Regularizer based on Pixel Neighboring Relationship for Deep Convolutional Neural Network in Image Segmentation

Lukman Hakim

Supervisor: Prof. Takio Kurita

Department of Information Engineering

Hiroshima University

This dissertation is submitted for the degree of

Doctor of Engineering in Information Engineering

Graduate School of Engineering January 2022

Acknowledgements

I would like to deeply thank my supervisor, **Prof. Takio Kurita**, for his great support, guidance, invaluable advice, knowledge, discussion, and supervision during my Ph.D. study and research work. I really have learned a lot based on his suggestions and ideas about research, finding problems, proposing ideas, writing papers, interpreting results, and others. Searching for research problems and ideas would be much more difficult without his intellectual guidance. I was inspired by his motivation and passion for working in research with pleasure and happiness.

I would also like to thank all Pattern Recognition and Machine Learning (PRML) lab members during my studies for valuable discussions and enjoyable laboratory life.

I would also like to thank Japan Student Services Organization (JASSO), Hiroshima International Center (HIC), Yahata Memorial Scholarship Foundation for granting me a generous scholarship to complete my studies in Japan. All the teachers and staff of Hiroshima University, particularly the Graduate School of Engineering, have provided dedicated service, hospitality, and assistance during my Ph.D. studies.

Lastly, I would like to thank my family members: my parents, my wife, and my children, who gave me freedom and support during my studies in Japan.

Abstract

Image segmentation is an important task in the field of computer vision. The task of image segmentation is to classify each pixel in the image based on the appropriate label. Recently, many deep learning techniques have been developed for image segmentation that offer high accuracy and deep architecture. However, the deep learning technique uses a pixel-wise loss function for the training process. Using pixel-wise loss neglected the pixel neighbor relationships in the network learning process. The neighboring relationship of the pixels is essential information in the image. Utilizing neighboring pixel information provides an advantage over using only pixel-to-pixel information. This study presents regularizers to provide the pixel neighbor relationship information to the learning process. The regularizers are constructed by the graph theory approach and topology approach: By graph theory approach, graph Laplacian is used to utilize the smoothness of segmented images based on output images and ground-truth images. By topology approach, Euler characteristic is used to identify and minimize the number of isolated objects on segmented images. Experiments show that our scheme successfully captures pixel neighbor relations and improves the performance of the convolutional neural network better than the baseline without a regularization term.

Table of contents

List of figures

List of tables

Chapter 1

Introduction

1.1 Motivation

Image segmentation is an essential part of the image recognition system. This technique divides the image into several segments or objects. Image segmentation can be categorized as a pixel classification problem, where each pixel in the image will be classified with a specific label. There are several categories of image segmentation: Semantic Segmentation and Instance Segmentation. Semantic segmentation classifies each pixel into a specific label and treats multiple objects of the same class as a single entity. On the other hand, instance segmentation treats multiple objects of the same class as distinct individual objects (or instances).

Image segmentation applications have been widely used, such as medical imagery (for example, Retinal Vessel Segmentation, Tumor Segmentation, Breast Cancer Detection), Autonomous Vehicles (Example: Vehicle Detection, Pedestrian Detection), Counting Objects, and Farm Industry. Several Image segmentation techniques have been developed, such as thresholding, k-means clustering, watershed, active contour, region growing, and graph cut methods.

Recently, Deep learning has greatly impacted computer vision with multiple practicals in image segmentation. Deep learning has advantages to learning features from data and is more robust to appearance variations. Long [43] proposed the first deep learning works for semantic image segmentation, using a fully convolutional network (FCN) by modifying the existing CNN architectures such as VGG16 and GoogleNet. Ronnerberger [61] proposed the U-Net for medical image segmentation. The U-Net is inspired by FCN and encoder-decoder models. The author uses skip connections between the encoder and decoder parts. This skip connection uses to transmit the detailed information from the encoder to the decoder part, which is lost due to the downsampling layer. Another work was done is SegNet proposed by Badrinarayanan [4]. SegNet is based on an encoder-decoder network, which is identical to the VGG16 network. SegNet's main innovation is how the decoder does non-linear upsampling on its lower resolution input feature map(s); notably, it leverages pooling indices generated in the corresponding encoder's max-pooling step.

The main issue of image segmentation tasks is providing a technique or algorithm with high accuracy and high efficiency. However, in reality, it is very challenging. Some methods offer the advantage of high accuracy but require high computational costs. The other methods provide high efficiency but with less accuracy.

For training purpose, many deep learning techniques in image segmentation has focused on deeper architectures while still using a pixel-wise loss, such as Binary Cross-Entropy(BCE)[43], [6], Mean Squared Error (MSE) [13], [29], [35] or DICE Loss which is loss calculated based on pixel-to-pixel on predicted output and ground-truth.

However, learning features by using pixel-wise loss fail to preserve the neighboring relationship of the pixel in images[26],[71]. We found that pixel-wise loss difficult to deal with structure boundary issue and tiny region. Ideally,

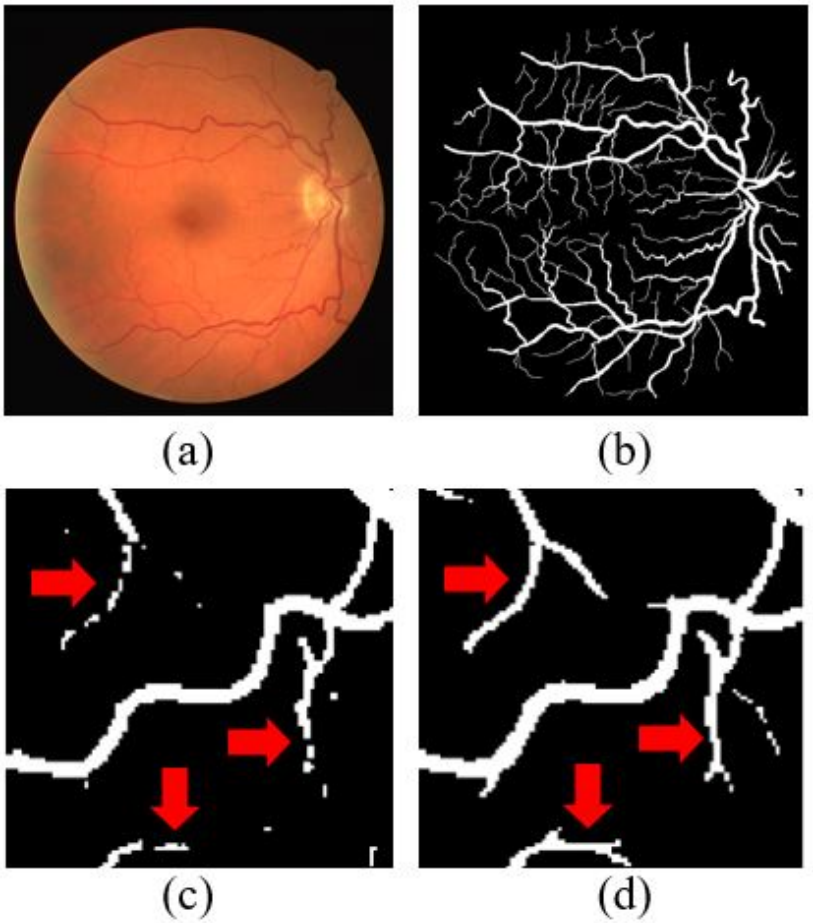
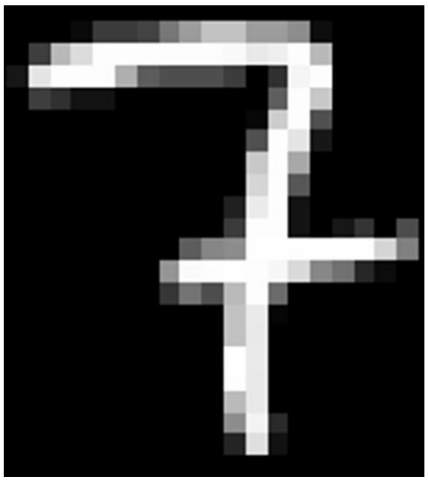


Fig. 1.1 Illustration of the unclear problem and isolated object in vessel segmentation task. (a) input fundus image. (b) groundtruth segmentation of the vessels. (c) result of the baseline without regularizer. (d) proposed isolated object regularizer (red color arrow indicates retaining the width of the small and disjointed vessel connectivities)

the boundary between different classes is clear. However, as shown in Fig. 1.1, the CNN baseline output tends blurred in the boundary area and fails to segment the tiny region. In the segmentation task, this leads to the prediction error at the structure boundary as a false positive, leading to errors in the evaluation process. In super-resolution task, it will make failed to generate detailed part of images.

To address this problem, we propose several works to including pixel relationships information along pixel-wise loss to neural network. The neighboring relationship of a pixel contains rich information about the spatial structure,



	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	10	59	59	67	119	156	194	194	156	156	141	14	0	0	0	0
2	0	62	185	220	253	254	253	253	253	253	249	233	239	253	185	0	0	0	0
3	23	229	253	253	253	175	92	78	78	78	60	0	48	245	253	0	0	0	0
4	0	62	49	19	19	0	0	0	0	0	0	0	94	253	200	0	0	0	0
5	0	0	0	0	0	0	0	0	0	0	0	8	205	253	80	0	0	0	0
6	0	0	0	0	0	0	0	0	0	0	0	80	254	229	23	0	0	0	0
7	0	0	0	0	0	0	0	0	0	0	0	200	253	168	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0	0	0	214	253	87	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0	0	32	235	253	19	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0	0	59	253	253	19	0	23	53	0	76
11	0	0	0	0	0	0	0	0	93	134	141	254	254	254	254	255	254	213	128
12	0	0	0	0	0	0	0	149	250	253	254	253	247	218	135	114	39	12	0
13	0	0	0	0	0	0	0	54	115	78	186	253	107	0	0	0	0	0	0
14	0	0	0	0	0	0	0	0	0	0	194	234	6	0	0	0	0	0	0
15	0	0	0	0	0	0	0	0	0	0	194	233	0	0	0	0	0	0	0
16	0	0	0	0	0	0	0	0	0	0	255	234	0	0	0	0	0	0	0
17	0	0	0	0	0	0	0	0	0	0	254	233	0	0	0	0	0	0	0
18	0	0	0	0	0	0	0	0	0	0	186	233	0	0	0	0	0	0	0
19	0	0	0	0	0	0	0	0	0	0	141	244	42	0	0	0	0	0	0
20	0	0	0	0	0	0	0	0	0	0	36	223	18	0	0	0	0	0	0
21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Fig. 1.2 Representation of simple 2D image (left), the image taken from the MNist dataset[34]. The image consists of the pixel, which is each pixel has a numerical value between 0-255 (right).

local context, and structural knowledge. Preserve the pixel relationship information is very important and generally needed to make network learning more robust which is motivation of this study. The propose methods also have a better trade-off between accuracy and efficiency. We offer methods with high accuracy capabilities with less computational costs.

1.2 Pixel Relationship

Pixel is the picture element that is the basis of all digital image processing[66]. Pixel is the smallest element in a digital image and contains an intensity value indexed as (x, y) where x is a column, and y is a row representing the pixel's location in the image. Figure 1.2 shows an example of a simple representation of a 2-dimensional image with a size of 19x22 pixels. Each pixel has a numeric value in the range 0-255. In this example, the value 0 represents black, and the value 255 represents white. Meanwhile, 0-255 represents a different level of gray according to the tendency of the pixels value.

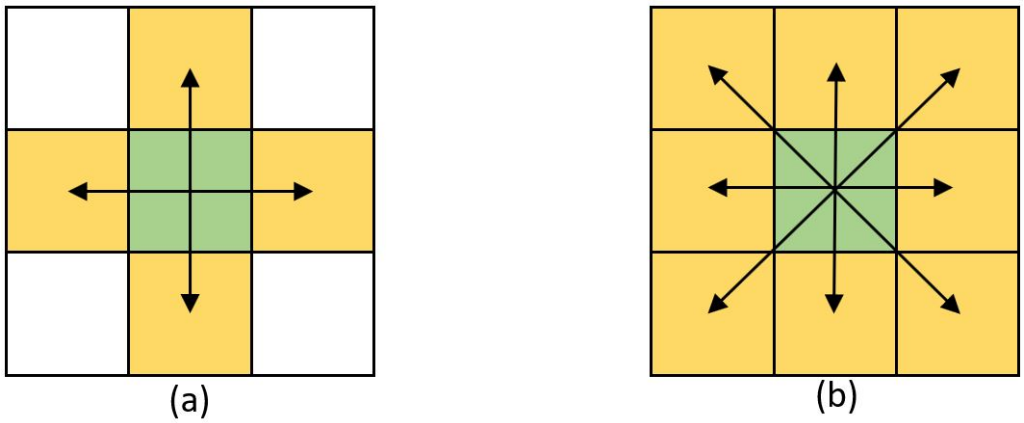


Fig. 1.3 Pixel relationships : (a) 4-connectivity, (b) 8-connectivity

Each pixel in the image has a side-by-side neighbor. Each pixel $p(x,y)$ has a neighbor with coordinates $(x,y-1),(x,y+1),(x-1)$ and $(x+1,y)$ called the 4 neighbors of the p. Each pixel $p(x,y)$ also has neighbors with coordinates $(x+1,y+1),(x+1,y-1),(x-1,y+1)$ and $(x-1,y-1)$ which are called 4 diagonal neighbors of p. The concept of the neighboring pixel relationship is illustrated in Figure 1.3.

Pixels neighborhood relationship approach recently studied to solve the several problem in computer vision. In image processing, pixel relationship information is used as a kernel or convolution matrix for blurring, sharpening, embossing, edge detection, and more. All task is performed by preparing a convolution between the kernel and an image.

Recently, pixel relationships information has been combined with convolutional neural network (CNN) to provide structure reasoning for semantic or instance segmentation. Wei[84] presented an affinity space for semantic segmentation that highlights structure by using co-occurring output patterns between neighboring pixels. Ke[26] introduces the concept of Adaptive Affinity Fields (AAF) to preserve and fit the relationships between neighboring pixels in the label space.

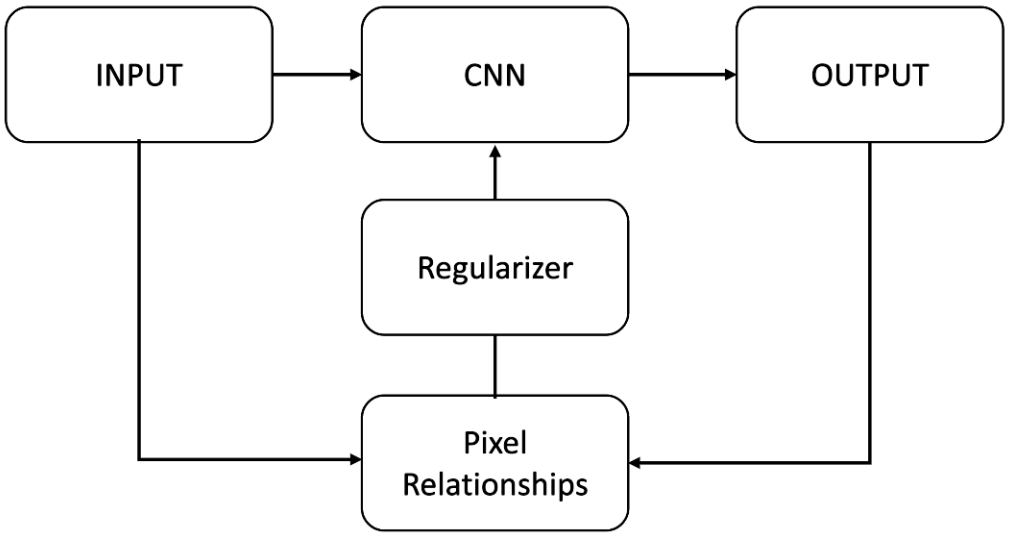


Fig. 1.4 Illustration of the general system of all architectures uses in this study. Image as basic information uses as input for CNN, and then CNN learns the feature of the image to generate the output. Furthermore, the pixel relationships are extracted from the output and input and used as a regularizer for CNN.

Recently, several methods based on the graph theory approach have also been used to learn the pixel relationships information to its neighbors[81], [42], [59]. The relationship information between neighboring pixels is captured from the image by being represented as a graph, where a pixel represents a node, and the intensity similarity between pixels is represented as a weighted edge between two nodes. However, all previous study constructs graphs and defines weighted edges only from predicted images while ignoring ground truth in defining graphs. Groundtruth is very important to be involved in defining a graph for an accurate learning process.

In this study, we proposed methods for utilize pixel neighborhood information as additional constrain to make deep learning methods more robust by using graph theory approach and morphological approach. We proposed graph-based smoothing regularizer that treated the network paying attention to small and tiny vessel connections. It treated the image as two graphs by calculating the graph laplacians on the vessel region and background regions.

Furthermore, we proposed regularizers based on Euler Characteristic, which calculate the number of isolated objects for accurate segmentation of vessel regions. Figure 1.4 show the illustration of the general system of all architectures uses in this study. Image as basic information uses as input for CNN, and then CNN learns the feature of the image to generate the output. Furthermore, the pixel relationships are extracted from the output and input and used as a regularizer for CNN.

1.3 Regularization in Deep Learning

Regularization allows for efficient fine-tuning of network layers, resulting in improved efficiency. This is consistent with previous research, which found that using semantic regularization in a deep network improves accuracy and convergence speed[53].

Generally, the learning process of the neural network is maximized using task-specific regularization techniques to avoid overfitting and push the sparsity of the network. The authors in [71] proposed a regularizer with scribbles, combining partial cross entropy and normalized cut for weakly-supervised segmentation.

Regularization based on graphs is used in many applications in the literature. Zeng [79] employs graph Laplacian regularization into a deep learning architecture for real image noise removal. On a patch-by-patch basis, the graphs are built from the CNN outputs. Dinesh [12] proposed a signal-dependent feature graph Laplacian regularizer (SDFGLR) for denoising imperfection 3D point cloud due to the acquisition process means that point clouds are often corrupted with noise. The author assumes that the normal surfaces calculated from point coordinates are smooth in parts with regard to the signal-dependent graph Laplacian matrix. However, the studies need to modify the network archi-

tecture to deploy graph Laplacian as a layer. Changing the architecture makes the complexity increase. Therefore, this study only adds the proposed methods on the last part of baseline CNN without changing the architecture.

Ando [2] introduced generalization limitations to learn graphs utilizing the characteristics of the graph in Laplacian regularization. This study showed in particular the relevance of laplacian normalization and a decrease in graphic design dimensions. The author stated that the typical L-scaling standard procedure is unsatisfactory, because the factors of normalization might vary substantially throughout a pure component.

In the segmentation task, Li[38] applies graph convolution into the semantic segmentation task use Laplacian to perform reasoning directly to the feature space. Lu[44] generates a neighborhood graph that shows the relationship for each point's neighboring points and then filters the neighborhood graph using Chebyshev polynomials.

Regularization based on the graph is also used in the Super-resolution task. Xu[75] proposed the graph Laplacian regularization to preserve the spectral information in Hyperspectral image. Liu[41] employe graph Laplacian regularizer to capture the essential piecewise smooth characteristic of the depth map, which has desirable filtering properties. In Hu et al.[21] utilize graph Laplacian regularization for generalized piecewise-smooth image denoising and super-resolution. However, the studies mentioned above are not considering the ground truth image in defining the graph. The graph is only derived from input images or predicted images.

Different from existing methods, in this study, we proposed a graph based smoothing regularizer that considers the image into two regions by calculating graph laplacians on vessels and its background areas. The proposed regularizer is used as a objective function in the deep CNN framework makes the network can efficiently learn the pixel connectivity on image segmentation results. And

then, we propose Graph Laplacian Regularization based on the Differences of Neighboring Pixels (GLRDN) by constructing graph laplacian from both prediction image and ground-truth image. The graphs use the pixel as a node and the edge defined by the "differences" of a neighboring pixel instead of the similarity between pixels. The basic idea is, if pair-wise pixels belonging a similar class, the differences are small. Otherwise, the differences are big if pair-wise pixels belonging a different class. Then, CNN uses a ground-truth image as a reference to make the differences of the prediction image more similar to the ground-truth image by minimizing the differences between the two images. Specifically, by using the differences of neighboring pixels as a Laplacian regularization term into a deep learning framework, we are successfully capturing pixel relationships information than pure CNN-based approaches.

Furthermore, we propose regularization based on the topological information for image segmentation. This regularizer utilizes the Euler characteristic (EC) to define isolated objects on predicted output and ground truth. By define the number of isolated object, regularizer based on EC learn to minimize the number of isolated object on segmented outputs.

1.4 Application in Retinal Blood Vessel Segmentation

To evaluate the effectiveness of the proposed methods, we implemented the proposed method on medical image segmentation. In deep learning domain, segmentation area has been widely utilized to support medical image analysis [25, 23].

Image segmentation plays an important role in medical imaging applications, specifically in retinal blood vessel on fundus image segmentation.

Characteristics of blood vessels in retina guide an ophthalmologist to diagnose pathologies of different eye anomalies such as age-related macular degeneration (ARMD) and diabetic retinopathy (DR)[67][1].

Diabetic retinopathy (DR) is one of the causes of visual impairment in aging population and signifies an increased risk of coronary heart disease [32]. As it involves changes in the shape and structure of blood vessels in the fundus, visual impairment is often detected by analyzing fundus images. Fundus images are color images that represent the inner surface of the human eye and are often used by doctors to study at the retina and the anatomical components of the eye.

Therefore, accurate identification of large and small retinal vessels with diameters of less than a pixel can be used as an early bio-marker for the diagnosis of DR [74][67]. Additionally, it helps to identify several physiological problems, specially hypertension and some other cardiovascular diseases [65].

However, blood vessel segmentation on fundus images has several challenges. Commonly fundus images are affected by noise. Moreover, examining the small blood vessels is difficult. It is also time-consuming to identify the disease-caused blood vessels, especially the changes of states of small vessels and their characteristics. The separation of the blood vessel is not an easy task because of the small and fragmented structure in a low contrasting retinal image.

