of the pixel at location (x, y), which lies in the interval $[0, L - 1]$, where L is the number of gray levels. Let the number of pixels with gray value i be n_i and N be the total number of pixels in I. The histogram is considered a probability distribution:

$$p_i = \frac{n_i}{N}$$

Then, the average gray value of I is:

$$\mu_I = \sum_{i=0}^{L-1} i \cdot p_i$$

As a result of thresholding, the pixels of I are divided into two classes, $C_1 = \{(x,y) \mid 0 \leq f(x,y) \leq k\}$ and $C_2 = \{(x,y) \mid k+1 \leq f(x,y) \leq L-1\}$, where k is the selected threshold value. Normally, C_1 is the foreground and C_2 is the background, or vice versa. The probability of class occurrence and the average gray value of each class are, respectively, computed as:

$$\omega_1(k) = \sum_{i=0}^{k} p_i, \qquad \omega_2(k) = \sum_{i=k+1}^{L-1} p_i$$

and

$$\mu_1(k) = \sum_{i=0}^{k} \frac{i \cdot p_i}{\omega_1(k)}, \qquad \mu_2(k) = \sum_{i=k+1}^{L-1} \frac{i \cdot p_i}{\omega_2(k)}$$

In Otsu's method, the performance of the result threshold is measured by considering the distinction between the foreground and background. Using this criterion, the optimal threshold k^* must maximize between-class variance:

$$k^* = \underset{0 \leq k \leq L-1}{\arg\max} \{\omega_1(k)(\mu_1(k) - \mu_I)^2 + \omega_2(k)(\mu_2(k) - \mu_I)^2\}$$

There are several improvements to the Otsu's method. The basic idea is the addition of a weight W to the objective function for adjusting the output threshold. For the intuition purpose only, we only discuss the original approach here. The earliest weighting scheme for the Otsu's method is the valley-emphasis method [8], which can be summarized as the following equation:

$$k^* = \underset{0 \leq k \leq L-1}{\arg\max} \{(1 - p_k)(\omega_1(k)\mu_1^2(k) + \omega_2(k)\mu_2^2(k))\}$$