Several automated systems of segmentation of blood vessels were developed to rectify the subjective detection of retinal blood vessels. Hence, automatic characterizing of retinal blood vessels is essential for detecting DR. To this end, computer-aided segmentation of retinal blood vessels, based on pattern recognition, and supervised and unsupervised machine learning, has been successfully proposed [60, 14, 30].

Chakraborti et al [5] presented a self- adaptive matched filter by combining the vesselness filter with the matched filter for the detection of blood vessels on the retinal fundus image. Tagore et al [69] presented a new algorithm for retinal blood vessel segmentation by using the intensity information of red and green channels of color fundus image. It helped to distinguish between vessels and its background in the phase congruency image.

Recently, several researchers have been implementing convolutional neural network (CNN) for retinal blood vessel segmentation. Various studies in recent times have reported improve performance of blood vessel segmentation based on deep learning techniques [57, 19, 45, 76, 48, 16, 54].

In [9], the authors developed a multi-label method based on supervised structured for the segmentation of retinal vessels. They applied pre-processing methods before feeding the image into the network. Ortiz et al [57], in contrast, proposed a deep CNN that ignored pre-processing steps and directly used raw RGB image as input. In [19], the segmentation network was divided into two steps: using multiscale CNN and using fully connected conditional random fields (CRFs).

Thick and thin vessel segmentation were considered using segment-level loss and pixel-wise loss [76]. Similarly, deep vessel segmentation was proposed that achieved high quality vessel probability map using a CNN network and a CRFs layer [16]. The fully CNN network generates probability maps, whereas carry out segmentation using dense global pixel correlation. The authors in [54] utilized stationary wavelet transform with a multiscale fully connected CNN to adapt with the varying width and direction of the vessel structure in the retina. However, aforementioned architectures used multiple steps that increases misclassification error and computational complexity. Several medical image segmentation methods using U-Net-based CNNs have been proposed for solving the segmentation problem and reducing the error rate [16].

The U-Net structure takes care of the sampling that is required to check class-imbalance factors. Furthermore, it is capable of scanning an entire image in just one forward pass, which enables it to consider the full context of the image [78, 24]. Hence, this study also considered using U-Net-like CNN architecture to evaluate the efficiency of the proposed regularizers.

Melinscak [51] presented retinal vessel segmentation system using ten layers of CNN. In [40], a structured prediction scheme was used to highlight the context information, while testing a comprehensive set of architectures. Fu et al [17] combined a typical 7-layer CNN with a conditional random field and reformulated a recurrent neural network to model long-range pixel interactions. Li et al [37] considered the vessel segmentation task as a cross-modality data transformation problem in a deep learning model. Dasgupta [10] proposed a neural network framework that iteratively classify pixels from the fundus image.

Several U-Net models based on convolutional neural networks (CNNs) have been developed by modifying the number of layers in the encoder and decoder architectures and have achieved better results for retinal blood vessel segmentation [45, 22, 39, 61]. Although the localization of the vessels has improved significantly with CNN, the fragmented small vessel identification is still a challenging task. Because sometimes it is located at the end of the vessel branch and failed to maintain the connectivity. Furthermore, it is challenging to detect isolated vessels in the low contrast background. Besides that, these approaches considered large and medium vessels as it appears in the ground truth. They failed to characterize the damaged tiny or small vessels, and hence unreliable diagnosis may occur. In this study, we aimed to delineate, in addition to the thick vessels, damaged tiny or small vessels that do not appear in the ground truth but appear in the original fundus images.

To address the small vessel connectivity problem, we introduced a graph-based smoothing (GS) regularizer that pushes the network to consider small vessel. The GS regularizer hold smoothness level by calculating graph laplacians between the vessel and surrounding background area.

We also proposed a different approach to penalize the network to pay more attention to small vessels by utilizing the number of isolated objects, as shown in the Figure 1.1. Furthermore, we observe that most of the existing architectures fail to accurately detect the vessel connections especially on the thinner and small branches of the vessels (Figure 1.1c).

However, for accurate segmentation, it important to preserve the vanished blood vessels as isolated objects so as to use them for finding entire vessel connections. Hence, this study is attempted to trace the retinal vessel connections using Euler characteristics (EC). The EC is an essential topological invariant for realizing the number of components and number of enclosed cavities to utilize the EC as a demonstration of vessel connectivity [7, 11].

We introduced a regularizer based on EC to minimize the number of isolated objects, especially in the small branches and tiny retinal vessel regions. The proposed method is based on the criterion of a vessel structure that is fully connected or if consisting of minimal number of isolated objects. Specifically, we proposed two regularizers, minimum number of isolated objects (MISO) and differences of number of isolated objects (DISO) between predicted and true segmentation groundtruth based on EC, to delineate tiny vessel regions. We developed a U-Net-like CNN architecture to evaluate the performance of the proposed regularizers in delineating small vessel connections. The interesting key point in this study is implementing mathematical topology approach in the neural network architecture.

The contribution of this study consist of three sections summarized as follows : (1) We proposed a regularizer based on the graph to capture the

smoothness of segmented regions. The graph is generated from the connectivity between adjacent pixels. (2) We proposed a regularizer based on graph Laplacian to capture the differences of neighboring pixels between prediction images and ground-truth images. (3) We proposed a regularizer based on Euler Characteristic to define the number of isolated objects in segmented images. The Euler characteristic is constructed from connectivity between adjacent pixels.

1.5 Structure of Thesis

The structure of this thesis is as follows:

In chapter 2, we introduced a regularizer based on laplacian graphs. The laplacian graph is constructed based on the information of neighboring pixels in the ground-truth image. We calculate the laplacian graph on the object and background parts separately. This regularizer is then added to the objective function of U-Net. We implemented this regularizer on the segmentation of the retinal vessels in the fundus image.

In chapter 3, we proposed a regularizer based on differences in neighboring pixels. From the differences of neighboring pixels information, a laplacian graph is then constructed. In contrast to our work in chapter 1, this study computes a laplacian graph of the difference between neighboring pixels in the output and ground-truth images. We also implemented this regularizer on retinal blood vessel segmentation on the fundus image.

In chapter 4, we proposed a regularizer based on a topological approach. We use the Euler characteristic to capture neighboring pixel information to define the number of isolated objects in the segmentation results. We propose this regularizer to enforce the network to minimize the number of isolated

objects. We implement this regularizer on retinal blood vessel segmentation, assuming the structure of blood vessels is connected.

Chapter 5 summarizes the work, limitations of the proposed approach, and potential developments for future work.

Chapter 2

Graph Laplacian Regularizer based on Prediction Images

This chapter explain the proposed graph laplacian regularizer based on prediction images called graph based smoothing (GBS) regularizer. GBS considers the image into two regions by calculating graph laplacians on vessels and its background areas. The proposed regularizer is used as a objective function in the deep CNN framework makes the network can efficiently learn the pixel connectivity of the small or isolated blood vessel structure. The effectiveness of our proposed regularization term was evaluated and compared using U-net architecture and baseline U-net. The performance of the proposed approach was also compared with the state-of-the-art networks model in reconstructing the small and isolated vessel regions.

In parallel to track large vessel we are interested to reconstruct the small or isolated vessels. Paying attention to segment the small vessels in the fundus region, we considered to define a regularizer that is based on the graph laplacian smoothing method. We evaluate the effectiveness of our proposed regularizer using U-Net architecture [62]. The schematic diagram in Fig. 2.1, describes the

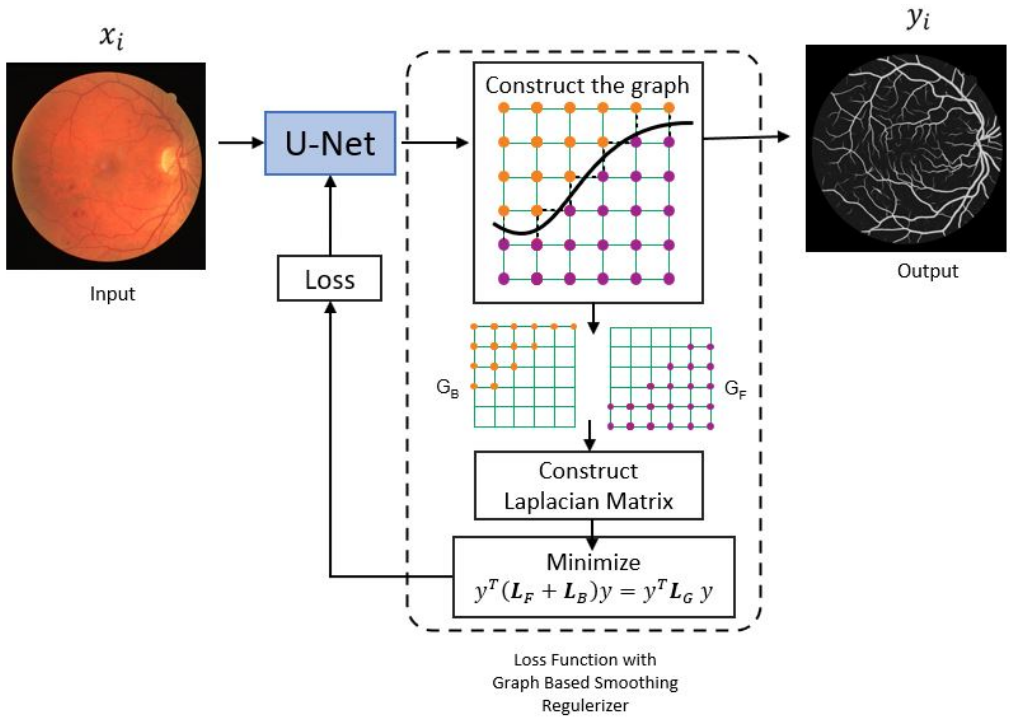


Fig. 2.1 U-Net framework using graph laplacian regularizer

proposed graph based smoothing regularizer for small vessel reconstruction in the U-net framework.

2.1 Pixel Adjacency Graph

Graph theory has a long history in mathematics and has been used in many fields of science and engineering. A graph is a representation of the set of elements and the set of relationships between these elements. These elements are called nodes or vertices, and their relations are called edges[47]. In this paper, graph is used to modeling every pixel in image data as node and use neighboring pixel information as edge structure for each node.

The proposed graph based smoothing regularizer is based on the graph laplacian matrix. Graph laplacian can be obtained by constructing the adjacency graph and diagonal matrix. Here we briefly explained the computation of graph laplacian matrix in reconstructing the small vessel by using a simple

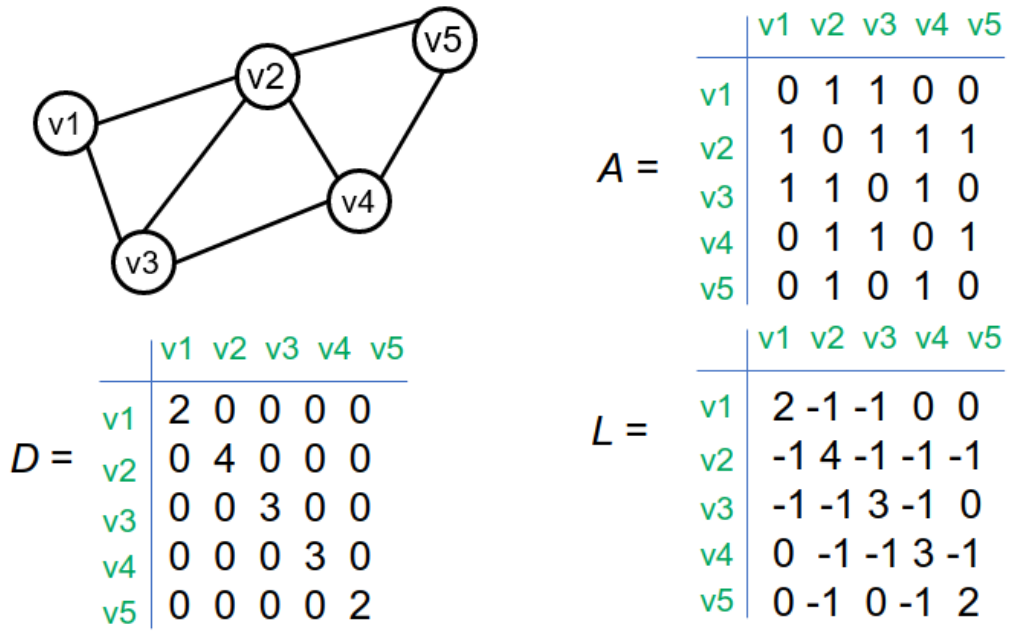


Fig. 2.2 Computation of laplacian matrix from a simple graph

graph. It includes four vertices and five edges as shown in Fig. 2.2. The graph can be transformed into a matrix by counting the number of edges between the two adjacent vertices. The two vertices can be said to be adjacent if it is connected by an edge. The example graph in figure 4 has 4 vertices and hence the adjacency matrix A of size 4x4 can be constructed. If the vertices are connected to other vertices, then it is accounted as an edge and fill the matrix.

Let us consider the matrices $v1$ and $v2$, which are connected by an edge and hence the matrix is filled with some value which corresponds to the number of edges of $(v1, v2)$. The vertices $v1$ and $v2$ are said to be adjacent if their pair is an edge. If the two vertices are not connected then it is filled to be zero. Graph laplacian matrix L can be calculated by using the formula $L = A - D$. The parameter D represents the diagonal matrix and it can be obtained by calculating the summation of the row values of the adjacency matrix. In addition, to make the representation easier, we constructed the image into a pixel adjacency graph. Thus the set of vertices of the graph are used to indicate

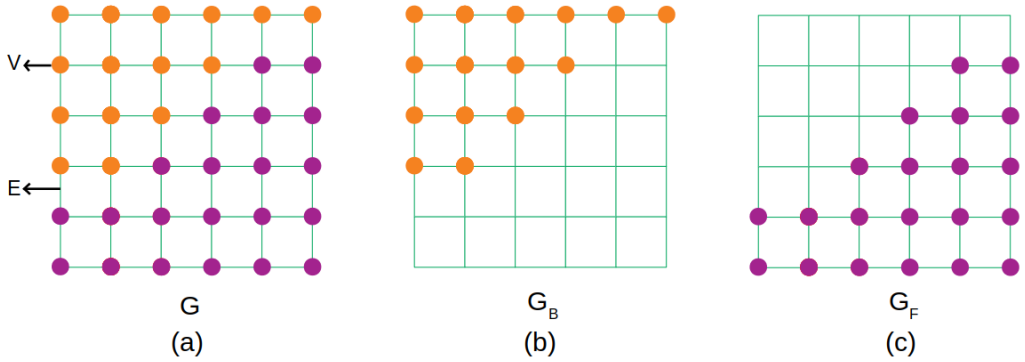


Fig. 2.3 Representation of set of nodes of the pixel graph to compute boundary between two regions, a) image as an graph, b) background region graph and c) foreground region graph.

the set of pixels on the image. The edges are used to link the neighboring pixels. The graph size depends on the image size.

In this study, we formally defined two graphs G_F and G_B for foreground and background, respectively as shown in Fig. 2.3. G_F and G_B includes a pair of (V_F, E_F), (V_B, E_B), respectively. The parameters V_F and V_B are finite set of elements called vertices, and $(E_F = \{(j_F, k_F) | j_F \in V_F, k_F \in V_F\}$ and $(E_B = \{(j_B, k_B) | j_B \in V_B, k_B \in V_B\}$ are edges.

2.2 Graph Based Smoothing as Regularizer

The segmentation of the blood vessel especially in detecting the small or thin isolated vessels is not an easy task. With the help of graph laplacian, the image pixels can be interpreted as node. Every node is connected with every other node in the graph. Let us consider $(x_i, t_i) | i = 1, ..., M$, where x_i is a i^{th} input data from the training dataset X, and t_i is a i^{th} from data target or label T. The number of training samples and labels is denoted by M and N, respectively.

The proposed CNN based U-Net architecture is trained to predict the output image pixels y_i from a given input image pixels x_i. For each edge of foreground $(j_F, k_F) \in E_F$ and background $(j_B, k_B) \in E_B$ of the pixel graph, the similarity

$\beta_{(j_F,k_F)}$ and $\beta_{(j_B,k_B)}$ is defined as

$$\beta_{(j_F,k_F)} = 1 - |t_{j_F} - t_{k_F}| \tag{2.1}$$

$$\beta_{(j_B,k_B)} = 1 - |t_{j_B} - t_{k_B}| \tag{2.2}$$

We introduced the regularization term for smoothing S based on foreground region F and background region B as

$$\sum_{(j_F,k_F)\in G_F} \beta_{j_F,k_F}(y_{j_F} - y_{k_F})^2 = y^T(D_F - A_F)y = y^T L_F y \tag{2.3}$$

$$\sum_{(j_B,k_B)\in G_B} \beta_{j_B,k_B}(y_{j_B} - y_{k_B})^2 = y^T(D_B - A_B)y = y^T L_B y \tag{2.4}$$

where, L_F and L_B is Laplacian graph. The adjacency and diagonal matrices is defined as following

$$\begin{pmatrix} A_F = \beta_{(j_F,k_F)}, D_F = \sum_{j_F=1}^{N} \beta_{j_F,k_F} \\ A_B = \beta_{(j_B,k_B)}, D_B = \sum_{j_B=1}^{N} \beta_{j_B,k_B} \end{pmatrix} \tag{2.5}$$

Graph Based Smoothing regularizer S can be written as

$$S = y^T(L_F + L_B)y = y^T L_G y \tag{2.6}$$

The objective function O applied in this study is the summation of the binary cross entropy of each label with the regularization term using graph based smoothing, which is defined as

$$O = \sum_{i}^{M} \{t_i log(y_i) + (1 - t_i) log(1 - y_i)\} + \lambda \sum_{i}^{M} S \tag{2.7}$$

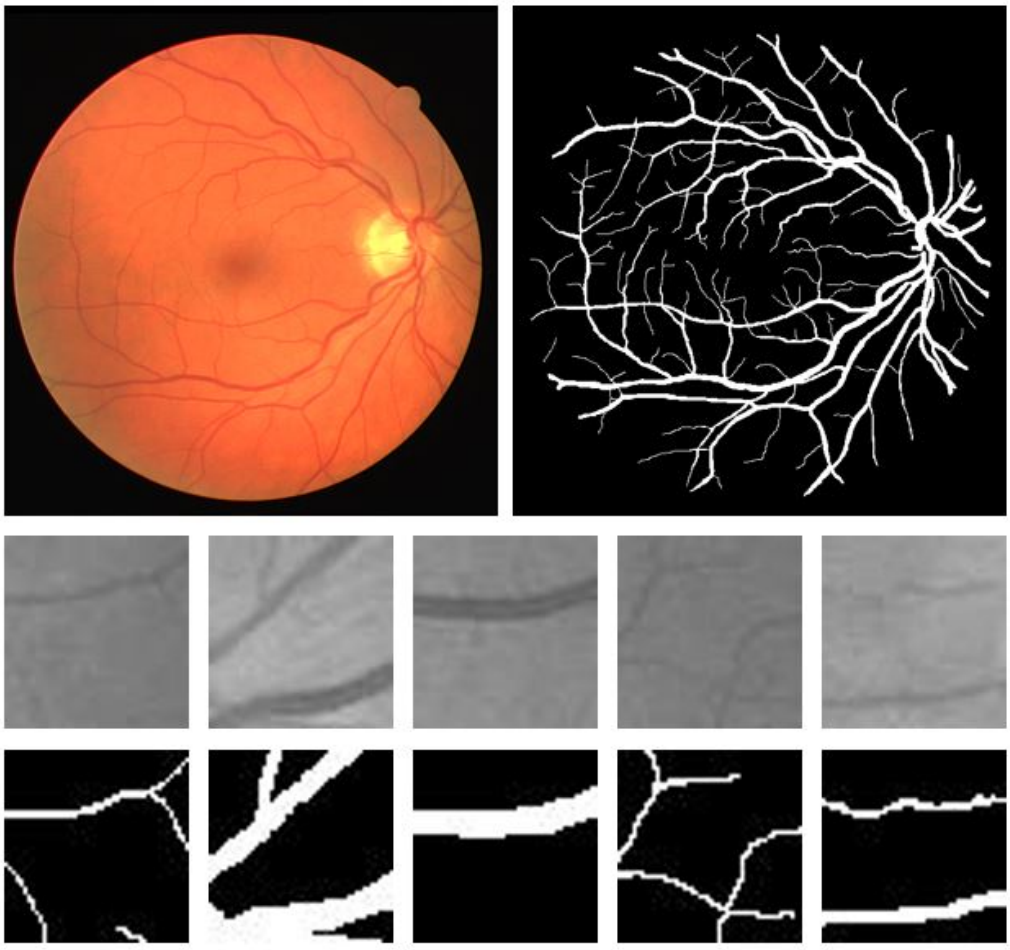


Fig. 2.4 Example of fundus image: The first row shows the full of fundus image with the ground truth, the second row shows patches in grayscale mode and their ground truth in the third row.

One of the complexity of the graph construction is depends on the size of images. Therefore, we calculate laplacian graph randomly on background and foreground respectively to reduce the complexity.

2.3 Dataset

In this study, the proposed method is evaluated on the DRIVE, STARE, and CHASEDB1 datasets. The DRIVE dataset consists of 40 fundus images. The manual segmentations of the vessels is provided for the all the datasets. The vessels of small width or isolated pixels were defined as small vessels. The size

of the image is 565 x 584 pixels with 8 bits per color channel. The number of training and testing data used in this study is 20 and 20, respectively. The DRIVE dataset is captured by Canon CR5 nonmydriatic 3CCD camera at 45° field of view, which is comprised of 40 images. The size of images is 768x584 pixels with 8 bits per color channel. The STARE dataset was captured by TopCon TRV-50 fundus camera at 35° field of view, which is consisted of 20 images. Each image has a size of 605x700 pixels and has 24 bits per pixel. The CHASEDB1 dataset included 40 images with the size 999x960 pixels. Considering our aim was evaluating the proposed segmentation error, manual segmentation by an ophthalmologist was used as a reference ground truth image. An example of a fundus image along with the ground truth is shown in the Fig. 2.4 with some parts enlarged as patches to show in detail the parts of the small ship.