where the weight W is $(1 - p_k)$. The key-point of this valley-emphasis method is the complement of probability of occurrence $(1 - p_k)$. The smaller the value of p_k (i.e., the lower the probability of occurrence of gray-level value), the larger the weight W becomes. This weight ensures that the output threshold will always be in the valley or bottom rim of the histogram. Otsu's method produces acceptable results when objects in the images are large and sufficiently different from the background. More specifically, it works for images whose histograms have bimodal or multimodal distributions.
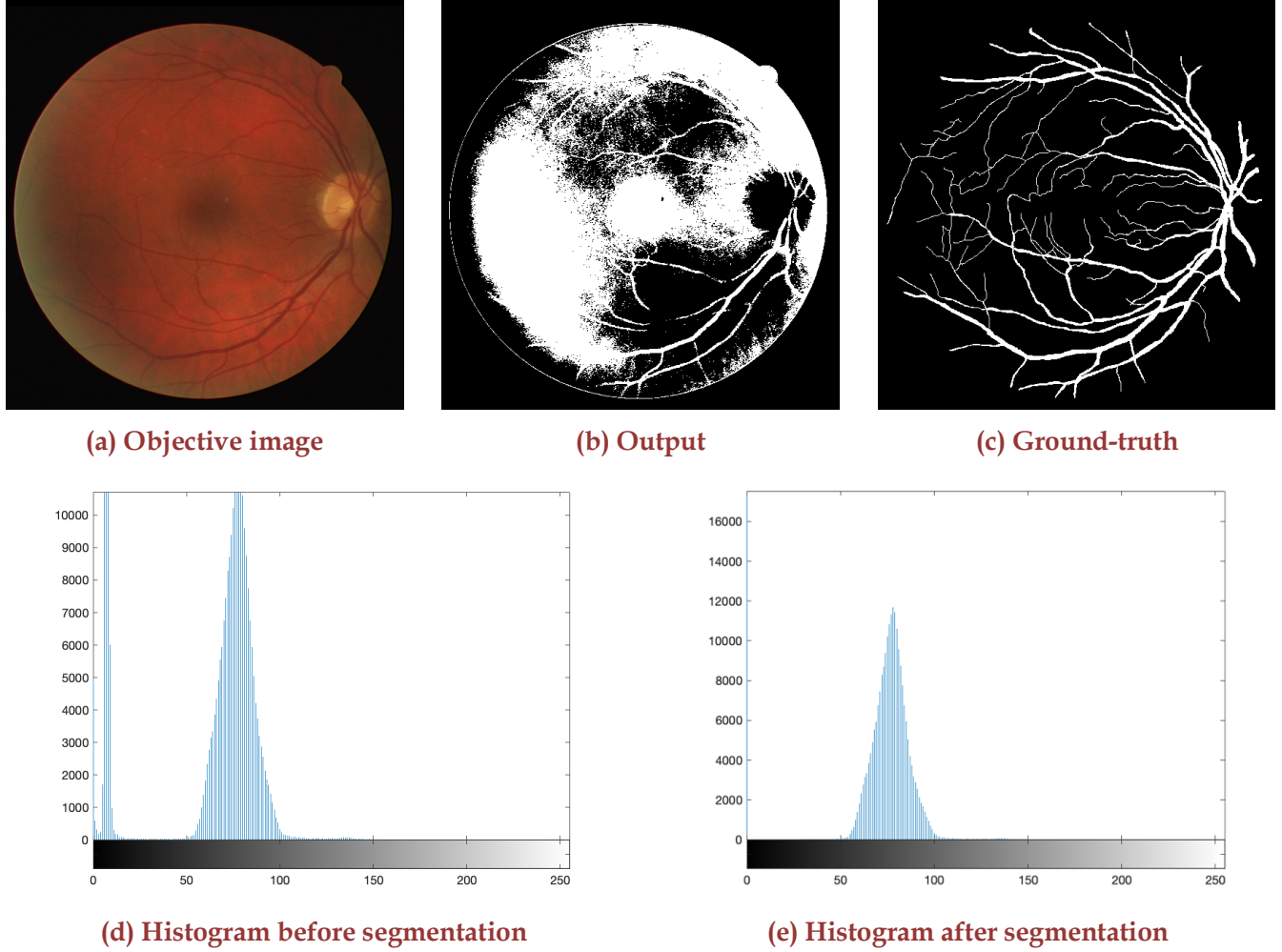
(a) Objective image　　　　**(b) Output**　　　　**(c) Ground-truth**

(d) Histogram before segmentation　　　　**(e) Histogram after segmentation**

Figure. 4 Segmentation by the Otsu's method

For the retinal vessel segmentation task in this project, both of the fundus and the vessels appear in orange and scarlet colors, and thus it is hard for the Otsu's method to segment them. The experimental result is shown in Figure. 4 (b). Obviously, comparing with the ground-truth, only the vessels at the bottom of the expecting region are segmented, and the vessels in the remaining areas are disturbed by a lot of point-like noise particles. The histograms of the retinal vessel images before and after segmentation are shown in Figure. 4 (d) and Figure. 4 (e), respectively.

D. The Ensemble Method

Inspired by the concept of ensemble learning, we are now going to combine the thresholding strategy of the Ostu's method with the strategy of the baseline method. Specifically, we first
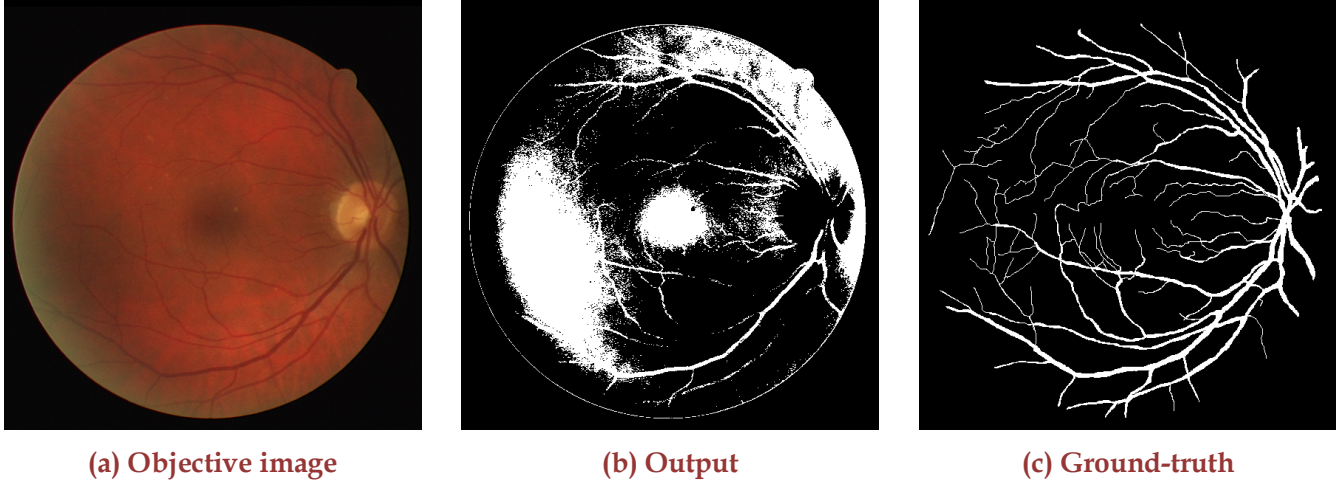
(a) Objective image   (b) Output   (c) Ground-truth