The Cartoon Set[Set] and Manga109[50] datasets also use for additional experiments in super-resolution tasks. Cartoon Set is a collection of 2D cartoon avatar images that are selected randomly. With a total of 1013 potential combinations, the cartoons are divided into ten artwork categories, four color categories, and four proportion categories. We started our experiment on this simple dataset at first. Manga109 was collected by the Aizawa Yamasaki Matsui Laboratory from the University of Tokyo. The collection is designed for use in academic study on Japanese manga media processing. We resized all samples from Cartoon Set and Manga109 to 128x128 pixels. We took 1k images for training and 100 images for testing. We use Cartoon set and Manga109 dataset for evaluation on image super-resolution tasks.

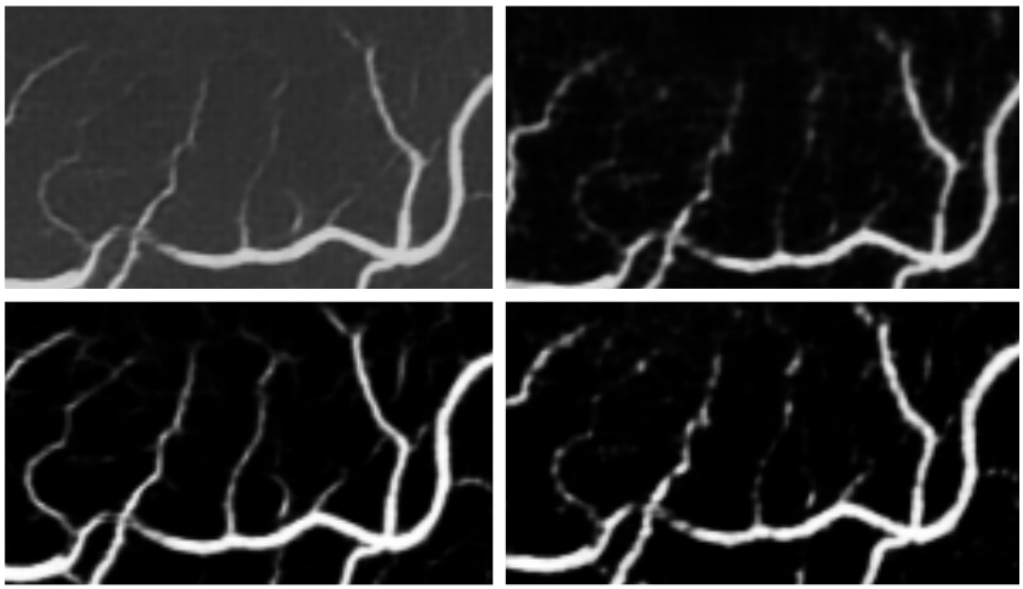


Fig. 2.5 First row : Left $\lambda = 0.0001$ Right $\lambda = 0.00001$. Second row : Left $\lambda = 0.000001$ Right $\lambda = 0.0000001$

2.4 Evaluation and Performance Measures

To evaluate the performance of proposed methods, several metrics measurement is used, such as sensitivity (Sn), specificity (Sp), Accuracy (Acc), Jaccard Index, Precision-Recall Curve (PRC) and Area Under the Curve (AUC).

Accuracy calculates the proportion of the predicted vessel pixels, which are true vessel pixels. Its can express as follows:

$$Acc = \frac{TP + TN}{TP + FN + TN + FP} \tag{2.8}$$

where TP is true positive, which defines the vessel pixel that is properly classified, FN is a false negative which defines the vessel pixel is classified as a non-vessel pixel. TN is a true negative which referred to as a non-vessel pixel that correctly labeled. FP is a false positive, which defines a non-vessel pixel that is classified as a vessel.

Sensitivity evaluates the proposed structure's ability to detect vessel pixels. Sensitivity (Sn) is calculate as follows:

$$Sn = \frac{TP}{TP + FN} \tag{2.9}$$

Specificity evaluates the ability of the structure to detect backround pixels. The Sp is defined as follows :

$$Sp = \frac{TN}{TN + FP} \tag{2.10}$$

Precision (Pr) indicates the rate at which positive predictions pixel are correctly classified and giving information about the validity of segmentation result. In this paper we define precision and then used for estimate Precision-Recall Curve (PRC). Precision is define as follow:

$$Pr = \frac{TP}{TP + FP} \tag{2.11}$$

Intersection-Over-Union (IoU) also known as the Jaccard Index is overlap area (O) between prediction output and ground truth divided by the union area (U) between prediction output and ground truth. the IoU can defined as follows:

$$IoU = \frac{O}{U} = \frac{TP}{TP + FP + FN} \tag{2.12}$$

Dice Coefficient (F1 Score) is 2 times the overlap area (O) divided by the total number of pixel (N) in prediction output and ground truth. F1 score is define as follow:

$$F1 \quad Score = \frac{2 * O}{N} = \frac{TP}{TP + \frac{1}{2}(FP + FN)} \tag{2.13}$$

We also calculated the AUC value using the receiver operating characteristic curve to evaluate the proposed capability to detect vessel pixels. The receiver

operating characteristic (ROC) curve is used to determine AUC. ROC curve is pixel comparison between the ground truth image and prediction output based on confusion matrix.

2.5 Experimental settings

In order to find a suitable λ value for the regularizer, we varied this parameter such as λ = 0.0001, 0.00001, 0.000001 and 0.0000001 for accurate small vessel construction. we used Adam optimizer with learning rate of 0.001 and 100 epoch to train the model. We chose binary cross entropy and dice loss because it greatly improved the performance of the model. In our experiments, we compared our proposed network with graph based smoothing regularizer over the baseline U-net without regularizer.

We use U-Net[62] model as our baseline model which is contains encoder modules and decoders. The encoder path has three units. In the first unit, two convolutional layers are followed by ReLU and max-pooling layer, with each 32 feature maps. In the second unit, two convolutional layers are followed by the ReLU and max-pooling layer with 32 and 64 feature maps. In the third unit, two convolutional layers with each 128 feature maps. The decoder path has four convolutional layers followed by ReLU and one convolution layer without ReLU. After the decoder path, the output is upsampled with the factor of 2. Lastly, the feature maps were joining with the encoder layers by skip connections. Sigmoid classifier function applies to probability map, and then we added GLRDN after the sigmoid function. We used 200 epochs and set the batch size to 16. The learning rate is set at 0.001 and decreased ten times every 25 epochs to ensure network convergence. We utilized Adam to optimize the network. Experiments have set the regularization parameters $lambda$ to 1e-5. The proposed method is implemented using the Pytorch and train using

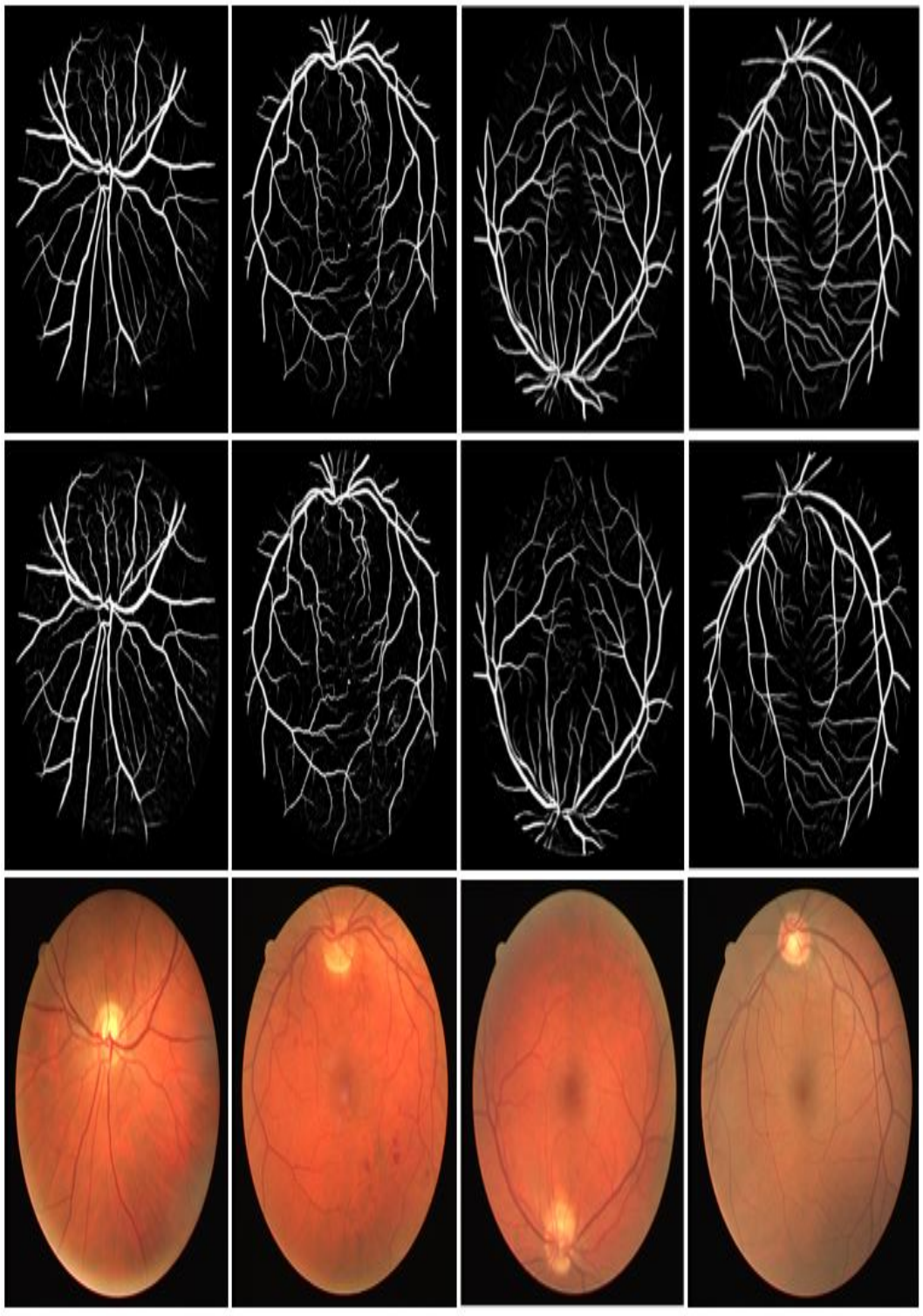


Fig. 2.6 Segmenting both large and small vessels. From left to right : Fundus image, results of network without regularizer, and with proposed graph based smoothing regularizer

GPU machine Nvidia GeForce GTX 1080/PCIe/SSE2 graphic card on Intel(R) Core(TM) i7-6700K CPU @ 4.00GHz Processor, 32 GB of RAM.

The number of training and testing data of the DRIVE dataset used in this study is 20 and 20, respectively. In order to ensure that there are enough training images, we divided the images into patches with the size of 48x48 pixels. Hence the total number of patches generated on every image is 4.750 patches. Thus a total of 950.000 patches was generated from the training images to increase the size of the dataset. We also use same setting for experiment on STARE datasets. Example gray scale image patches and their ground truth is shown in Fig. 2.4.

Before training process, we enhanced the dataset with some pre-processing steps. The image is converted into grayscale followed by data normalization and contrast-limited adaptive histogram equalization (CLAHE). We applied normalization to keep the image into the same scale and CLAHE technique to enhance the contrast of the grayscale image.

The graph based regularizer adopts a 4-neighbor strategy, so it requires a large memory to hold a 4× ground truth label and predicted probabilities when calculating its difference matrix. We use two approaches to overcome this problem: firstly, when choosing the neighbor pixels in a square region, we set the region size to be 3×3 as suggested. Secondly, we calculate the graph from n samples every minibatch randomly. In experiments, we use n=5 sample images for each minibatch. This approach reducing the complexity time significantly.

2.6 Results

The appropriate regularizer parameter value λ to reconstruct the width of the small vessel retained with adequate information was found with the value of 0.000001 (Fig. 2.5). The proposed graph based smoothing regularizer cou-

Table 2.1 Performance comparison of the U-Net without and with our proposed GBS regularizer on DRIVE datasets using binary cross entropy loss and dice loss.

Method	**Sn**	**Sp**	**Acc**	**AUC**	**Jaccard**	**F1-Score**	**PRC**
bce	0.7583	0.9826	0.9551	0.9691	0.6742	0.8054	0.8893
bce+gbs	0.7802	0.9854	0.9602	0.9817	0.7063	0.8279	0.9195
dice	0.7440	0.9856	0.9560	0.9665	0.6753	0.8061	0.8840
dice+gbs	0.8352	0.9789	0.9613	0.9797	0.7256	0.8410	0.9162

pling with U-net succeeds in reconstructing both large and small vessel pixels compared over U-net without regularizer is presented in Fig. 2.6.

Furthermore, the improvement of the disjointed vessel connectivity is clearly observed from the patch based fundus image as shown in Fig. 2.7.

The graph based smoothing regularizer in the U-net depends on the pixel connectivity criterion. Hence, when we compared the AUC value of architecture without regularizer, our approach resulting high AUC value with large number of vessels. The qualitative and quantitative results of our approach in segmenting most of the vessels is shown on the image patch examples (Fig. 2.8).

Our approach achieves significantly higher performance with high AUC value (0.979) in segmenting the small retinal blood vessels than the other state-of-the-art methods is presented in Table 2.2 . Sensitivity of our method is moderate and it is almost similar with other conventional methods. However, the proposed approach achieves considerably higher specificity (0.99) than all the other methods, which reconstructs to the segmentation of more vessels. When we considered accuracy, our approach scored higher value (0.95), and very close to [73] and [40] methods. Fig. 2.9 explains examples of the analysis of vessel reconstruction, where this study focused on the small vessels. The segmentation results are colorized to demonstrate the confusion matrix: green

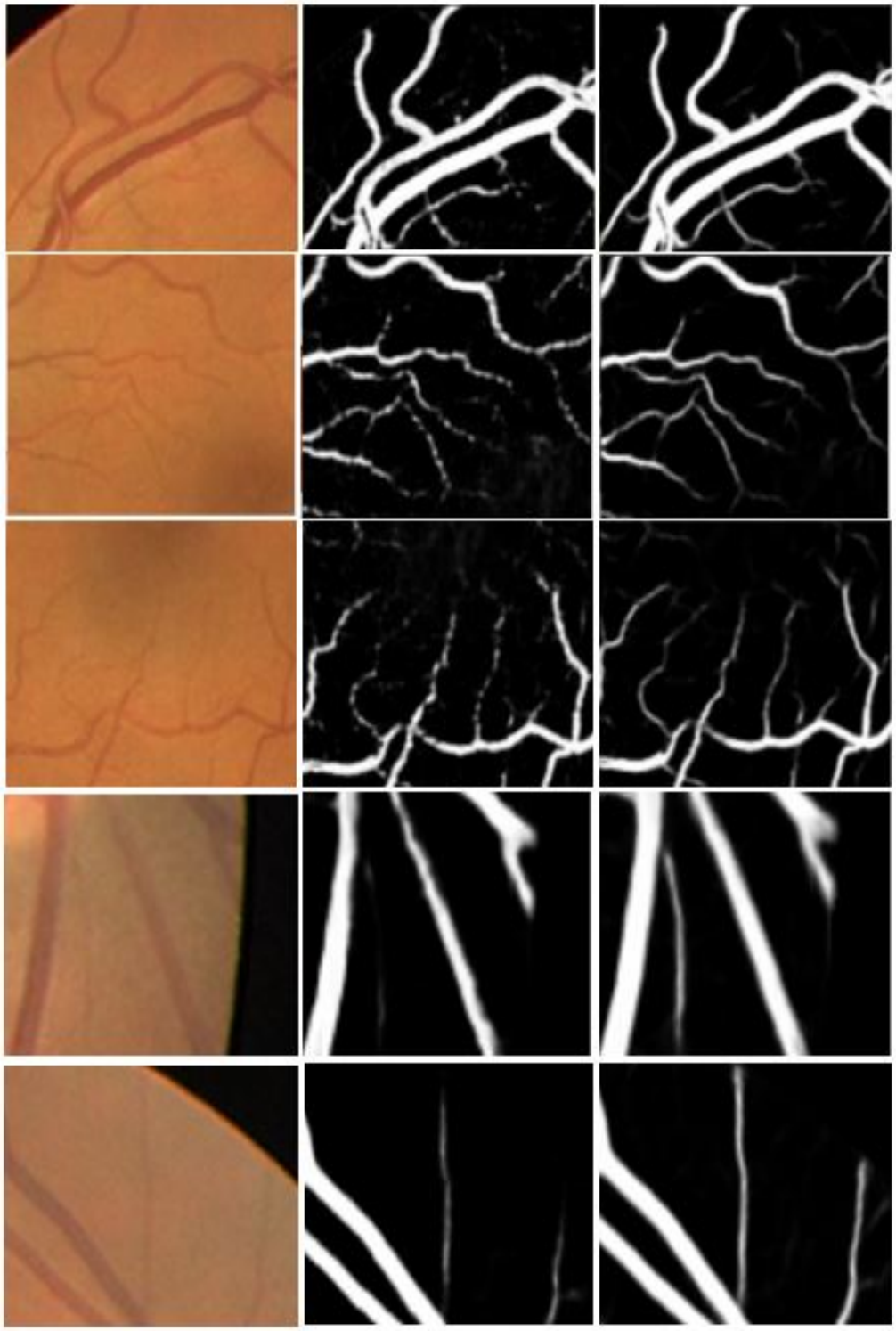


Fig. 2.7 Segmenting small vessels. From left to right :patch fundus image, results of network without regularizer and with proposed graph based smoothing regularizer

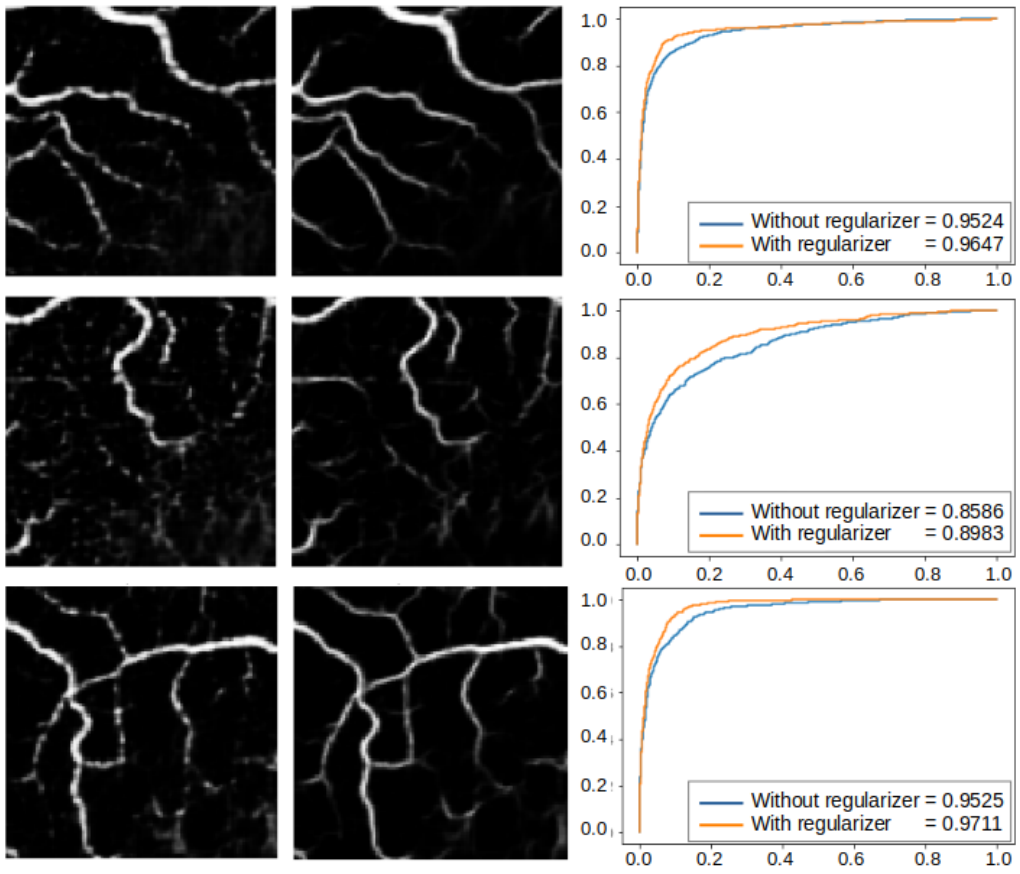


Fig. 2.8 Segmented small vessels. From left to right: without graph based smoothing regularizer, proposed graph based smoothing regularizer and AUC performance

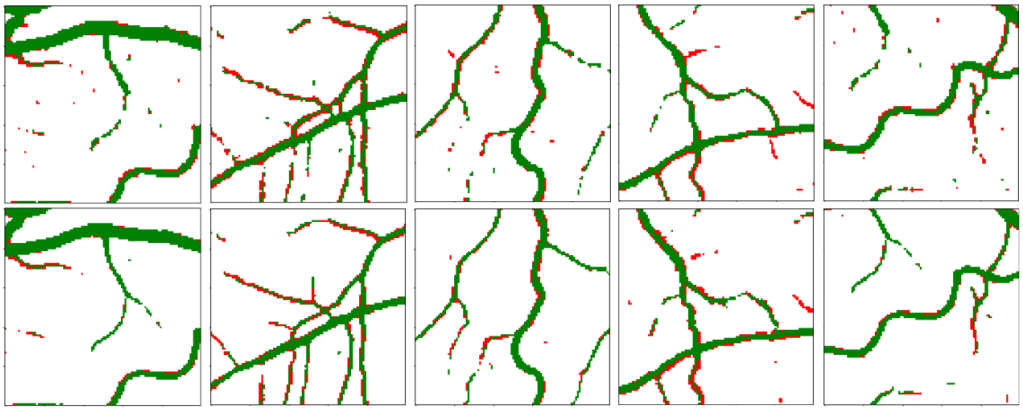


Fig. 2.9 Small vessel segmentation examples: Green, and red colors represented TP, and FN respectively. First row shows the segmented vessels without graph based smoothing regularizer and the second row shows the results of the proposed graph based smoothing regularizer.