Figure. 5 Segmentation by the ensemble method

apply the global thresholding strategy to find threshold T_1. After this, we use the Ostu's method to find threshold T_2. Thresholds T_1 and T_2 are combined by weighted sum as follows:

$$T_{new} = \alpha \cdot T_1 + \beta \cdot T_2$$

where weight factors α and β are simply chosen as 0.5 here. The experimental result of the ensemble method is shown in Figure. 5 (b). Comparing with the baseline method (Figure. 2 (b)) and the Otsu's method (Figure. 4 (b)), the ensemble method achieves better suppression capability of the point-like noise particles scattered than the Otsu' method, while it keeps more details of the vessels on the bottom area than the baseline method. Better result can be achieved by adjusting the weight factors.

In general, these thresholding methods show some capability of segmenting retinal vessel images. More sophisticated operations are needed to achieve better segmentation results.

III. PART II: MACHINE LEARNING

In the last several years, deep convolutional neural networks have outperformed the state-of-the-art in many visual recognition tasks. While convolutional networks have already existed for a long time, their success was limited due to the size of the available training sets and the size of the considered networks. The breakthrough by Krizhevsky *et al.* was due to supervised training of a large network with eight layers and millions of parameters on the ImageNet dataset with one million training images. Since then, even larger and deeper networks have been trained [8].

With the DRIVE dataset [7], we can now build deep neural networks with convolutional layers, pooling layers, and fully-connected layers and other functional layers to automatically
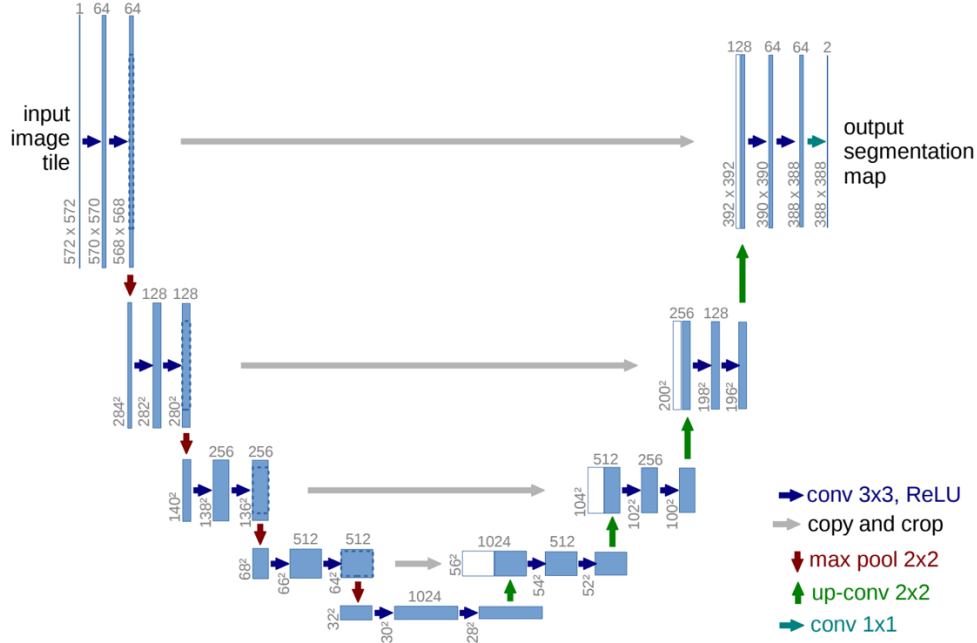
9