Table 2.2 Performance comparison of our approach with state-of-the art methods interns of sensitivity, specificity, accuracy and AUC

Year	**Method**	**Sn**	**Sp**	**Acc**	**AUC**
2016	Azzopardi et al. [3]	0.7655	0.974	0.9442	0.9614
2016	Khan et al. [27]	0.7373	0.9670	0.9501	-
2016	Zhao et al. [82]	0.7420	0.9820	0.950	0.8620
2018	Marin et al. [49]	0.7067	0.9801	0.9452	0.9588
2018	Orlando et al. [56]	0.7897	0.968	-	-
2016	Fu et al. [17]	0.7294	-	0.9470	-
2015	Wang et al. [73]	0.8173	0.9733	0.9533	0.9475
2016	Liskowski et al. [40]	0.7569	0.9816	0.9527	0.9738
-	Proposed method	**0.8352**	0.9789	**0.9613**	**0.9797**

pixels indicate the TPs and red pixels represent the FNs. Graph based network showed highly acceptable performance for small vessel reconstruction. Our method almost reconnecting all isolated vessels and it can be observed from the lager number of TP pixels. When we analyzed the network without our proposed regularizer, it produced large number of FN pixels.

We also demonstrated the superiority of our methods over baseline methods on all DRIVE test images dataset in figure 2.11-2.15. The ROC curve for all DRIVE test images dataset also presented in figure 2.16-2.20.

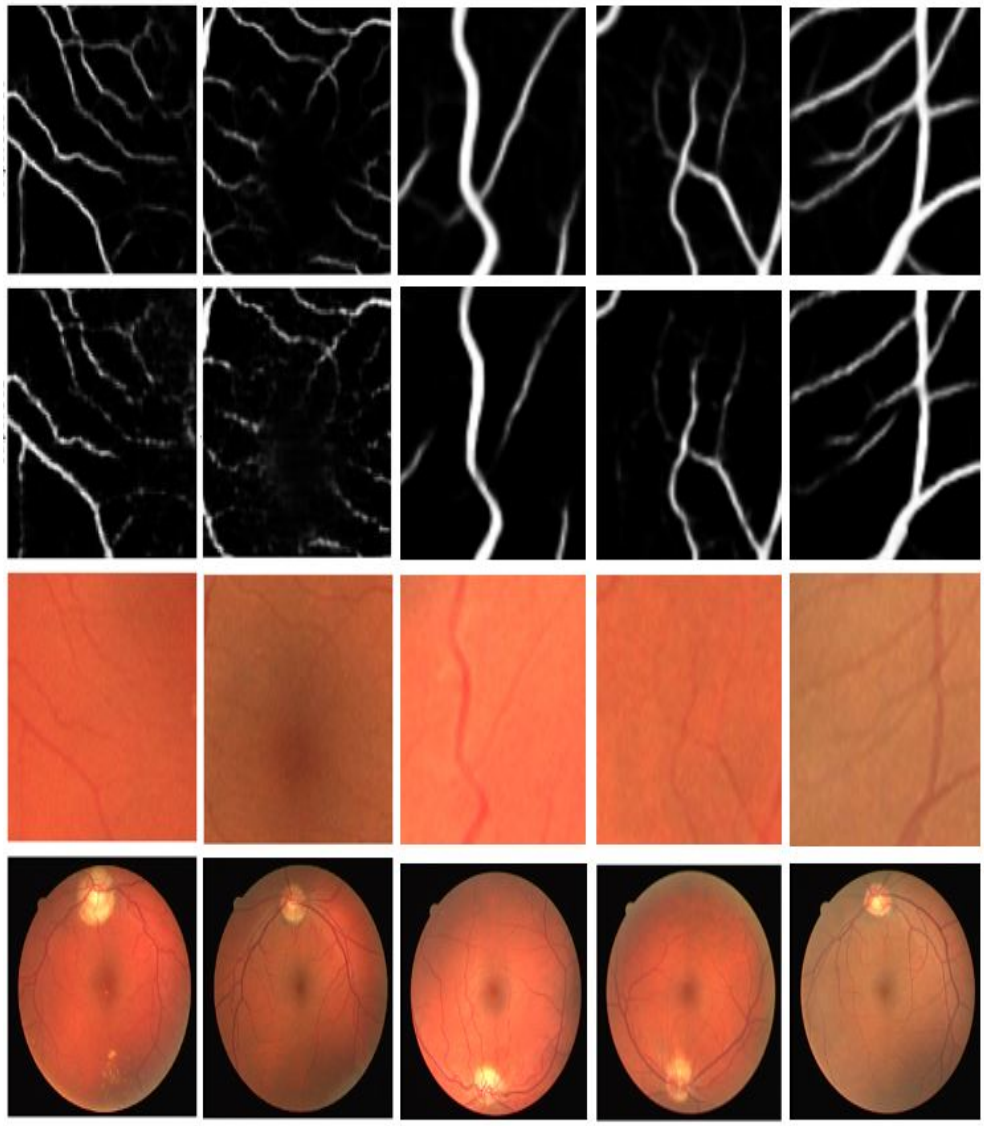


Fig. 2.10 Segmenting small vessels. From left to right: Fundus image, patch image, results of network without regularizer, and with proposed graph based smoothing regularizer

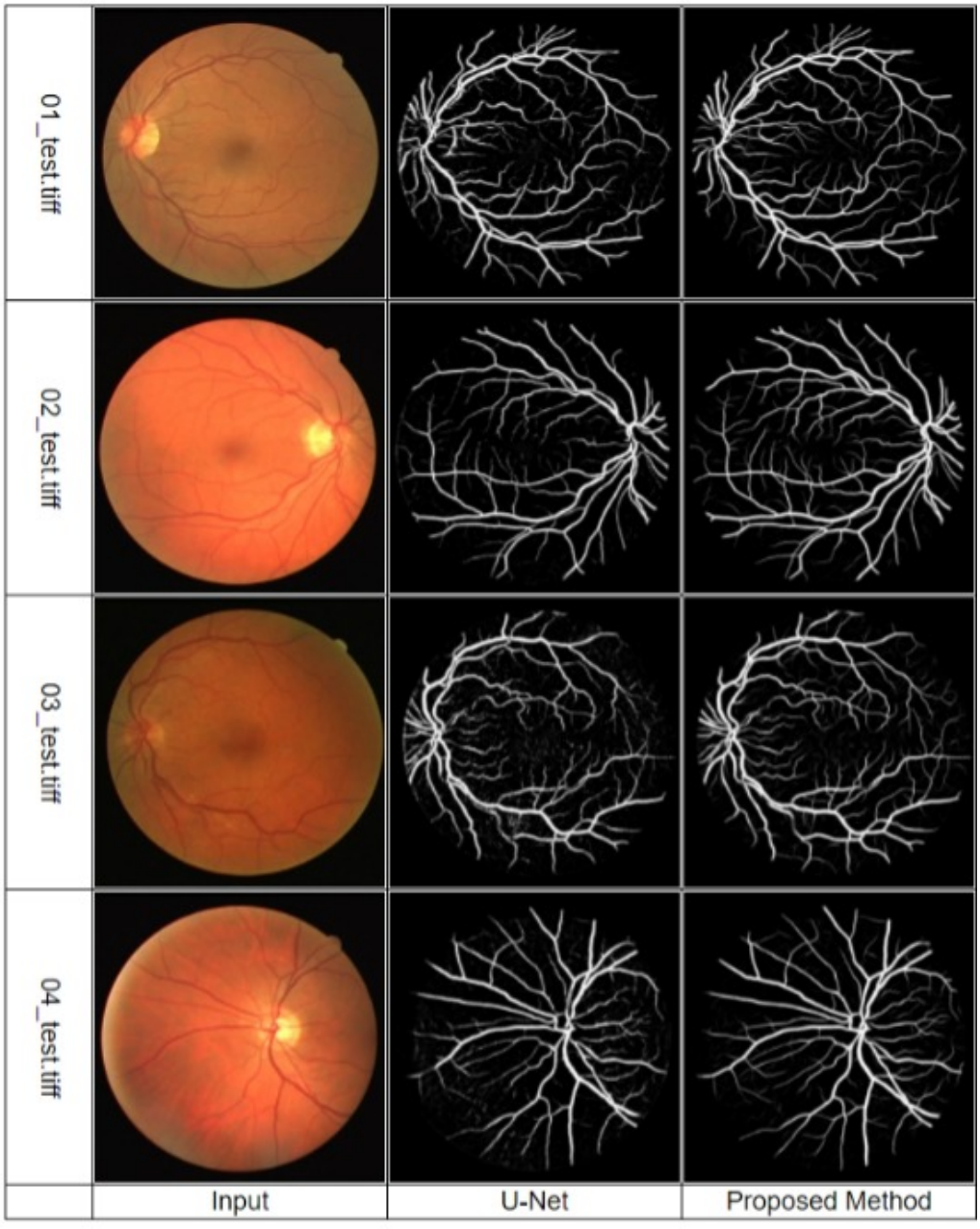


Fig. 2.11 Comparation of segmentation results between baseline and proposed method on DRIVE dataset (Cont.)

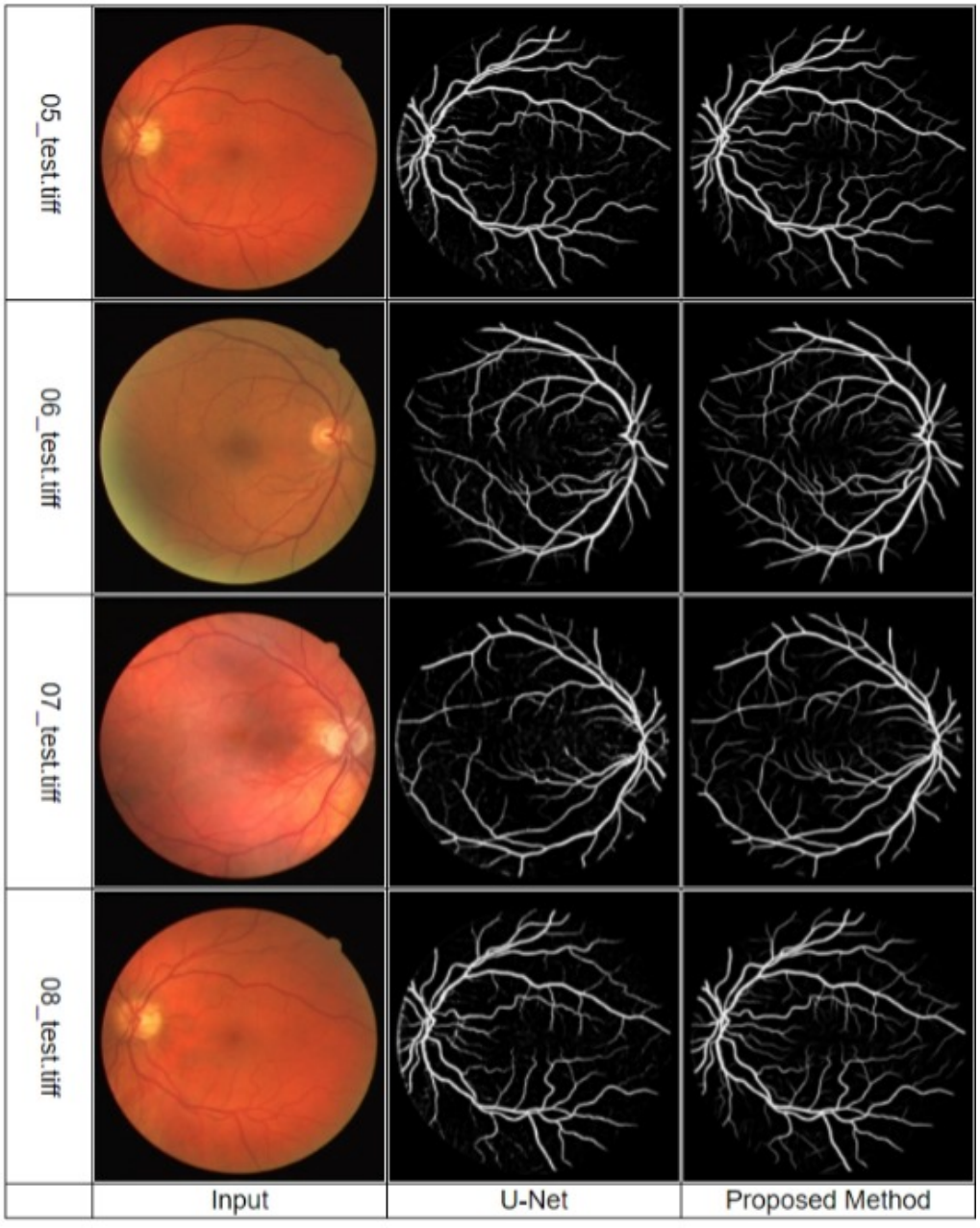


Fig. 2.12 Comparation of segmentation results between baseline and proposed method on DRIVE dataset (Cont.)

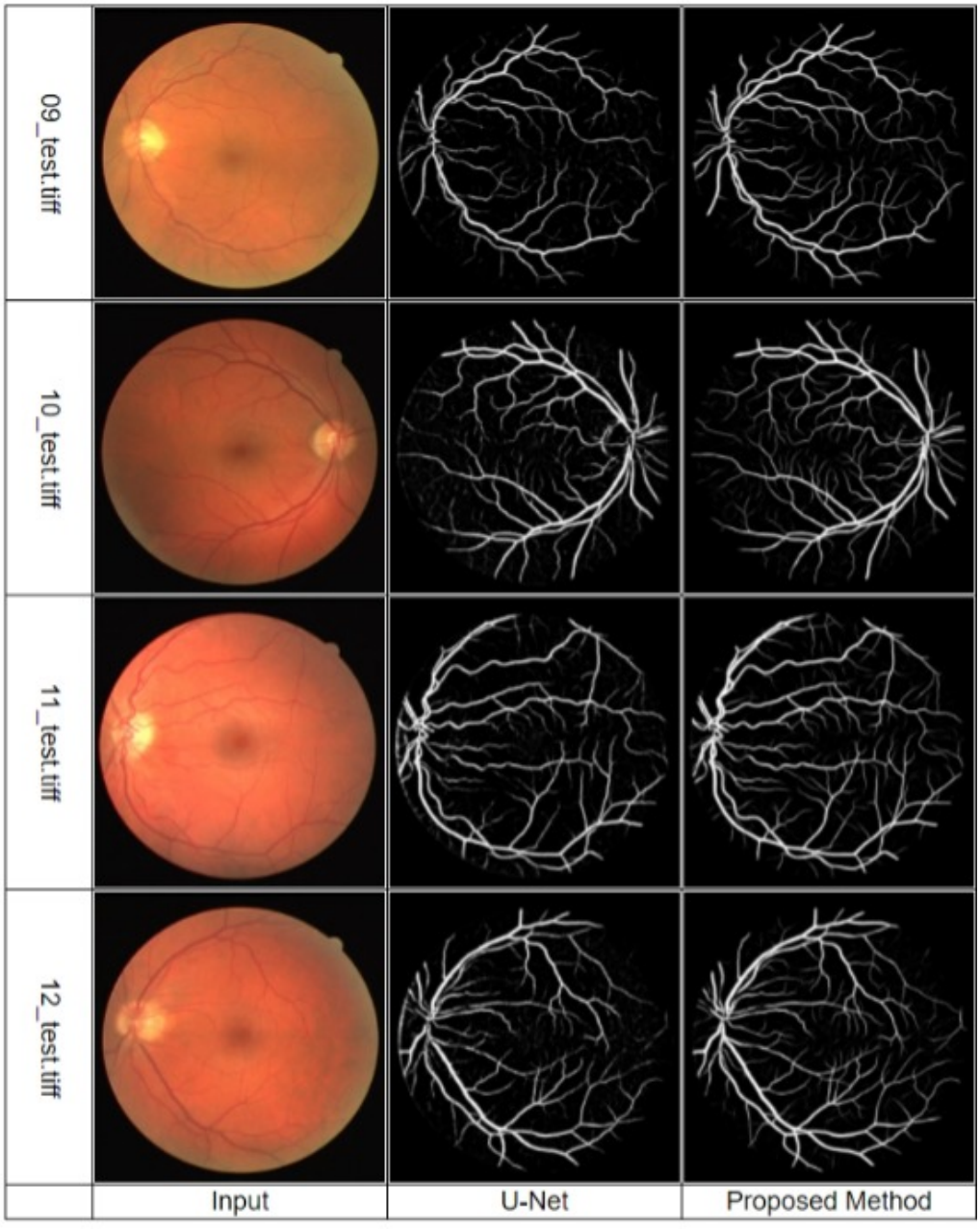


Fig. 2.13 Comparation of segmentation results between baseline and proposed method on DRIVE dataset (Cont.)

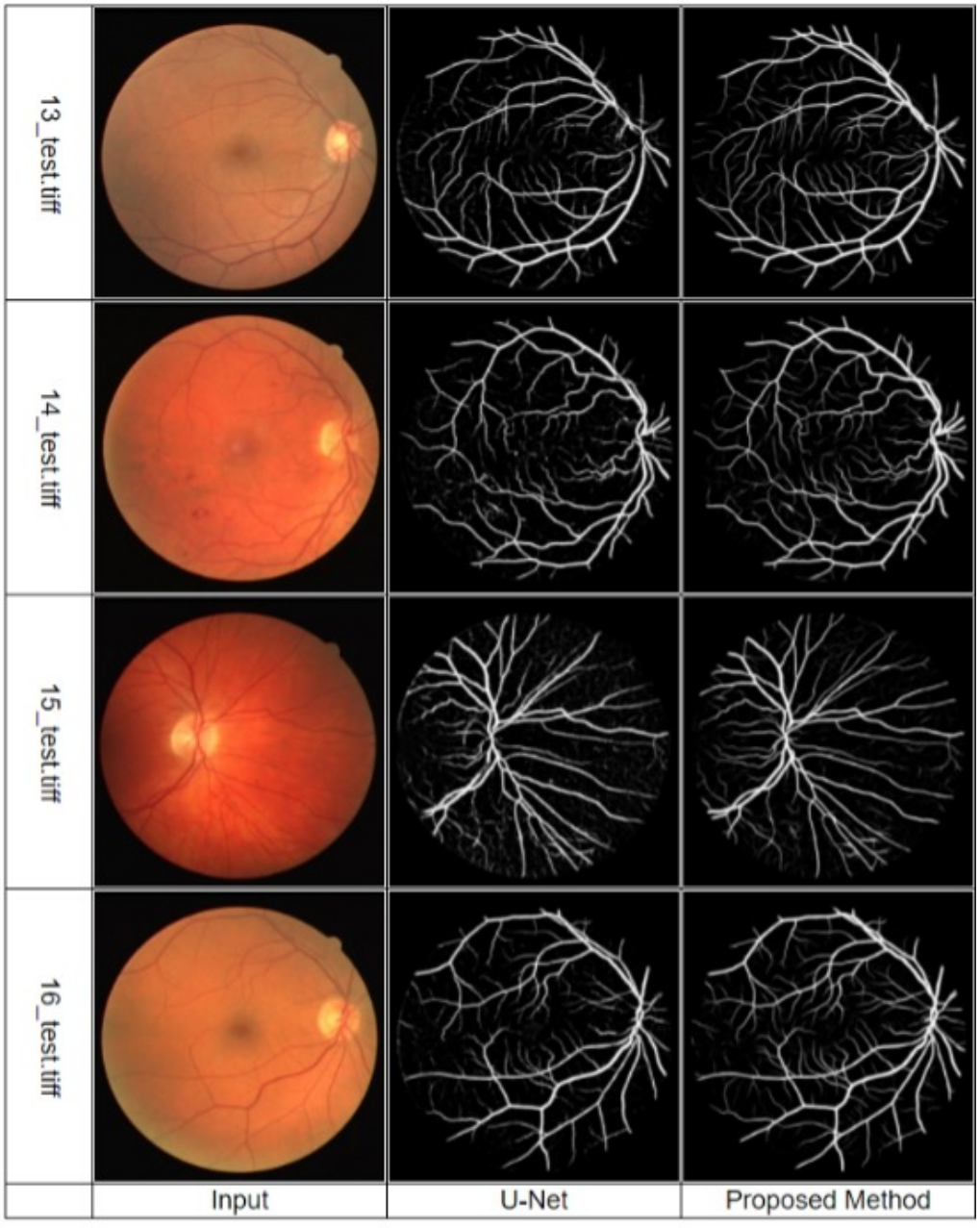


Fig. 2.14 Comparation of segmentation results between baseline and proposed method on DRIVE dataset (Cont.)

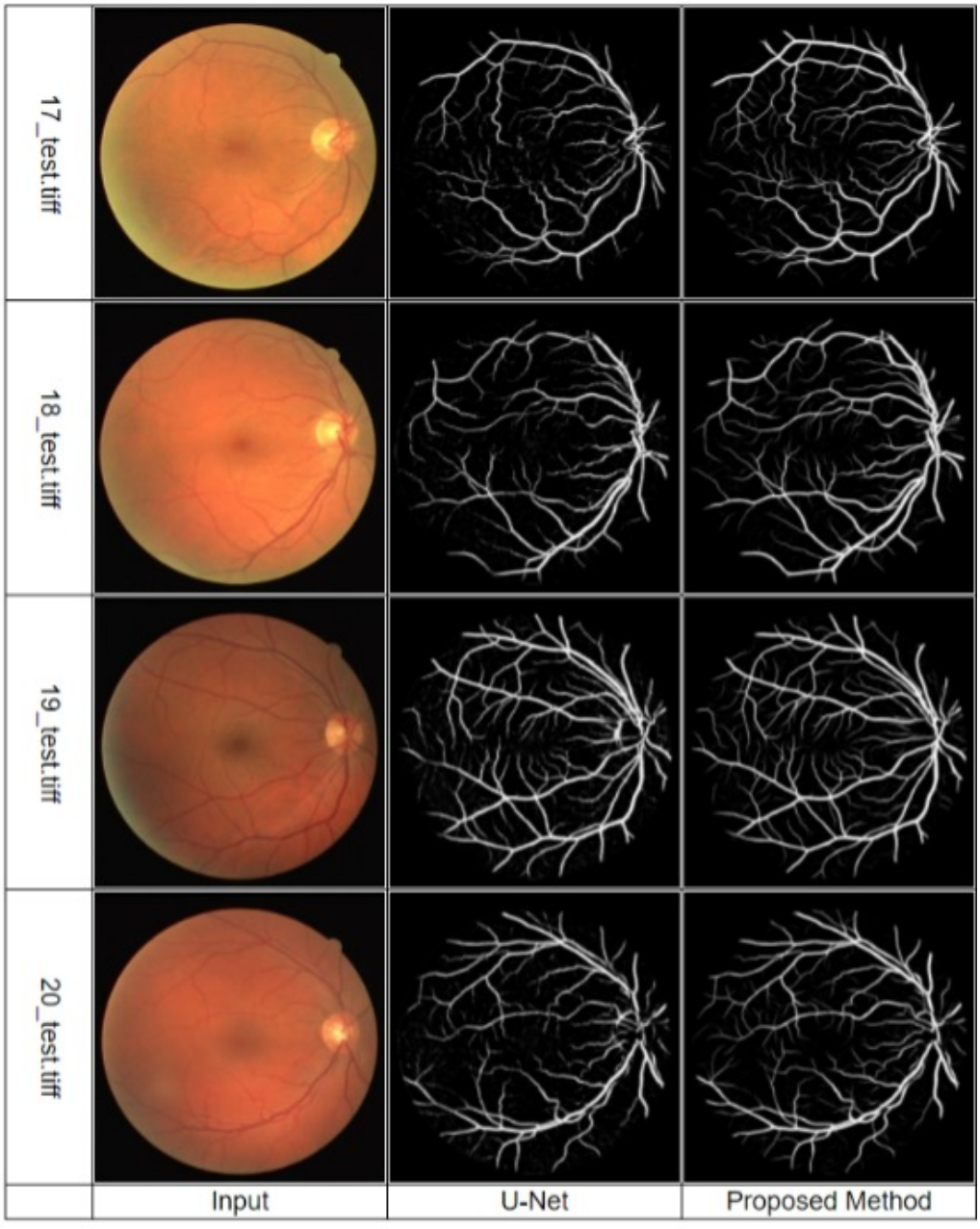


Fig. 2.15 Comparation of segmentation results between baseline and proposed method on DRIVE dataset (Cont.)