Figure. 6 The architecture of U-Net [4] for 32 × 32 pixels in the lowest resolution. Each blue box corresponds to a multi-channel feature map. The number of channels is denoted on top of the box. White boxes denote copied feature maps. The arrows denote the different operations.

segment/detect retinal vessels from images. In this project, we use VesselNet, which inherits the basic architecture of U-Net [4] and combines the idea of the dense blocks in DenseNet [5]. Let us first review the basic structures and characteristics of them.

A. U-Net

The typical use of convolutional networks is on the classification tasks, where the output to an image is just a single class label. However, in many biomedical image processing tasks, for example, the retinal vessel segmentation, the desired output will need localization. That is to say, the class labels are supposed to be assigned to each pixel. What is more, thousands of training images are usually hard to obtain in biomedical tasks due to the difficulty of annotations that requires expert knowledge. U-Net modifies and extends the architecture of the fully convolutional network such that it can work with very few training images and yields precise segmentation results [4].

U-Net consists of a contracting path and an expansive path as shown in the left side and the right side of Figure. 6, respectively, as follows:

- For contracting path: The contracting path follows the typical arrangement of a convolutional neural network. Repeated applications of two 3 by 3 unpadded

convolutions are used and each of them is followed by a ReLU unit and a 2 by 2 max pooling operation with stride of two. At each down-sampling step, the number of feature channels is doubled.

● For expansive path: Every step in the expansive path consists of up-sampling of the feature map followed by a 2 by 2 convolution that halves the number of feature channels, a concatenation with the correspondingly cropped feature map from the contracting path, and two 3 by 3 unpadded convolutions of each followed by a ReLU unit. The cropping is necessary due to the loss of border pixels in every convolution.

At the final layer, a 1 by 1 convolution is used to map each feature vector to the desired number of classes. In total, the network contains 23 convolutional layers. The input retinal vessel images and their corresponding segmentation maps are used to train the network with stochastic gradient descent. Due to the unpadded convolutions, the sizes of the output (segmented) images are smaller than the input images by a constant border width. The energy function is computed by a pixel-wise softmax over the final feature map combined with the cross-entropy loss function. The softmax is defined as follows:

$$p_k(x) = \frac{e^{a_k(x)}}{\sum_{k'=1}^{K} e^{a_{k'}(x)}}$$

where $a_k(x)$ denotes the activation in feature channel k at the pixel position x. K is the number of classes and $p_k(x)$ is the approximated maximum function. The cross-entropy then penalizes at each position the deviation of $p_{\ell(x)}(x)$ from 1 using:

$$E = \sum_x w(x)\log\left(p_{\ell(x)}(x)\right)$$

where $w(x)$ denotes a weight map that gives some pixels more importance during training and ℓ denotes the true label of each pixel. The separation border is computed using morphological operations. The weight map $w(x)$ is then computed as follows:

$$w(x) = w_c(x) + w_0 \cdot e^{-\frac{(d_1(x)+d_2(x))^2}{2\sigma^2}}$$

where $w_c(x)$ denotes the weight map to balance the class frequencies, $d_1(x)$ denotes the distance to the border of the nearest cell, and $d_2(x)$ denotes the distance to the border of the second nearest cell. w_0 and σ are tunable parameters.

Data augmentation is essential for teaching deep neural networks for retinal vessel image segmentation. Especially random elastic deformations of the training samples seem to be the key concept to train a segmentation network with very few annotated images. In U-Net, the smooth deformations are generated using random displacement vectors on a coarse 3 by 3 grid. The displacements are sampled from a Gaussian distribution with 10 pixels standard
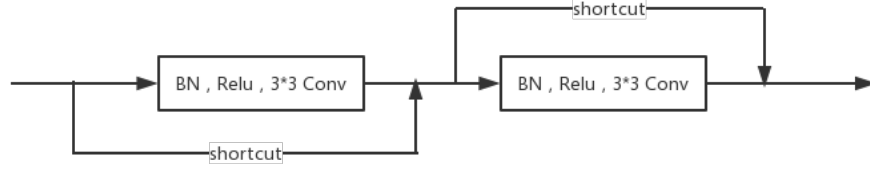
Figure. 7 Convolutional layers in VesselNet

deviation. Per-pixel displacements are then computed using bicubic interpolation. Dropout layers at the end of the contracting path perform further implicit data augmentation.

B. DenseNet

The densely connected convolutional networks (DenseNet) [5] proposed by Huang *et al.* gives a better way of training deep neural networks with the dense block. For each layer, the feature maps of all preceding layers are used as inputs into all subsequent layers. Consider a single retinal vessel image $r(x, y)$ that is passed through a L-layer convolutional network. Each layer implements a non-linear transformation $H_l(\cdot)$, where l is the index of layer. $H_l(\cdot)$ can be a composite function of operations such as the batch normalization (BN), the ReLU unit, the pooling operation, etc. The output of the l-th layer is denoted as r_l.

The traditional convolutional feed-forward networks connect r_l as the input to the $(l + 1)$-th layer, which gives rise to the following layer transition:

$$r_l = H_l(r_{l-1})$$

To further improve the information flow between layers, DenseNet introduces the dense connectivity as follows:

$$r_l = H_l([r_0, r_1, \dots, r_{l-1}])$$

where $[r_0, r_1, \dots, r_{l-1}]$ denotes the concatenation of the feature maps produced in layers 0, ..., $l - 1$. Experimental results verify that DenseNet tends to yield consistent improvement in accuracy with growing number of parameters, without any signs of performance degradation or overfitting [6]. Under multiple settings, it achieved state-of-the-art results across several competitive datasets.

C. VesselNet

With the basic building blocks of U-Net [4] and the dense layers [5], we are now going to make some modifications. The VesselNet model has the same structure of U-Net, and hence it inherits its capability of automatically segmenting retinal vessel images and other biomedical

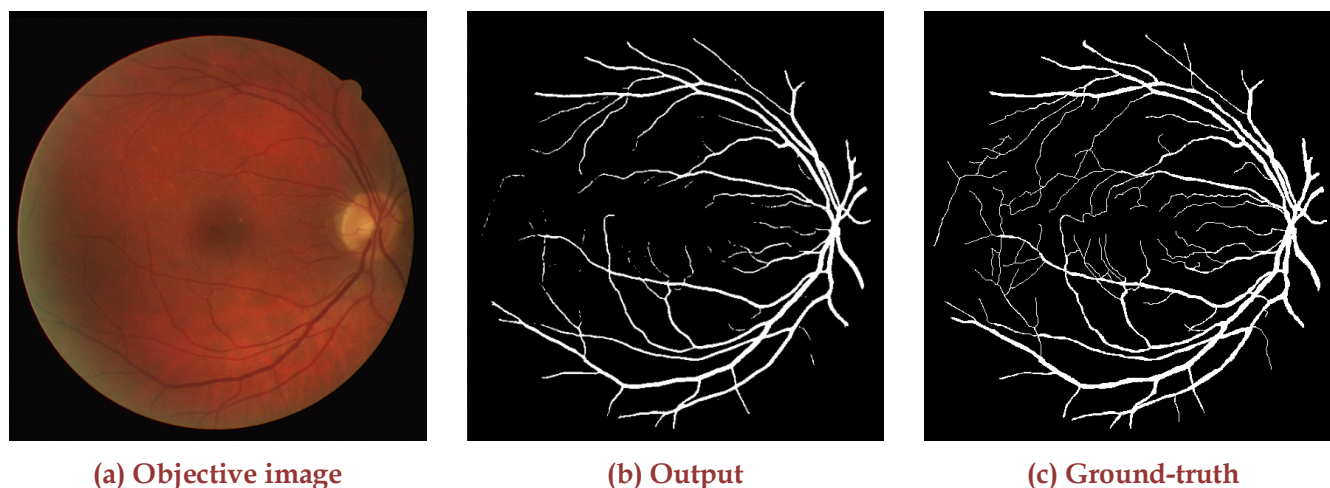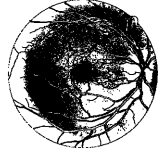**(a) Objective image**   **(b) Output**   **(c) Ground-truth**

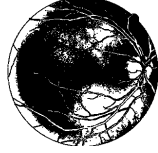**Figure. 8 Segmentation by VesselNet**

images. To get better segmentation results, the convolutional layers in U-Net are replaced by dense layers (see Figure. 7).