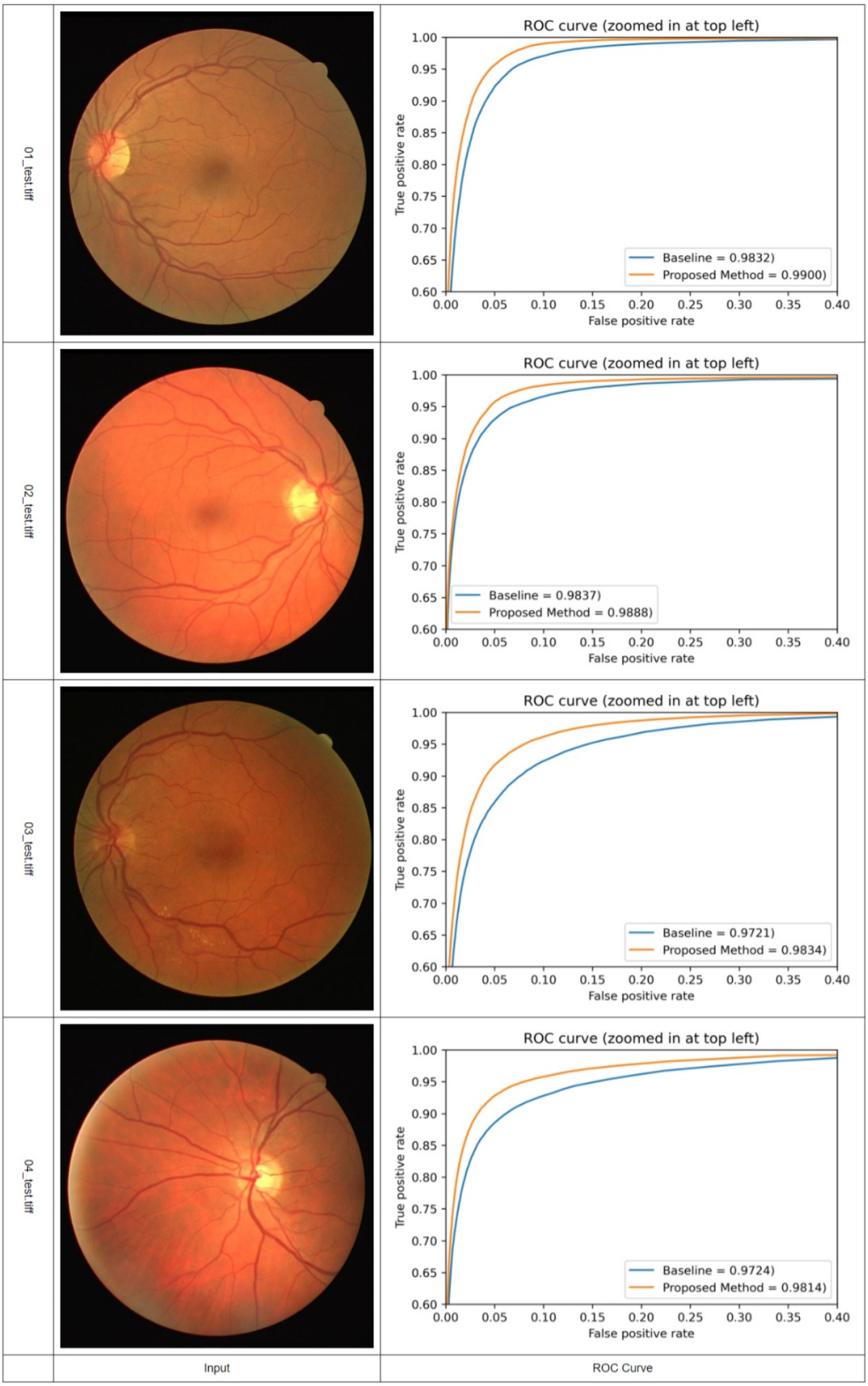


Fig. 2.16 Comparation of ROC Curve between baseline and proposed method on DRIVE dataset (Cont.)

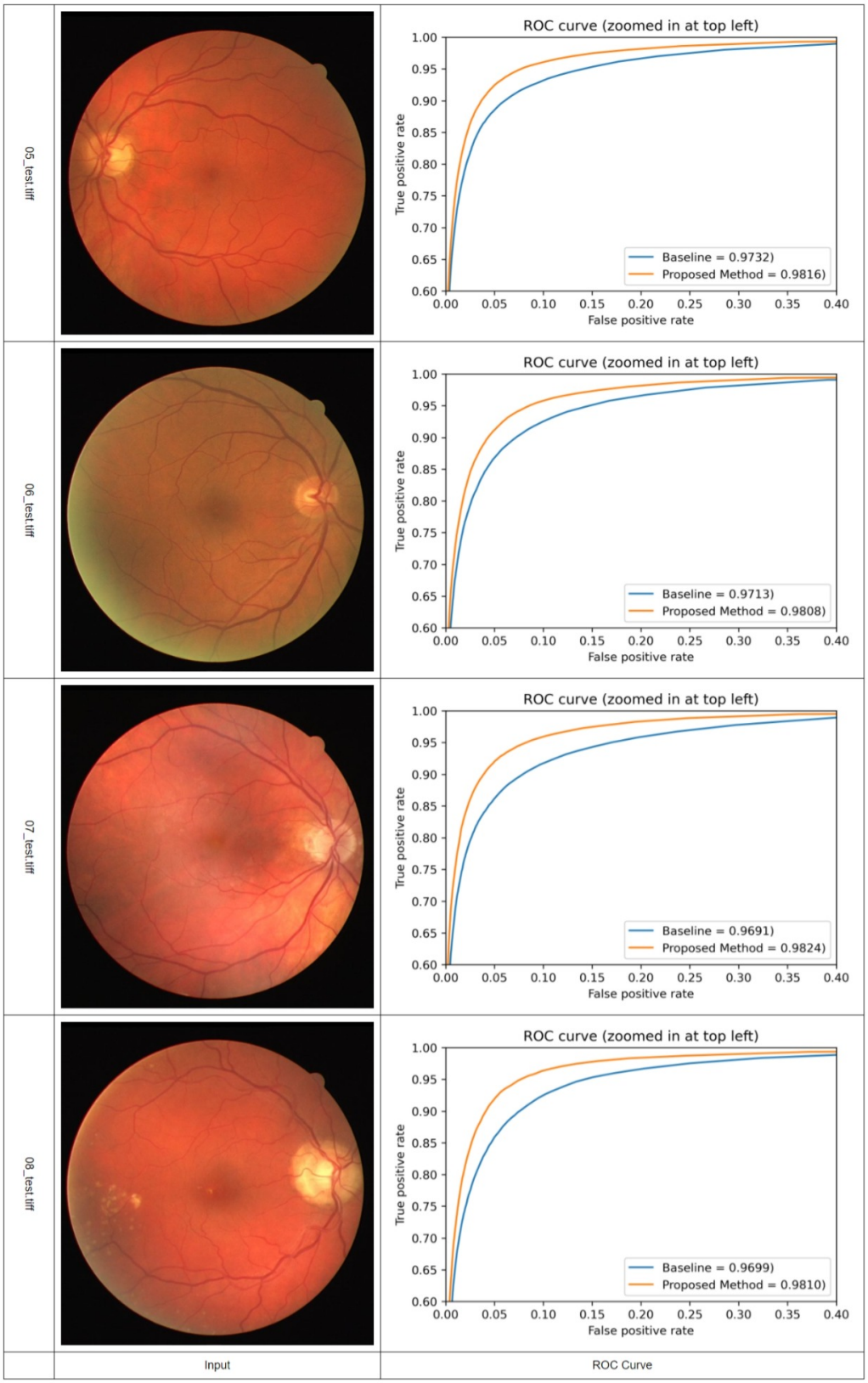


Fig. 2.17 Comparation of ROC Curve between baseline and proposed method on DRIVE dataset (Cont.)

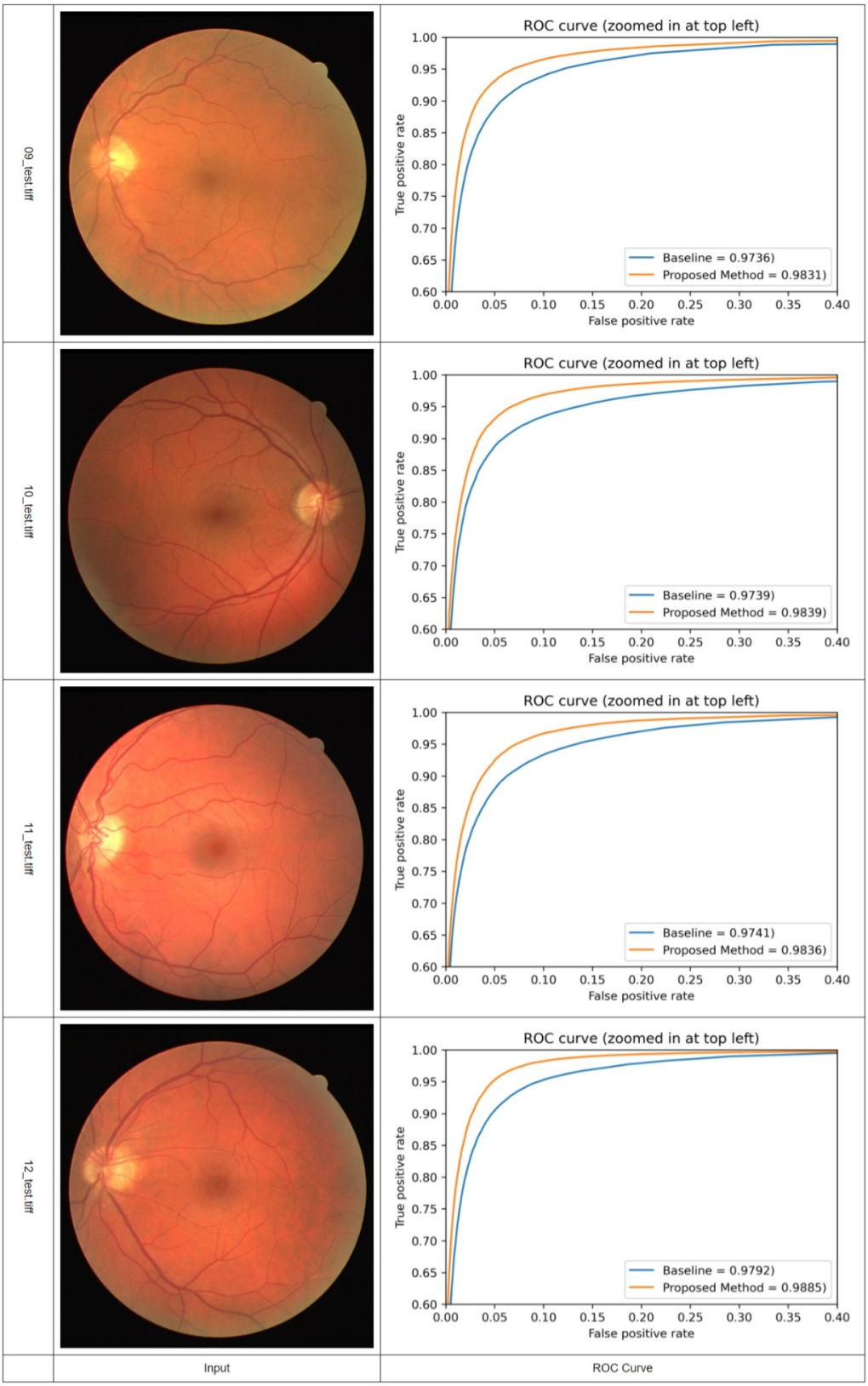


Fig. 2.18 Comparation of ROC Curve between baseline and proposed method on DRIVE dataset (Cont.)

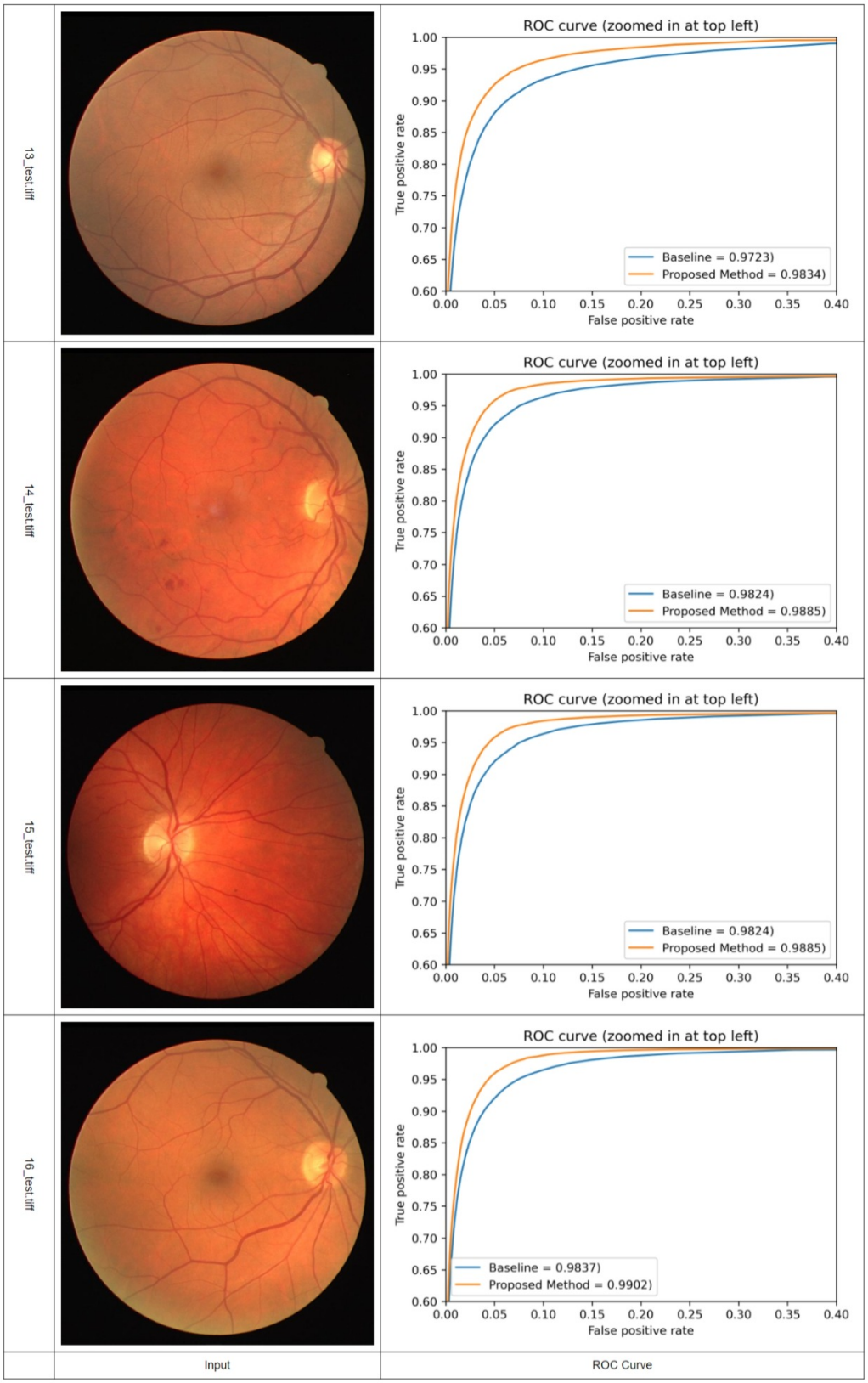


Fig. 2.19 Comparation of ROC Curve between baseline and proposed method on DRIVE dataset (Cont.)

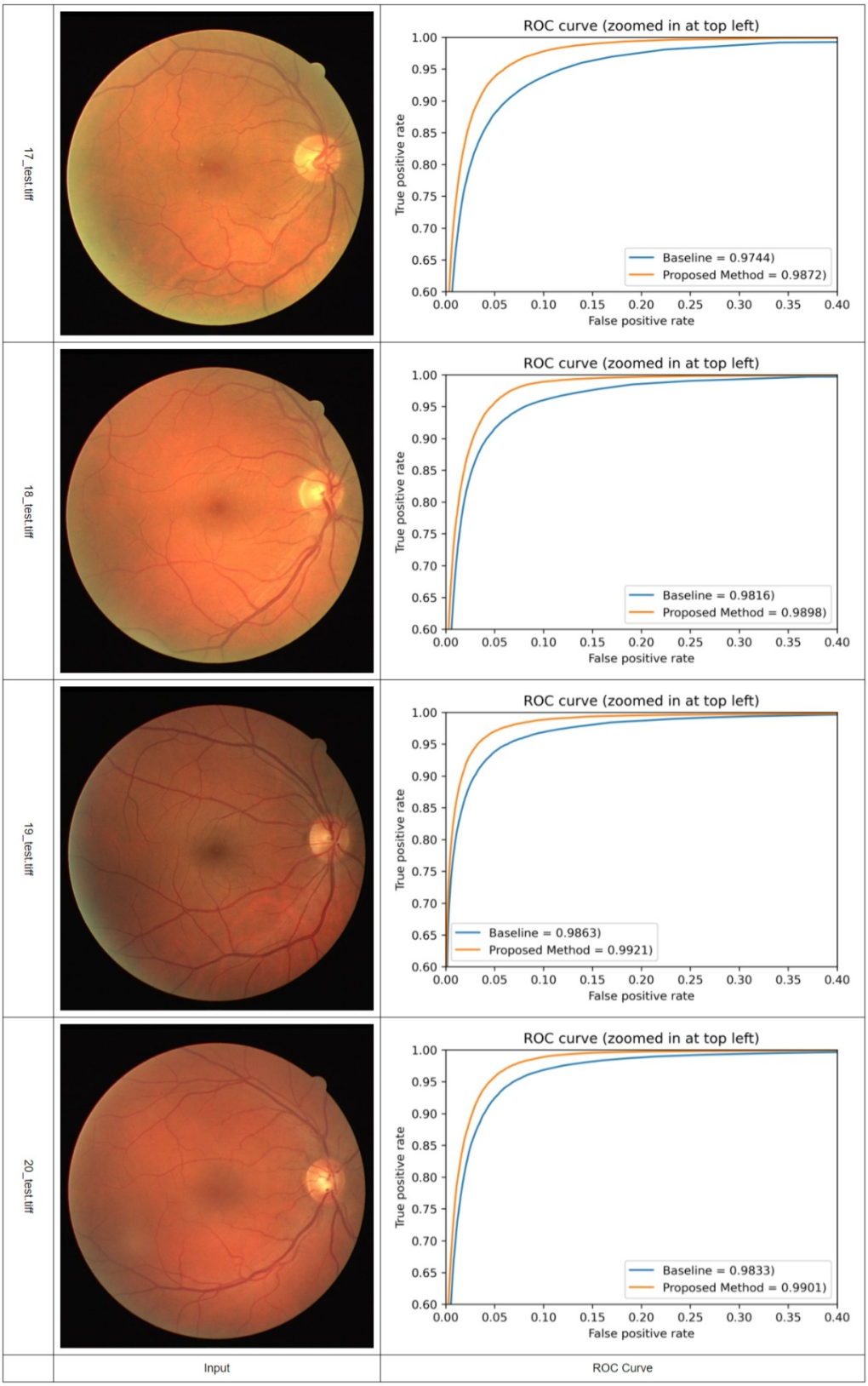


Fig. 2.20 Comparation of ROC Curve between baseline and proposed method on DRIVE dataset

Chapter 3

Graph Laplacian Regularizer based on Prediction and Groundtruth Images

This section presenting our work on a Graph Laplacian Regularizer based on Pixel Differences (GLRDN). Generally, our proposed approach is shown in the Fig. 3.1. In this work, we construct the graph laplacian deriving by two images: from prediction and ground-truth, which the adjacency matrix defined by differences of the neighboring pixel instead of the similarity between pixels.