The VesselNet in this project is implemented using Keras. The first twenty images in the DRIVE dataset [7] is used for training, the other nineteen images are used for validation, and the remaining one image (i.e., the objective image 29_test.tif) is used for testing.

The segmentation result is shown in Figure. 8 (b). As we can see from this figure and comparing the ground-truth with Figure. 8 (c), VesselNet achieves excellent segmentation performance. Specifically, most of the retinal vessels in the objective image have been segmented/detected, while the output has a clear view without point-like noise particles scattering around retinal areas like the output of the Ostu's method (see Figure. 4 (b)) or the baseline method (see Figure. 2 (b)). However, there are still some fine-grained details of the vessels that cannot be segmented by VesselNet. These details are too sophisticated that even human eyes cannot segment/detect them directly from the objective image (see Figure. (a)).

To obtain better segmentation results, some modifications can be added to VesselNet. First, we need more training data. In the deep learning community, forty training samples are not enough to training a network with complex structure. This can be achieved by adding training samples from other retinal vessel image datasets or using transfer learning. Second, we can fine-tune the hyperparameters of VesselNet to make it have more robust generalization ability. What is more, we can apply some pre-processing techniques to the objective images to make them more suitable for training. In general, with the power of deep learning, the segmentation results have been greatly improved. In the near future, with the emergence of more and more datasets and the development of faster computing power, methods based on supervised learning will become more accurate and efficient.

Table. I Summarization of Segmentations

Method	Accuracy	Comparison
Baseline	57.705%	
Modified baseline	76.928%	
Otsu's [3]	49.846%	
Ensemble	63.711%	
VesselNet	98.675%	