3.1 Graph Laplacian based On Differences of Neighboring Pixels

Let $G = (V, E)$ be a pixel adjacency graph where $V = \{i | i = 1, \ldots, N\}$ is the set of the pixel indices which represented as nodes with N pixels and the $E = \{(i, j) | i, j \in V\}$ is the neighboring relations between the pixels which representation of edges. Then the adjacency matrix of this pixel adjacency graph is

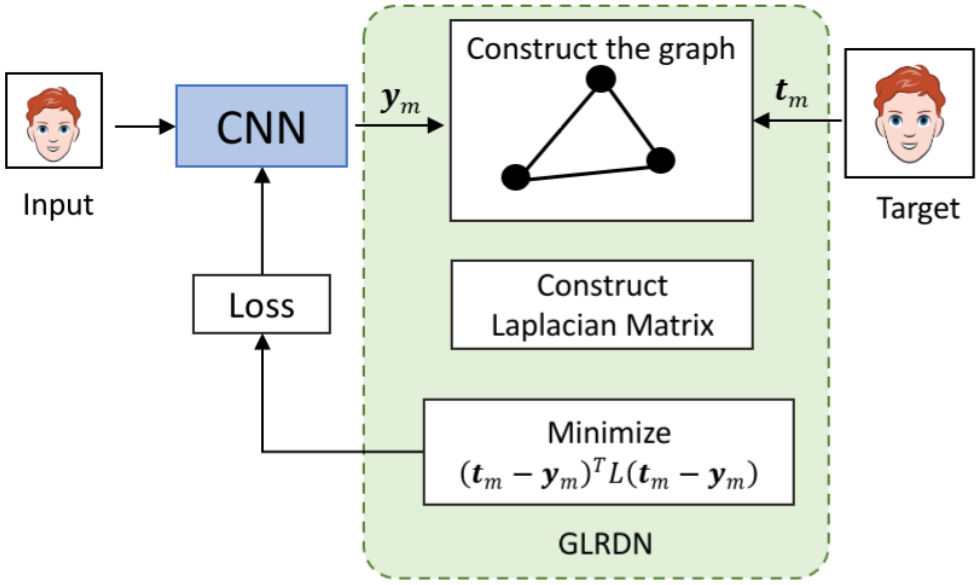


Fig. 3.1 Visualization of our proposed approach, the graphs constructed from prediction image and ground-truth image. Then, the differences metrice between prediction image and ground-truth image is uses as graph Laplacian regularization term.

defined as $A = [a_{ij}]$ where the (i, j) element of the matrix is given as

$$a_{ij} = \begin{cases} 1 & \text{if } (i,j) \in E \\ 0 & \text{otherwise} \end{cases} \tag{3.1}$$

The degree matrix of this graph is defined as the diagonal matrix $D = \text{diag}(d_j = \sum_{i=1}^{N} a_{ij})$. Then we can define the graph Laplacian matrix as $L = D - A$.

Let us consider the set of training samples $X = \{(\boldsymbol{x}_m, \boldsymbol{t}_m) | m = 1, ..., M\}$ where $\boldsymbol{x}_m$ is a m^{th} input image and $\boldsymbol{t}_m$ is the m^{th} target image. The number of training samples is denoted by M.

In deep convolutional neural network, the network is trained to predict the output image $\boldsymbol{y}_m$ from the m^{th} input image $\boldsymbol{x}_m$.

To define the GLRDN, we consider to utilize the differences of the differences of neighboring pixels between the target image $\boldsymbol{t}_m$ and the estimated

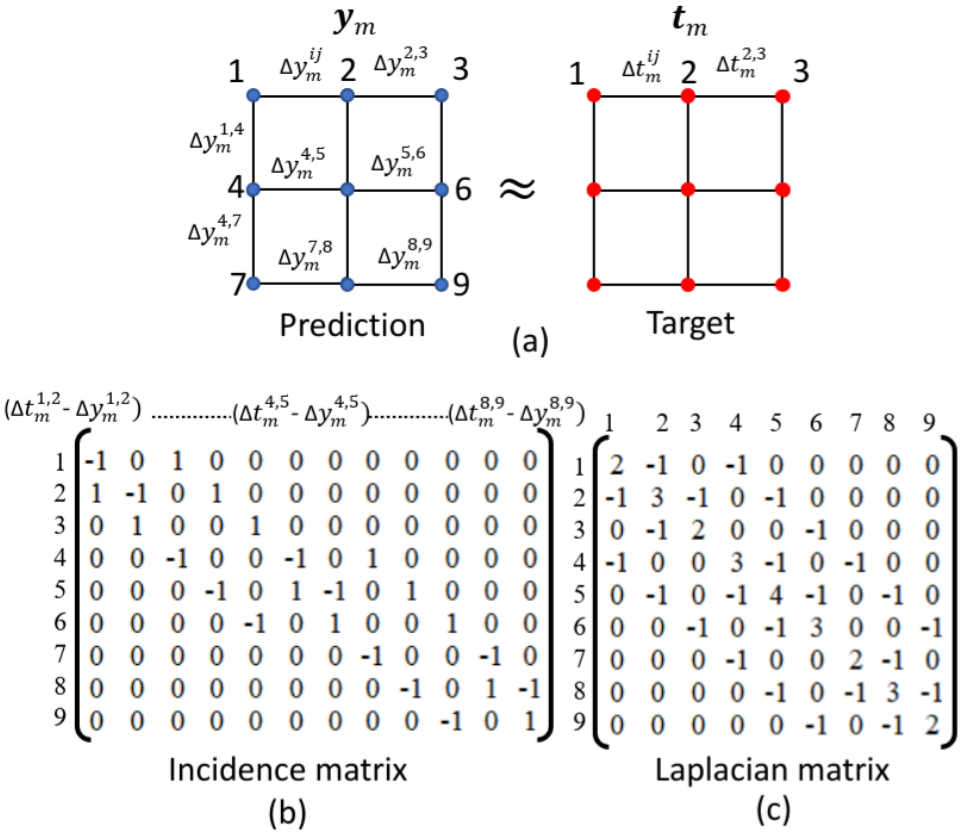


Fig. 3.2 Differences of neighboring pixels calculation. (a) The basic idea to measure the differences between pixels. (b) Incident matrix calculations example from (a). (c) Laplacian matrix calculated from incident matrix.

images $\boldsymbol{y}_m$ as

$$
\begin{aligned}
S_G(\boldsymbol{t}_m, \boldsymbol{y}_m) &= \sum_{(i,j)\in E} \{(t_{mi} - t_{mj}) - (y_{mi} - y_{mj})\}^2 \\
&= \sum_{(i,j)\in E} (\Delta t_m^{ij} - \Delta y_m^{ij})^2 \\
&= (\Delta \boldsymbol{t}_m - \Delta \boldsymbol{y}_m)^T (\Delta \boldsymbol{t}_m - \Delta \boldsymbol{y}_m) \\
&= (B\boldsymbol{t}_m - B\boldsymbol{y}_m)^T (B\boldsymbol{t}_m - B\boldsymbol{y}_m) \\
&= (\boldsymbol{t}_m - \boldsymbol{y}_m)^T B^T B (\boldsymbol{t}_m - \boldsymbol{y}_m) \\
&= (\boldsymbol{t}_m - \boldsymbol{y}_m)^T L (\boldsymbol{t}_m - \boldsymbol{y}_m)
\end{aligned} \tag{3.2}
$$

where B is the incident matrix. It is noticed that $B^T B$ is equal to the graph Laplacian matrix L. For example, the incident matrix B and the graph Laplacian matrix L for the simple graph in the Fig. 3.2(a) are given as Fig. 3.2(b-c). Graph Laplacian matrix assumes that the differences of neighbor pixel between target images and estimated images, denoted as $(\boldsymbol{t}_m - \boldsymbol{y}_m)$ is smooth with respect to concerning corresponding graph G. In particular, it enforce the value of $(\boldsymbol{t}_m - \boldsymbol{y}_m)^T L(\boldsymbol{t}_m - \boldsymbol{y}_m)$ should be small.

For M training samples, we can define the graph Laplacian regularization term as

$$S_G = \sum_{m=1}^{M} S_G(\boldsymbol{t}_m, \boldsymbol{y}_m) = \sum_{m=1}^{M} (\boldsymbol{t}_m - \boldsymbol{y}_m)^T L(\boldsymbol{t}_m - \boldsymbol{y}_m) \tag{3.3}$$

This measure S_G becomes small if the estimated output images are similar with the target images.

3.2 The GLRDN with objective function

The neighboring pixel relationship is a fundamental information in an image. Any task that involves images in the learning process, especially computer vision, utilizing neighboring pixel information provides an advantage over using only pixel-to-pixel information by adding the GLRDN as regularization term into pixel-wise loss function. In this section, we explain how to utilize the GLRDN combining with pixel-wise loss function such as MSE and BCE. MSE is common used in image super-resolution tasks. The MSE is given by

$$E_{mse} = \frac{1}{M} \sum_{m=1}^{M} (\boldsymbol{t}_m - \boldsymbol{y}_m)^2 \tag{3.4}$$

Table 3.1 Comparison of architectural performance without and with our proposed GLRDN on the super image resolution task. Performance was measured using PSNR and SSIM on Cartoon Set and Manga109 datasets.

Dataset	**Method**	**Scale**	**PSNR**	**SSIM**
Cartoon Set	EDSR	x2	25.02	0.8743
	EDSR+GLRDN	x2	**30.51**	**0.9604**
Manga109	EDSR	x2	20.54	0.7791
	EDSR+GLRDN	x2	**21.89**	**0.8263**

For the training of the parameters of the network, we combine the MSE loss with the regularization term as

$$Q_{sr} = E_{mse} + \lambda S_G \tag{3.5}$$

where λ is a scaling parameter to control the effect of the regularization.

In the image segmentation task, the binary cross-entropy (BCE) loss is used as the objective function. The BCE loss is given by

$$E_{bce} = -\sum_{m=1}^{M}\sum_{i=1}^{N}\{t_{mi}\log(y_{mi}) + (1-t_{mi})\log(1-y_{mi})\} \tag{3.6}$$

Similarly, we combine the objective function with the proposed regularization term for M training samples as

$$Q_{is} = E_{bce} + \lambda S_G \tag{3.7}$$

where λ is a scaling parameter to control the effect of the regularization.

We evaluate the improvement of our method using the measurements we have mentioned in the section 2.4. In addition, we measure additional experiments in images super-resolution task using PSNR and SSIM. PSNR (Peak Signal to Noise Ratio) is the ratio calculated between the highest potential

Table 3.2 Comparison of architectural performance without and with our proposed GLRDN on the image segmentation task. Performance is measured using Sn, Sp, Acc, AUC, Jaccard, F-1 Score and Area Under Precision-Recall (PRC) on DRIVE dataset.

Model	**Sn**	**Sp**	**Acc**	**AUC**	**Jaccard**	**F1-Score**	**PRC**
bce	0.7583	0.9826	0.9551	0.9691	0.6742	0.8054	0.8893
dice	0.7440	0.9856	0.9560	0.9665	0.6753	0.8061	0.8840
bce+glrdn	0.8092	0.9834	0.9621	0.9826	0.7235	0.8396	0.9239
dice+glrdn	0.8271	0.9806	0.9618	0.9797	0.7263	0.8414	0.9162

Table 3.3 Comparison of architectural performance without and with our proposed GLRDN on STARE dataset.

Model	**Sn**	**Sp**	**Acc**	**AUC**	**Jaccard**	**F1-Score**	**PRC**
bce	0.7625	0.9624	0.9326	0.9497	0.6283	0.7717	0.8576
bce+glrdn	0.8040	0.9580	0.9350	0.9561	0.6491	0.7872	0.8757

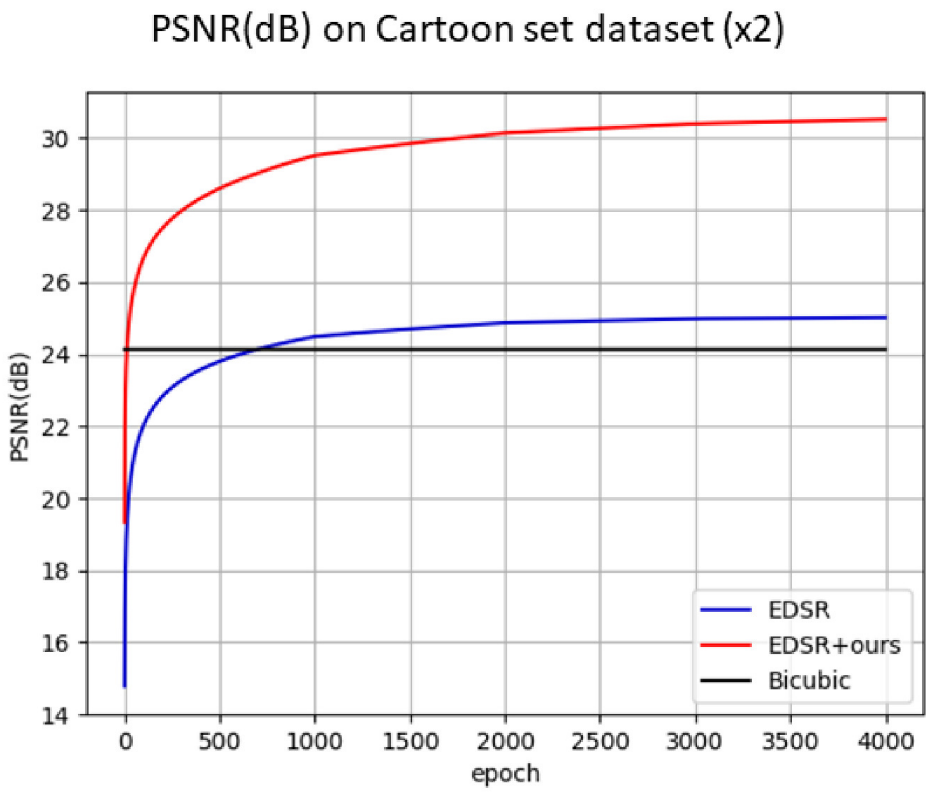


Fig. 3.3 PSNR comparison of our proposed method on Cartoon set dataset.

signal and distorting noise that influences its representation accuracy. Given a

Table 3.4 Comparison of the proposed approach performance on image super-resolution task with state of the art methods.

Methods	Scale	Cartoon set		Manga109	
		PSNR	SSIM	PSNR	SSIM
bicubic	x2	24.13	0.8972	19.95	0.7887
SRCNN	x2	24.15	0.8936	20.07	0.7811
WDSR	x2	24.86	0.8760	20.31	0.7758
EDSR	x2	25.02	0.8743	20.54	0.7791
EDSR+ours	x2	**30.51**	**0.9604**	**21.89**	**0.8263**
bicubic	x3	20.06	0.7531	16.47	0.6120
SRCNN	x3	20.41	0.7543	16.79	0.6085
WDSR	x3	21.50	0.7725	17.05	0.6188
EDSR	x3	21.74	0.7737	17.15	0.6244
EDSR+ours	x3	**25.96**	**0.8983**	**17.88**	**0.6749**
bicubic	x4	17.95	0.6398	14.70	0.4915
SRCNN	x4	18.47	0.6491	15.25	0.4879
WDSR	x4	19.45	0.6880	15.46	0.5056
EDSR	x4	19.68	0.6793	15.53	0.5074
EDSR+ours	x4	**22.66**	**0.8061**	**15.85**	**0.5397**

Table 3.5 Comparison of the proposed approach performance on image segmentation with state-of-the-art methods.

Methods	Sp	Sn	Acc	AUC
Orlando[55]	0.9684	0.7897	0.9454	0.9506
Hu[20]	0.9793	0.7772	0.9533	0.9759
Khan[28]	0.9670	0.7373	0.9501	-
Fraz[15]	0.9807	0.7406	0.9480	0.9747
Zhou[83]	0.9803	0.7262	0.9475	-
Yan[77]	0.982	0.7631	0.9538	0.975
Chen[8]	0.9735	0.7426	0.9453	0.9516
Zhang[80]	0.9712	0.7861	0.9466	0.9703
Wang[72]	0.9736	0.7986	0.9511	0.9740
Strisciuglio[68]	0.9702	0.7777	0.9454	0.9597
Proposed method	**0.9834**	**0.8092**	**0.9621**	**0.9826**

ground-truth image and a predicted image, the PSNR is defined by:

$$PSNR = 10\log_{10}(255^2/MSE) \tag{3.8}$$

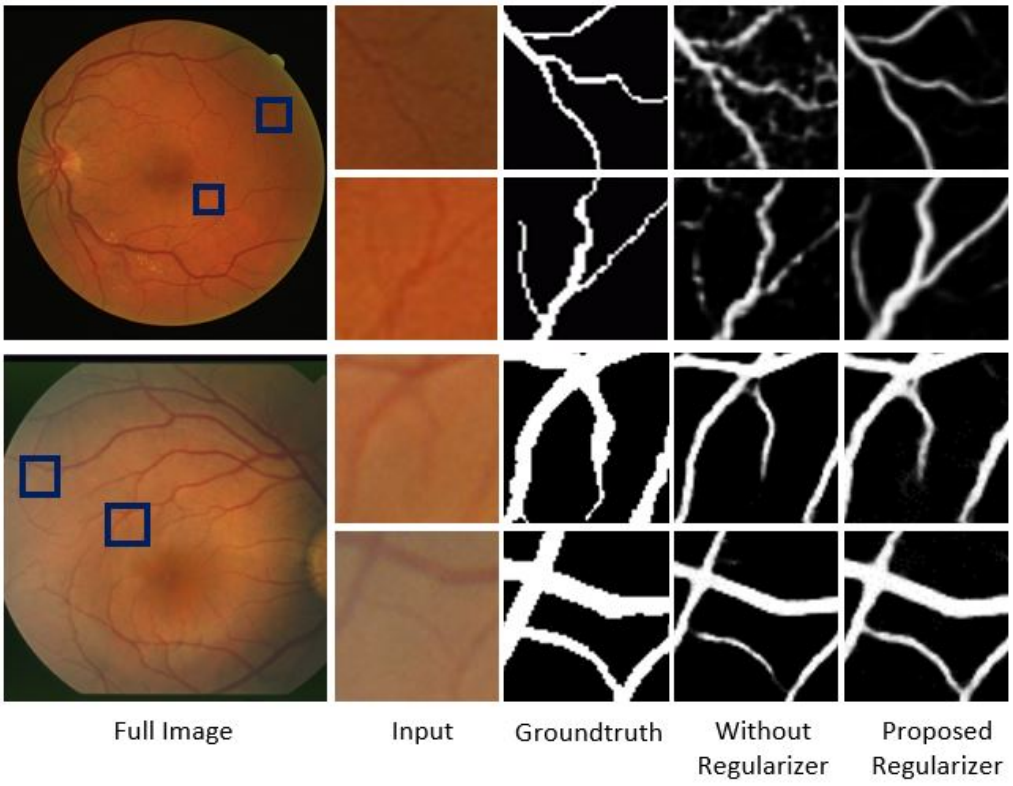


Fig. 3.4 Comparison results of blood vessel segmentation between our proposed regularizer and without proposed regularizer on the DRIVE (top row) and STARE (bottom row) datasets.

SSIM(Structure Similarity Index Method) is used for measuring the similarity between two images. The SSIM is built by modeling image distortion as a combination of three factors: lack of correlation, luminance distortion, and contrast distortion. The SSIM is defined as :

$$SSIM = l(x,t)c(x,t)s(x,t) \tag{3.9}$$

where l, c, and t are luminance distortion, contrast distortion, and lack of correlation, respectively.

3.3 Results and Discussion

The proposed network with the GLRDN succeeds in reconstructing the detail of boundary structure compared over network without the GLRDN is presented in this section. On Manga109 dataset, the CNN architecture without the GLRDN compared to the original HR image, the line portion of the bicubic and EDSR results appear jagged and not smooth. Coupling EDSR with GLRDN, the boundary structure is better and the line is more smooth close to the HR image. On Cartoon set, the architecture without GLRDN result seen not clear. For example, the color is inconsistent in the face skin region that has the same color. Using proposed network with GLRDN, the result showing the face skin color has more consistent and clearly. Based on the results, the proposed approach offers the advantage of smoother line sections, and more consistent sections of the same color.

Some selected qualitative results on the DRIVE and STARE dataset are shown in Fig. 3.4. Segmentation results of proposed network with GLRDN have more accurate connectivity and detail boundary than results without GLRDN. This demonstrates that GLRDN can definitely capture the relationship between pixels in an image by using differences between pixels. Thus, the prediction of the model with GLRDN has better visual qualities. It is because GLRDN considered the differences between neighboring pixels, which means the pixels belonging same classes will force the pixel pairs to make their predictions more consistent. The neighbor pixels which belong to different classes will separate force on them to make their predictions more inconsistent.

Table 3.1 and Table 3.2 presents the quantitative performance comparison of our proposed approach in image super-resolution and image segmentation task, respectively. The proposed network with GLRDN performed well in capturing the relationship of neighboring pixels better than network without GLRDN.

With EDSR as baseline models, GLRDN can improve the PSNR from 25.02 to 30.51 on the Cartoon set dataset and 20.54 to 21.89 on the Manga109 dataset with upscaling factor of 2. Fig. 3.3 plotting an example of PSNR visualization on a Cartoon set dataset. In the SSIM measurement, GLDRN improved the result from 0.8743 to 0.9604 on Cartoon set, and 0.7791 to 0.8263 on Manga109 dataset.

Furthermore, Table 3.2 showing the performance results on DRIVE and STARE datasets, The GLRDN can improve the AUC score by 0.62 and 0.67 percent on the DRIVE and STARE datasets respectively. We can also see that the reproduced models with GLRDN showing a competitive performance of sensitivity by 6.5 and 5.4 percent on the DRIVE and STARE datasets, respectively. The results demonstrate the effectiveness of graph laplacian regularizer based on differences of neighbor pixels.

We also compare the results of the proposed approach with the results of state-of-the-art methods. For fair comparison, we separate the comparisons according to the task. Table 3.4 presents proposed network with GLRDN performance compare with state-of-the-art in image super-resolution tasks which is higher than the other existing methods on all upscaling factor. On upscaling factor x2, our proposed approach achieved 30.51 of PSNR and 0.9604 of SSIM on Cartoon set and 21.89 of PSNR and 0.8263 of SSIM. On upscaling factor x3, our proposed regularizer achieved 25.96 of PSNR and 0.8983 of SSIM on Cartoon set and 17.88 of PSNR and 0.6749 of SSIM. On upscaling factor x4, our proposed regularizer achieved 22.66 of PSNR and 0.8061 of SSIM on Cartoon set and 15.85 of PSNR and 0.5397 of SSIM.