IV. COMPARISON & CONCLUSION

In this project, five methods are used to segment retinal vessels from images, i.e., the baseline method, the modified baseline method, the Otsu's method, the ensemble method, and the VesselNet. The first four methods are based on image processing, while VesselNet embraces the capability of deep learning. The segmentation results of these methods, which compared with the ground-truth, are summarized in Table. I and Figure. 9. Obviously, the supervised learning-based method achieves much better segmentation than the traditional image-driven methods. An accuracy of 98.675% is achieved by VesselNet, while just 49.846% for the Otsu's method [3]. However, we should clearly realize that training a deep neural network requires a lot of computational power, and thus it extends the segmentation time from seconds to minutes and even hours. A tradeoff must be made before we choose which segmentation method to use. This is not a problem just appears in retinal vessel image segmentation, but a common problem in the computer vision community. Maybe we should not compare which
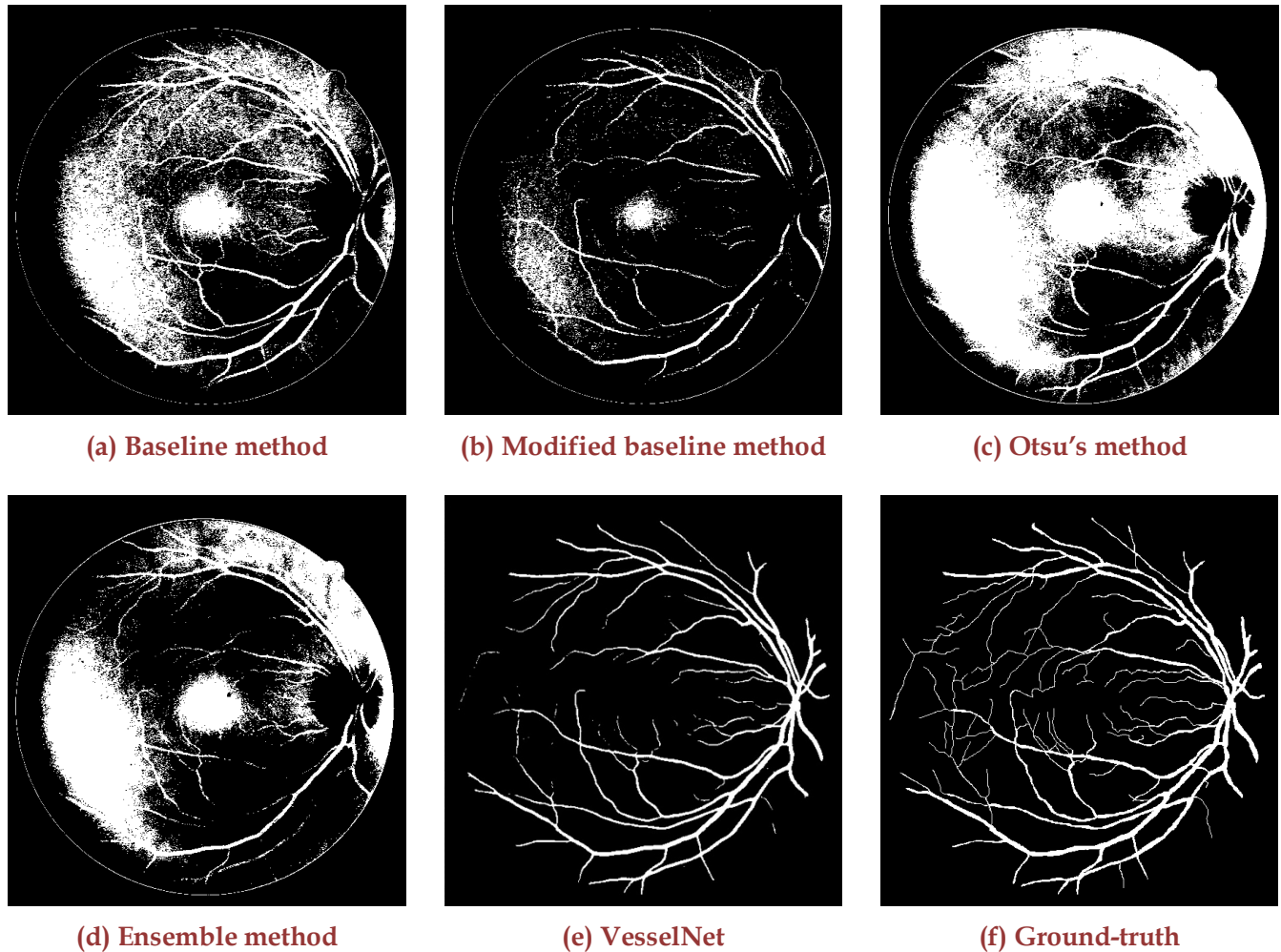
(a) Baseline method **(b) Modified baseline method** **(c) Otsu's method**

(d) Ensemble method **(e) VesselNet** **(f) Ground-truth**

Figure. 9 Comparison of five segmentation methods

method is more correct, but should compare which method is more appropriate for solving the problem.

REFERENCES

[1] X. Wang, X. Jiang, and J. Ren, "Blood Vessel Segmentation from Fundus Image by a Cascade Classification Framework," *Pattern Recognition*, vol. 88, pp. 331-341, 2019.

[2] X. Wang and X. Jiang, "Retinal Vessel Segmentation by a Divide-and-Conquer Funnel-Structured Classification Framework," *Signal Processing*, vol. 165, pp. 104-114, 2019.

[3] N. Otsu, "A Threshold Selection Method from Gray-Level Histograms," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 9, no. 1, pp. 62-66, 1979.

[4] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," *Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 234-241, Spring, Cham, 2015.

[5] G. Huang, Z. Liu, L. V. D. Maaten, K. Q. Weinberger, "Densely Connected Convolutional Networks," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'17)*, pp. 4700-4708, 2017.

[6] K. Akita and H. Kuga, "A Computer Method of Understanding Ocular Fundus Images," Pattern Recognition, vol. 15, pp. 431-443, 1982.

[7] M. Niemeijer, J. J. Staal, B. V. Ginneken, M. Loog, and M. D. Abramoff, "DRIVE: Digital Retinal Images for Vessel Extraction," *http://www.isi.uu.nl/Research/Databases/DRIVE*, 2014.

[8] H. F. Ng, "Automatic Thresholding for Defect Detection Algorithm," *Pattern Recognition Letter*, vol. 270, no, 14, pp. 1644-1649, 2006.

[9] A. Krizhevsky, I. Sustkever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," *Proceedings of the 27th Conference on Neural Information Processing Systems (NIPS'12)*, pp. 1106-1114, 2012.