Table 3.5 presents comparison of the proposed method performance with state-of-the-art methods on DRIVE datasets. The experiment results of all methods are taken from their published papers. The proposed method got an AUC of 0.9740 and an accuracy of 0.9561, which is nearly identical to the state-

of-the-art method. Although our approach is qualitatively superior to other methods, the ground truth-based quantitative calculations show the same accuracy and AUC value as current methods. This is because the ground truth is that there is no annotation of the small vein, which makes the calculation inaccurate, and the small vessel is considered a false positive. Good detection of small blood vessels contributes to good sensitivity. We can still get very high Sn at a very low Sp price by adjusting the decision thresholds, or vice versa. As a result, AUC is the most important performance predictor of the three, while Sn and Sp are primarily used for detail.

Although our approach proves its ability to enhance the boundary structure on overall datasets, the GLRDN has limitations that are calculated by treating pixel by pixel, which is a computational time cost. We plan to extend GLRDN by using a sparse graph approach to reduce computational costs in future work.

Chapter 4

Regularizer based on Euler Characteristic

This study proposed EC-based regularizers to estimate the number of isolated objects in U-Net-like deep CNN architecture for delineating small retinal vessel connections on a fundus image.

4.1 Constructing Euler characteristics for the number of Isolated Objects

Euler characteristic (EC) is a global topology, which invariant to all topological transformations such as rotation and scale. Generally, the EC of a two dimensional image is considered to be the number of connected component minus the number of holes. The objective function of the EC is computed based on the relationship between the number of vertices (P), sides (S), and faces (F) :

$$EC = F - S + P \tag{4.1}$$

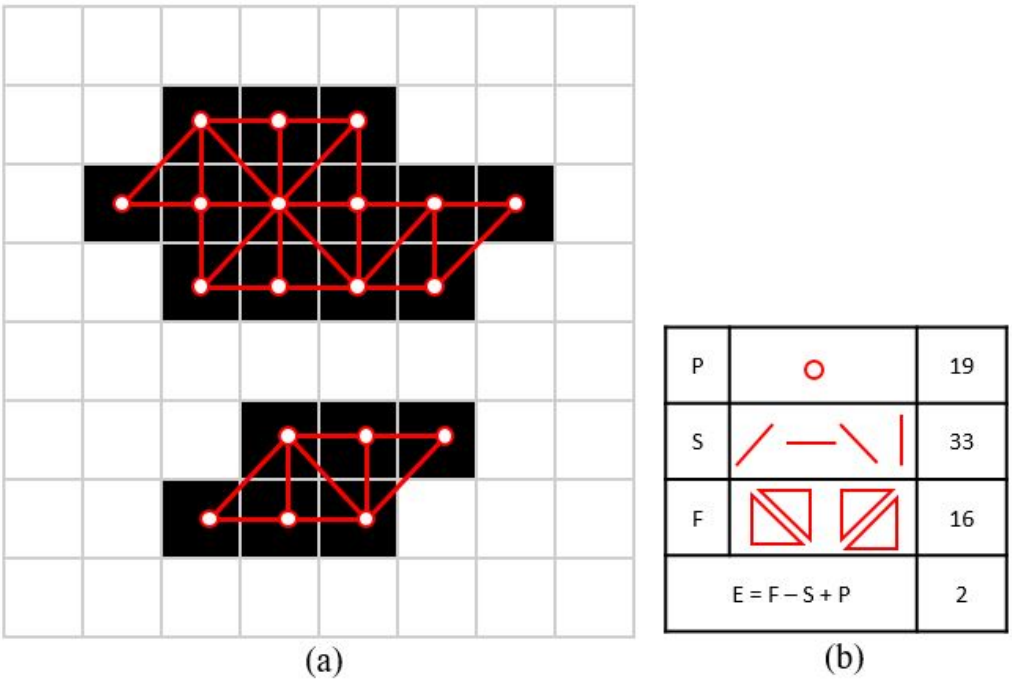


Fig. 4.1 An example of constructing Euler characteristics using mask patterns (a) illustration showing object (black) pixels based on 8x8 neighborhood pixels. (b) illustration calculating Euler characteristics for the number of isolated object according to the number of vertices (P), sides (S), and faces (F) from object pixels (a).

However, in the case of simple polygon without holes, the EC is equal to one. Here we briefly explained how to define the EC effectively by constructing vertices, sides, and faces on 8x8 neighbourhood pixels connectivity using simple mask patterns, as shown in figure 4.1. It calculates EC directly related to the number of isolated objects. The EC on the binary image based on 8x8 neighbourhood pixels can be constructed as follows; 1) transforming object pixel in the image to a vertice, 2) adding a side between the vertices with 8-connectivity without a cross side, 3) summing the number of vertices, 4) summing the number of sides, 5) summing the number of faces or triangles, and finally, 6) calculating EC using eq 4.1. As can be seen in figure 4.1, EC estimated the number of isolated objects as 2, with number of vertices, sides, and faces as 19, 33, and 16 respectively ($EC = F - S + P = 16 - 33 + 19 = 2$).

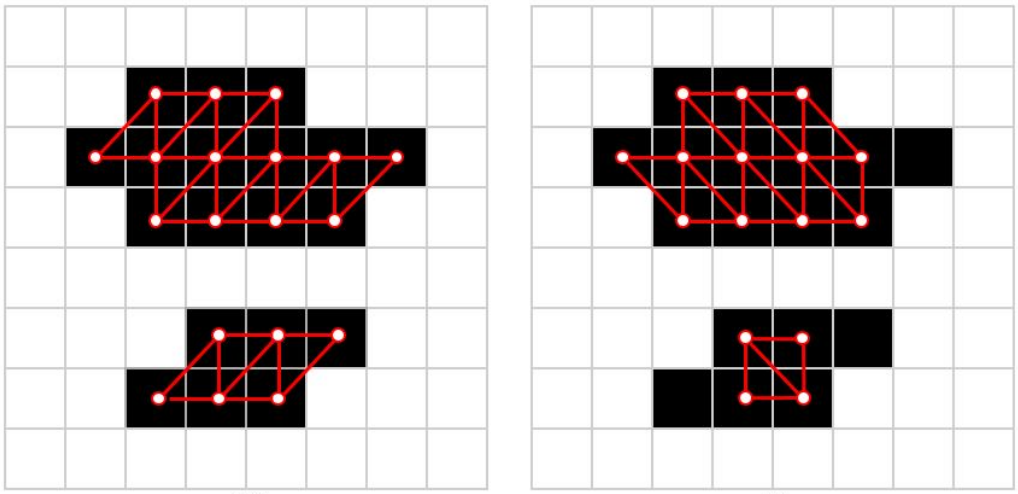


Fig. 4.2 Two ways of constructing Euler characteristics using mask patterns for triangulation. (a) and (b) show examples of triangulation in opposite direction with each other.

Furthermore, to address the invariant problem in preserving the vessel connections, we considered constructing vertices, sides, and faces in two different directions. It can be briefly explained using mask patterns on a 8x8 neighbourhood pixels connectivity, as shown in figures 4.2a and 4.2b . Then EC is estimated as $EC_1 = F_1 - S_1 + P_1$ and $EC_2 = F_2 - S_2 + P_2$ from 4.2a and 4.2b, respectively, which is directly related to the number of isolated objects. Finally, the total number of isolated objects E is considered by computing the average between these two directions, which is defined as

$$E = \frac{EC_1 + EC_2}{2} \tag{4.2}$$

4.2 Euler characteristic as Regularizer

The regularizer layer based on EC for small vessel connections is evaluated using U-Net-like deep CNN structure. The proposed architecture contains an encoder and decoder module, as shown on figure 4.3.

By following [58, 33, 31], we calculated EC and incorporated this into the network using one layer. Let us consider input image as $X \in \mathbb{R}^{M.N}$, and $T \in \{0,1\}^{M.N}$ be the corresponding groundtruth, with 1 indicating pixels in the vessel and 0 is indicating pixels in the background area. And let us consider f be a U-Net parameterized by weight W. Then the output image of the network is $Y = f(X,W) \in [0,1]^{M.N}$. The binary cross entropy (BCE) loss is used for calculating the vessel region segmentation, which is defied as

$$L_{BCE} = \sum_{i=1} t_i log(y_i) + (1-t_i) log(1-y_i) \tag{4.3}$$

Although U-Net predicts the vessel region, the BCE loss treats every pixel independently. Therefore, it fails to estimate the topological characteristics such as the number of isolated objects on a vessel region. It can be clearly observed as the misclassification of small vessel regions in figure 1.1. This could be accounted to the fact that some pixels exhibit low costs in terms of BCE loss, and thus have large impact on the topology of the predicted results. To address this problem, we proposed a regularizer based on EC that penalized BCE if have many isolated objects comprises the target task.

Therefore, we incorporated EC into the network that retrieved the number of isolated object (E) through the number of vertices, sides, and faces of the segmented regions using the eqs 4.1 and 4.2. We used E as a regularizer term with the cross entropy cost function to train the proposed network for precisely delineating the small vessel connections. It forced the network to minimize the number of isolated objects by minimizing the misprediction error there are large number of isolated objects, and thus we named it the $MISO$ regularizer. It is defined as

$$L_{MISO} = L_{BCE} + \alpha E_{OUT} \tag{4.4}$$

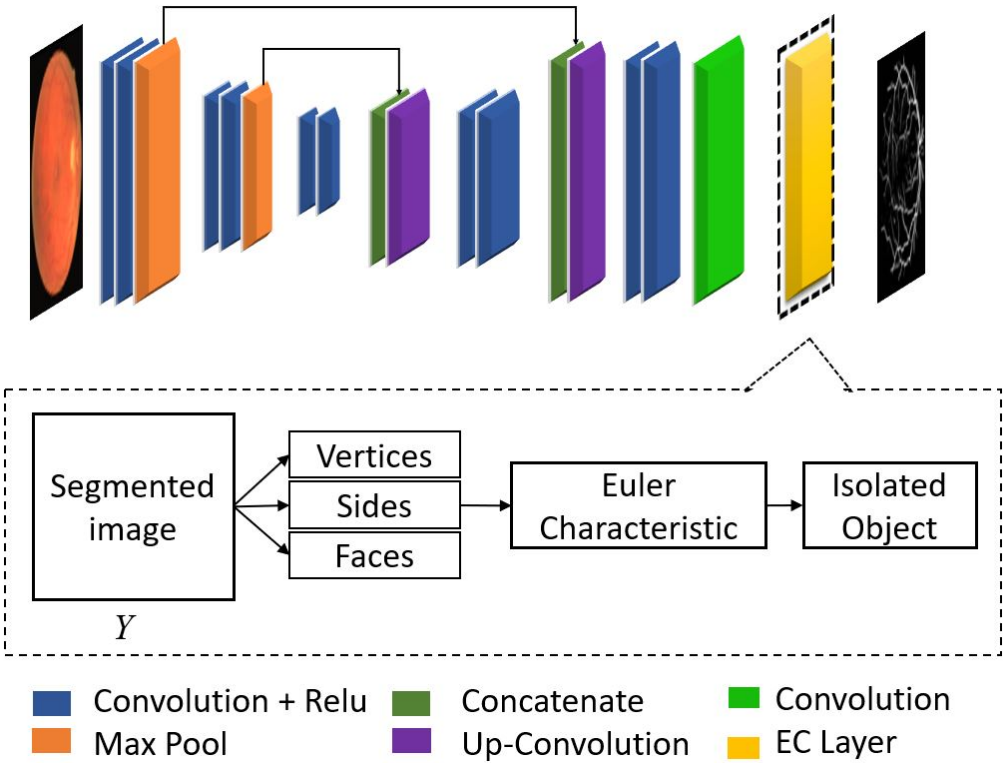


Fig. 4.3 Proposed architecture using Euler characteristics regularizer

where α is a scaling parameter to control the regularizer and E_{OUT} represents the number of isolated objects in the segmentation output.

4.3 Regularization based on the Differences of number of Isolated Objects between Prediction and Groundtruth

In this study, we also investigated the regularizer based on the differences of number of isolated objects between output and groundtruth (DISO) in delineating the vessel regions. Different from MISO, it forces the number of isolated objects between the output and the groundtruth to be equal or more closer. It is defined as

$$L_{DISO} = L_{BCE} + \alpha \left| E_{OUT} - E_{LABEL} \right| \tag{4.5}$$

where E_{OUT}, E_{LABEL} is number of isolated object of the predicted output and groundtruth, respectively. The DISO based object function leads to large misclassification error if the number of isolated objects between the output and groundtruth is not equal. Otherwise, it produces zero misclassification error in detecting the vessel regions.

4.4 Regularization based on Euler Characteristic with Graph based Smoothing

The regularizer based on GS using a graph laplacian matrix is considered the image into two graph laplacians for vessel regions and background area [18]. In GS, smoothness level on vessel regions (S) and background area is constructed and can be written as

$$S = y^T (L_F + L_B) y = y^T L_G y \tag{4.6}$$

where L_F and L_B indicates laplacian graph of vessel and background, respectively. In [18], we defined L_{GS} by incorporating GS with binary cross entropy loss, as indicated in the following equation

$$L_{GS} = L_{BCE} + \beta S \tag{4.7}$$

In this study, we attempted to evaluate the effectiveness of combining our proposed MISO-based EC with GS, thus utilizing the advantages of both the

Table 4.1 Performance comparison of the architecture without and with our proposed Euler characteristic-based regularizer on various datasets

Model	Sn	Sp	Acc	AUC	Jaccard	F1-Score	PRC
DRIVE							
L_{BCE}	0.7583	0.9826	0.9551	0.9691	0.6742	0.8053	0.8892
L_{MISO}	0.8463	0.9759	0.9600	0.9824	0.7218	0.8385	0.9226
L_{DISO}	0.8705	0.9700	0.9578	0.9825	0.7165	0.8349	0.9227
STARE							
L_{BCE}	0.6524	0.9824	0.9331	0.9370	0.5930	0.7445	0.8361
L_{MISO}	0.7401	0.9754	0.9403	0.9546	0.6418	0.7873	0.8715
L_{DISO}	0.8105	0.9680	0.9480	0.9691	0.6648	0.7987	0.8819
CHASEDB1							
L_{BCE}	0.6514	0.9788	0.9490	0.9567	0.5374	0.6991	0.7757
L_{MISO}	0.6080	0.9879	0.9533	0.9681	0.5724	0.7034	0.8167
L_{DISO}	0.7906	0.9770	0.9600	0.9786	0.6426	0.7824	0.8606

number isolated objects as well as smoothness level of vessel regions. The proposed MISO with GS (GISO) is defined as

$$L_{GISO} = L_{BCE} + \alpha E_{OUT} + \beta S \tag{4.8}$$

where α and β are scaling parameters to control the regularizer.

4.5 Results and discussion

The vessel regions with complicated vessel structure shows the intersection of the vessels with large and tiny vessels. It is shown with some representative examples in figure 4.4. We also demonstrated the superiority of our methods over baseline methods on all DRIVE test images dataset in Fig. 4.8-4.12. It demonstrated that the U-Net-like network without specific regularizer showed poor ability in delineating the vessel with complicated structure. However, if incorporated with our proposed regularizer based on EC, it can accurately identify the detailed vessel junctions and tiny vessels. Table 4.1 presents the performance comparison of our proposed regularizers in segmenting blood vessel connec-

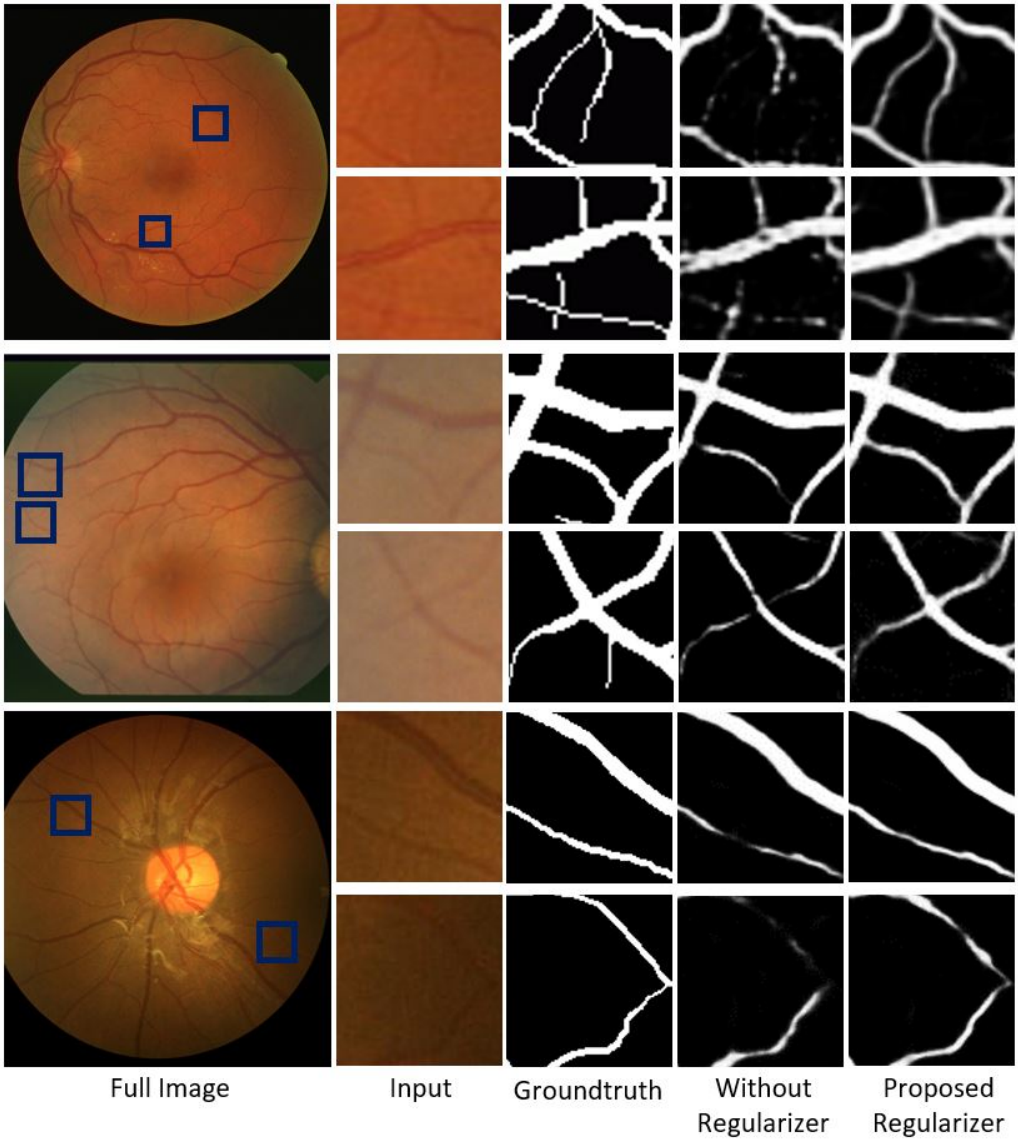


Fig. 4.4 Segmentation results of the architecture without and with our proposed regularizer for blood vessels on DRIVE (top row), STARE (middle row), and CHASEDB1 (bottom row) datasets.

tions, especially small isolated or thin vessel segmentation from background pixels over the architecture without regularizer on different datasets. The proposed network with MISO and DISO regularizer performed well in segmenting the disjoint vessel connections better than the architecture without regularizer. The CNN architecture without our proposed regularizer technique misses a lot of small or thin vessels, which can also be observed in the qualitative results on all dataset. The incapability of the classical BCE-based architecture in detecting

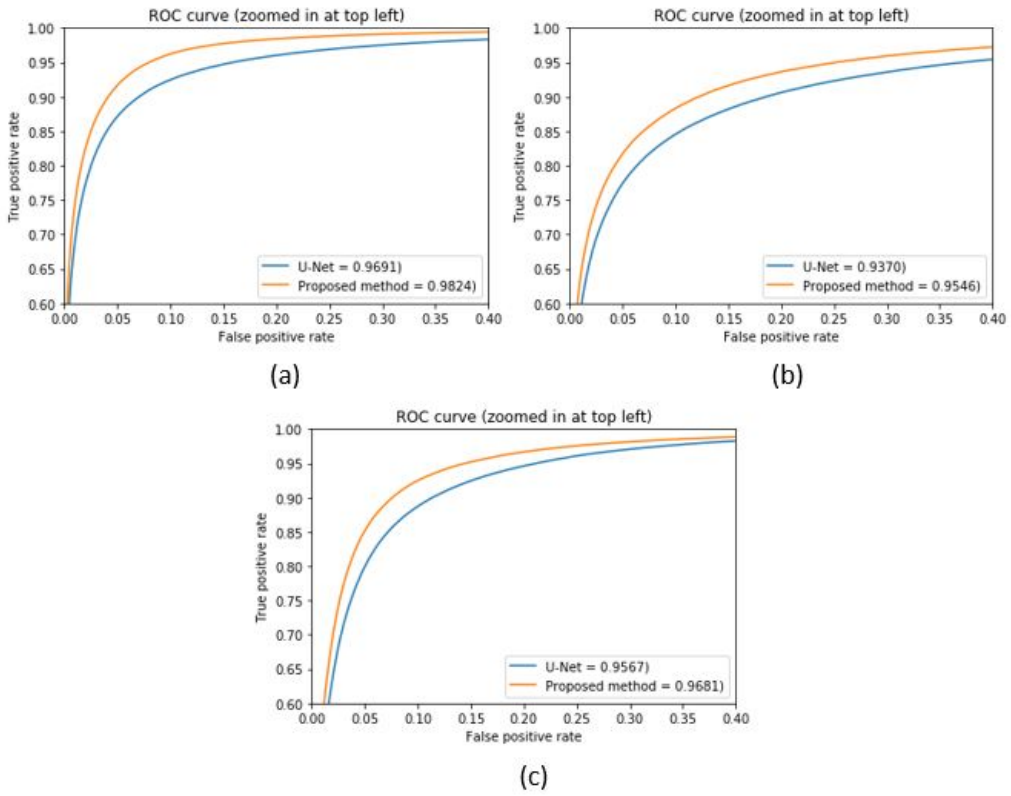


Fig. 4.5 Performance comparisons of the architecture without and with our proposed regularizer using receiver operating characteristics curve analysis on different datasets. (a) DRIVE (b) STARE and (c) CHASEDB1

thin vessels can also be found out through significantly low Sn values on all three datasets (Table 4.1). However, the proposed regularization technique efficiently connects the disjoint blood vessels with higher AUC values (< 95%), thus demonstrating the reliability of the proposed network in diagnosing the DR accurately (Figure 4.5). The ROC curve for all DRIVE test images dataset also presented in Fig. 4.13-4.17.

The performance of the U-Net-like CNN architecture with our proposed EC-based regularizers is compared over the GS and the combined EC with GS regularizers (Table 4.2). The performance of the combined GISO regularizer revealed almost similar performance with DISO regularizer, demonstrated that the proposed EC-based regularizer efficiently acquired both the smoothness level and isolated pixel level interpretation. Thus, EC-based regularizer shows

Table 4.2 Performance comparisons of the proposed regularizer over graph-based smoothing and combined graph-based smoothing with isolated object regularizer on various dataset.

Model	**Sn**	**Sp**	**Acc**	**AUC**	**Jaccard**	**F1-Score**	**PRC**
DRIVE							
L_{MISO}	0.8463	0.9759	0.9600	0.9824	0.7218	0.8385	0.9226
L_{DISO}	0.8705	0.9700	0.9578	0.9825	0.7165	0.8349	0.9227
L_{GS}	0.7802	0.9854	0.9602	0.9817	0.7063	0.8279	0.9195
L_{GISO}	0.8621	0.9721	0.9586	0.9825	0.7185	0.8362	0.9225
STARE							
L_{MISO}	0.7401	0.9754	0.9403	0.9546	0.6418	0.7873	0.8715
L_{DISO}	0.8105	0.9680	0.9480	0.9691	0.6648	0.7987	0.8819
L_{GS}	0.7739	0.9650	0.9365	0.9499	0.5852	0.7384	0.8661
L_{GISO}	0.7096	0.9806	0.9402	0.9548	0.6402	0.7806	0.8699
CHASEDB1							
L_{MISO}	0.6080	0.9879	0.9533	0.9681	0.5724	0.7034	0.8167
L_{DISO}	0.7906	0.9770	0.9600	0.9786	0.6426	0.7824	0.8606
L_{GS}	0.7427	0.9799	0.9583	0.9752	0.6184	0.7642	0.8495
L_{GISO}	0.7926	0.9746	0.9580	0.9757	0.6319	0.7744	0.8526

high potential in learning multiple properties by eliminating two different regularizers in diagnosing the blood vessel connections for DR. The performance of the combined GISO regularizer revealed higher performance than the regularizer using only with GS for vessel detection. It clearly demonstrated that the isolated object regularizer is beneficial in forcing the architecture to learn the segmentation region more effectively.

In addition, to prove the effectiveness of our approach with different α value, we used the MNIH road dataset [52] that has similar vessel-like structure as a fundus image. The performance comparison of our approach with baseline in terms of dice coefficient score (DCS) explains our approach performs better than the baseline with α value at 0.001 (Table 4.3).

Table 4.3 Performance comparison of our proposed method with baseline on the MNIH road dataset.

Regularizer	**Parameter α**	**DCS**
L_{DICE} (Baseline)	-	0.8782
$L_{DICE} + \alpha E_{OUT}$	0.01	0.8748
$L_{DICE} + \alpha E_{OUT}$	0.001	**0.8856**
$L_{DICE} + \alpha E_{OUT}$	0.0001	0.8775

Table 4.4 Performance comparison of the proposed approach with state-of-the-art methods

Methods	Sn	Sp	Acc	AUC
Azzopardi[3]	0.7655	0.9704	0.9442	0.9614
Li et al. [36]	0.7569	0.9816	0.9527	0.9738
Liskowski[40]	0.7763	0.9768	0.9495	0.972
Fu et al. [16]	0.7603	-	0.9523	-
Dasgupta et al. [9]	0.7691	0.9801	0.9533	0.9744
Roychowdhury[63]	0.725	**0.983**	0.952	0.962
Chen et al. [8]	0.7426	0.9735	0.9453	0.9516
Yan et al. [76]	0.7653	0.9818	0.9542	0.9752
Yan et al. [77]	0.7631	0.982	0.9538	0.975
Jin et al. [22]	0.7963	0.98	0.9566	0.9802
Proposed method	**0.8463**	0.9759	**0.9600**	**0.9824**

4.6 Performance comparison against state-of-the-arts

The proposed EC-based regularizer network is compared of state-of-the-art methods (Table 4.4). Our proposed approach achieved 0.9824 of AUC and 0.9600 of accuracy, which is higher than the other existing methods. This can be attributed to the fact that this study considered the branches of blood vessels as almost connected with each other, which means the vessel structure has only one isolated object in which all thin vessels are connected or exhibits minimum number of isolated objects. Hence, our proposed regularizer produced acceptable results by pushing the network to make the output region consisting minimum number of isolated objects.

Though the qualitative performance of our method is much better than the other methods, the quantitative computation based on the given ground truth shows almost similar accuracy and AUC value compared with the state-of-the-art methods. This is because of the missing annotation of the small vessels in the ground truth misleading the measurements and considering the predicted small vessels as false positives (figure 4.6). Moreover, experts ignored the small vessels due to the low contrast conditions or noise artifacts; our approach proved its efficiency in detecting small vessels even in the low quality images as well.

The other reasons of the proposed network showed lesser Sp than the other methods may be the regions such as fovea, optical disk, and lesion detected in unhealthy fundus image distracted the network training. However, including the other class information in the training may suppress the misclassification and could improve the Sp value. Compared with the existing architectures, our proposed architecture is simple and accurate. In [22], the U-Net architecture included deformable convolutional block layer in encoder and decoder part increased the computational complexity with large number of dimensions. Whereas, in our approach, we simply added one layer to evaluate the EC to calculate the number of isolated objects from the output of the last layer for accurate detection of vessel regions. Though our approach proved its capability in delineating the small branches and disconnected vessel regions better than the conventional architectures, the EC is estimated by treating pixel-by-pixel cost computational time. Thus, there is room for further improvement of the proposed regularizer to make it less computational costs.

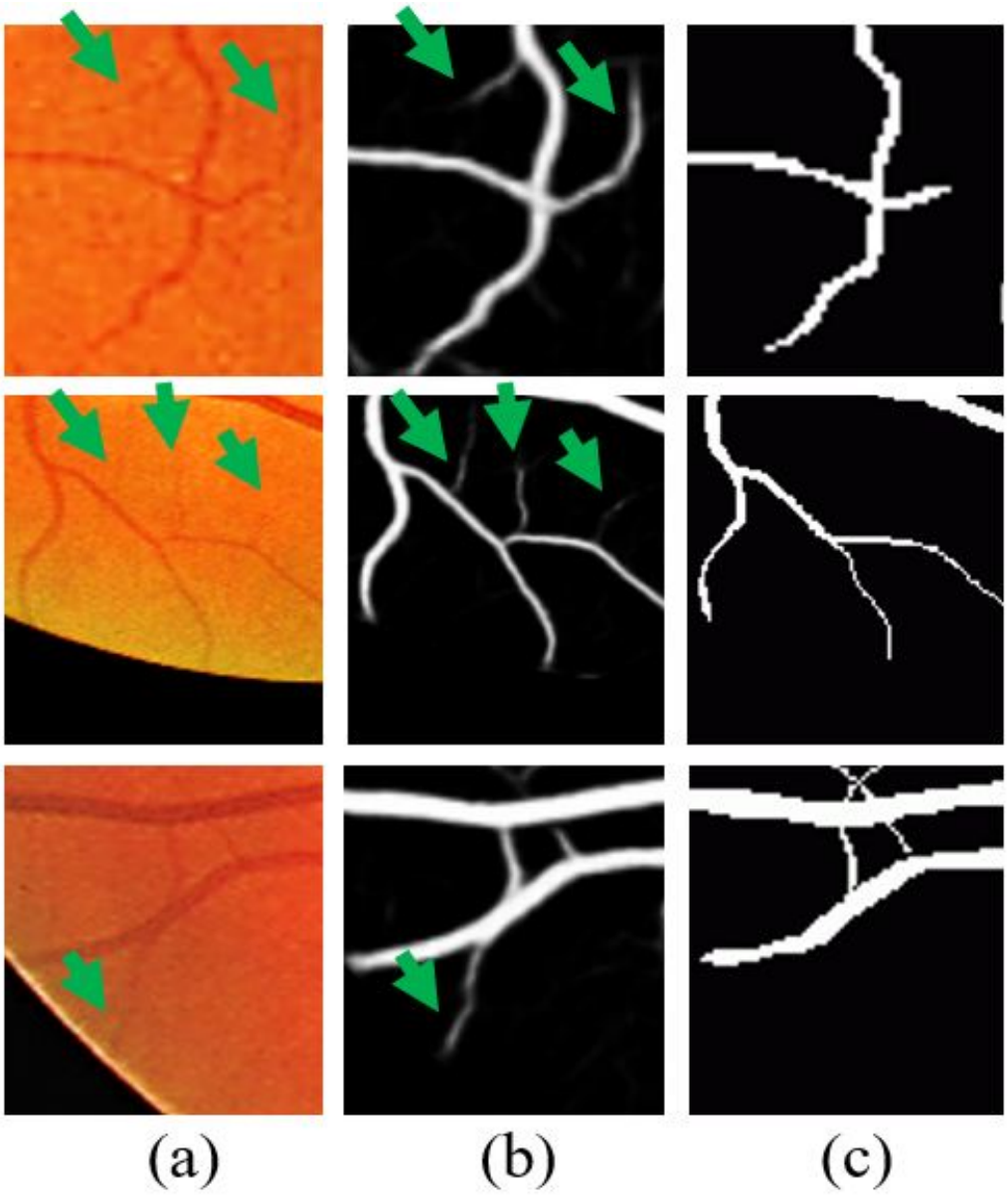


Fig. 4.6 Performance comparison of the proposed method with ground truth in predicting small vessels along with the large vessels. (a) input images, (b) prediction results (green arrow indicates small vessel detection) (c) ground truths (missing annotation of small vessels compared to the input image).

4.7 Sensitivity to noise evaluation for segmentation

We conducted experiments to evaluate the sensitivity of the segmentation achieved by our method to noise. We used the test set of the DRIVE dataset and added Gaussian noise with different levels of noise variances (from 0.01 to 0.1).

Table 4.5 Performance comparison of the measure of complexity of learning of the proposed approach with the state-of-the-art methods

Method	Time	Architecture
Azzopardi et al.[3]	10s	Non-Deep learning
Staal et al. [67]	15m	Non-Deep learning
Roychowdhury [63]	2.5s	Non-Deep learning
Liskowskiet al. [40]	92s	CNN
Luo et al. [46]	31.17s	SIFCN
Fu et al. [16]	1.3s	DeepVessel
Tan et al. [70]	10m	CNN
Jin et al. [22]	15.3s	DUNet
U-Net (Baseline)	23s	U-Net
Proposed Method	23s	U-Net+EC Regularizer

As shown in figure 4.7, our approach delineated the vessels with the AUC values ranging from 0.9824 to 0.9825, accuracy values ranging from 0.9577 to 0.9578 and sensitivity values ranging from 0.8693 to 0.8705. The small increasing variations of detection performance value (0.0001) with increasing levels of noise is observed, which demonstrates that the proposed approach is robust to noise.

4.8 Measure of complexity of learning

We investigated the measures of complexity of learning of our approach over the other methods during inference stage. We chose the computational complexity of deep learning and non deep learning methods (Table 4.5). We compared the computational complexity of learning between deep learning and non-deep learning methods and found that our approach does not show much differences compared with other methods. Furthermore, the measures of complexity of learning time of the baseline network is almost similar with ours. It indicates the addition of regularization term with the existing architecture does not tend

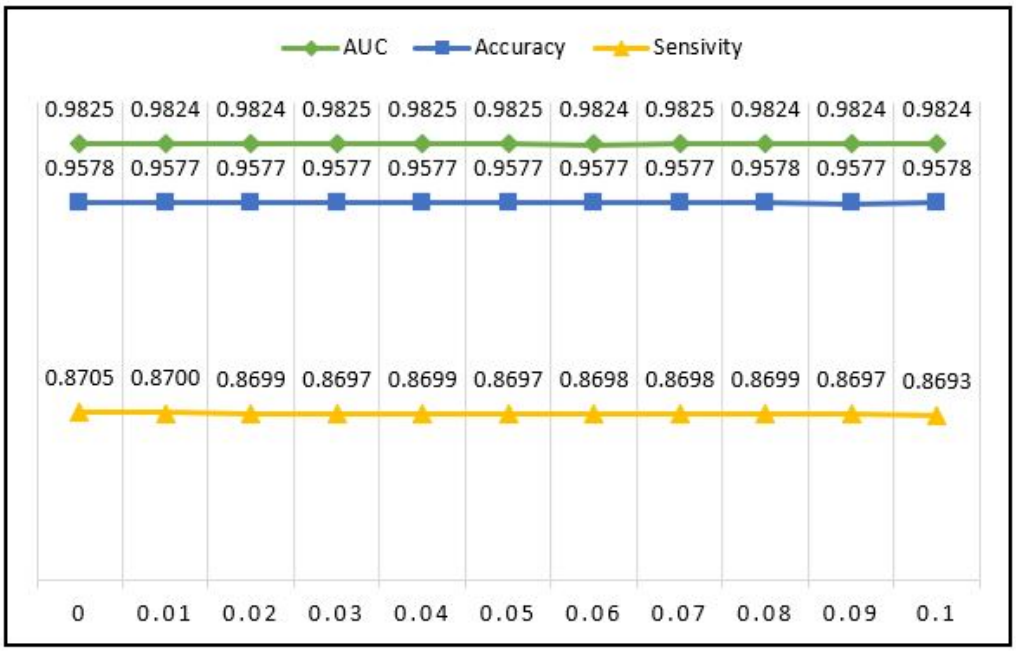


Fig. 4.7 Noise sensitivity evaluation on DRIVE dataset interms of AUC, accuracy and sensitivity. Evaluating Gaussian noise with the ranges of 0.01 to 0.1.

to increase the execution time and hence the complexity of learning of our approach is not sensitive to the performance of detection.

Though our approach is effective and accurate to identify the connectivity of pixels, it poses the restriction on implementing on the tree-like objects. The boundary cannot be well defined if the algorithm encounters any loop-like objects, which is the limitation of this study. However, it can be addresed using a suitable regularizer based on an high-quality ground truth. In future work, we plan to extend the proposed regularizer using the graph theory approach to estimate the number of isolated objects on image. Furthermore, it would be efficient to incorporate the vessel regions and lesions segmentation in one end-to-end network. This study also suggested that the proposed segmentation approach may be useful in non-medical applications that contains vessel-like structures, such as palmprints segmentation for biometric systems.

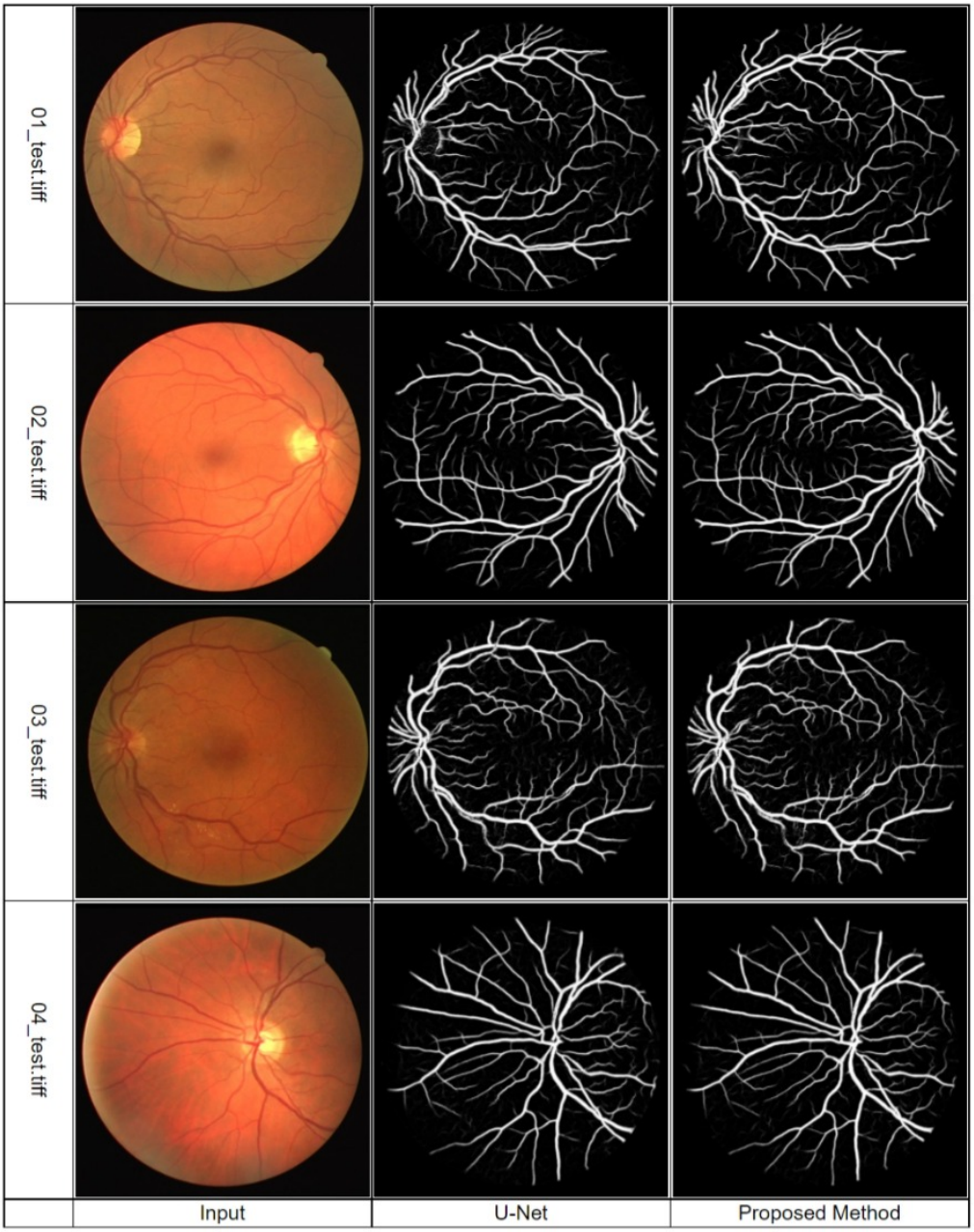


Fig. 4.8 Comparation result between proposed network with and without EC regularizer (Cont.)

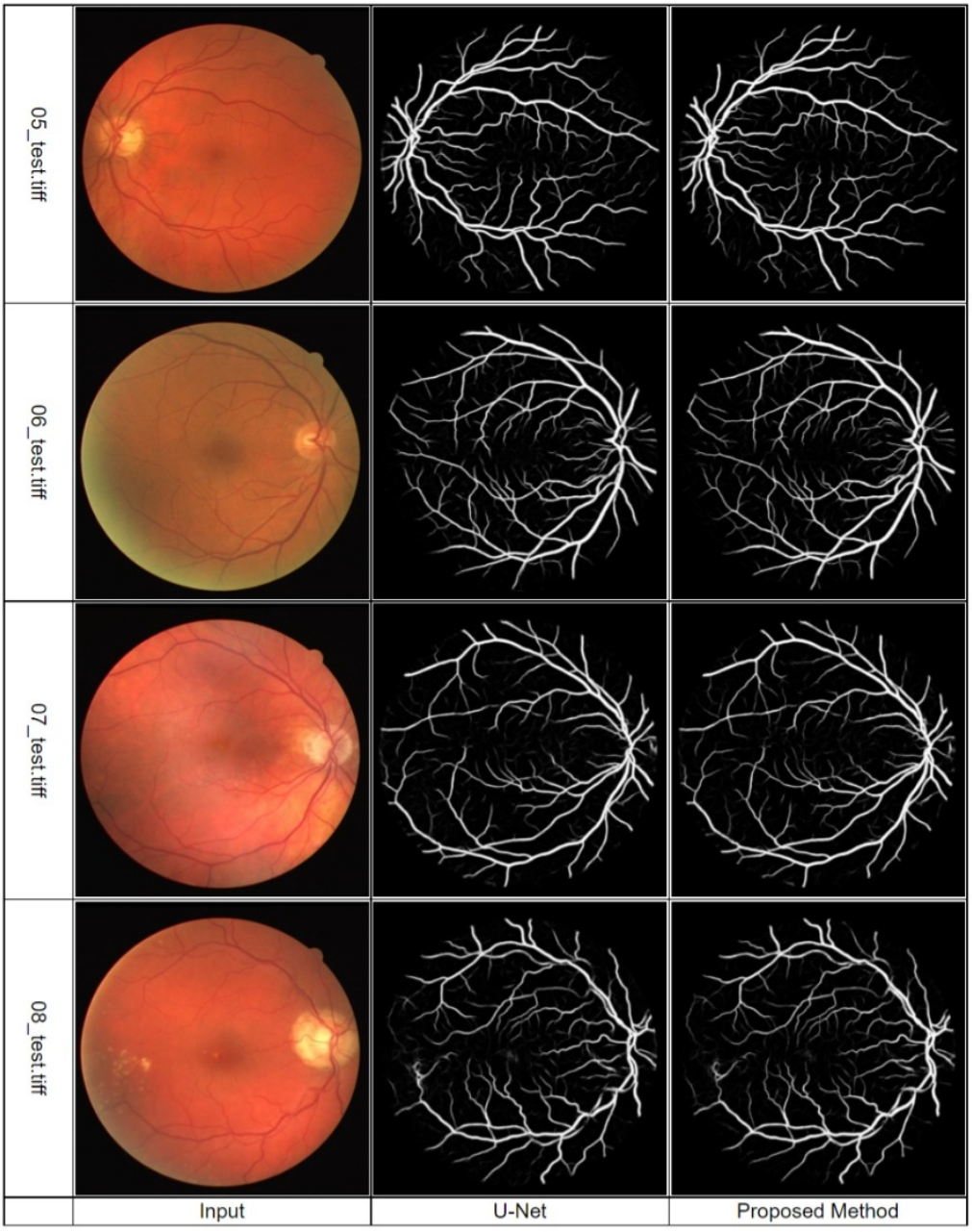


Fig. 4.9 Comparation result between proposed network with and without EC regularizer (Cont.)

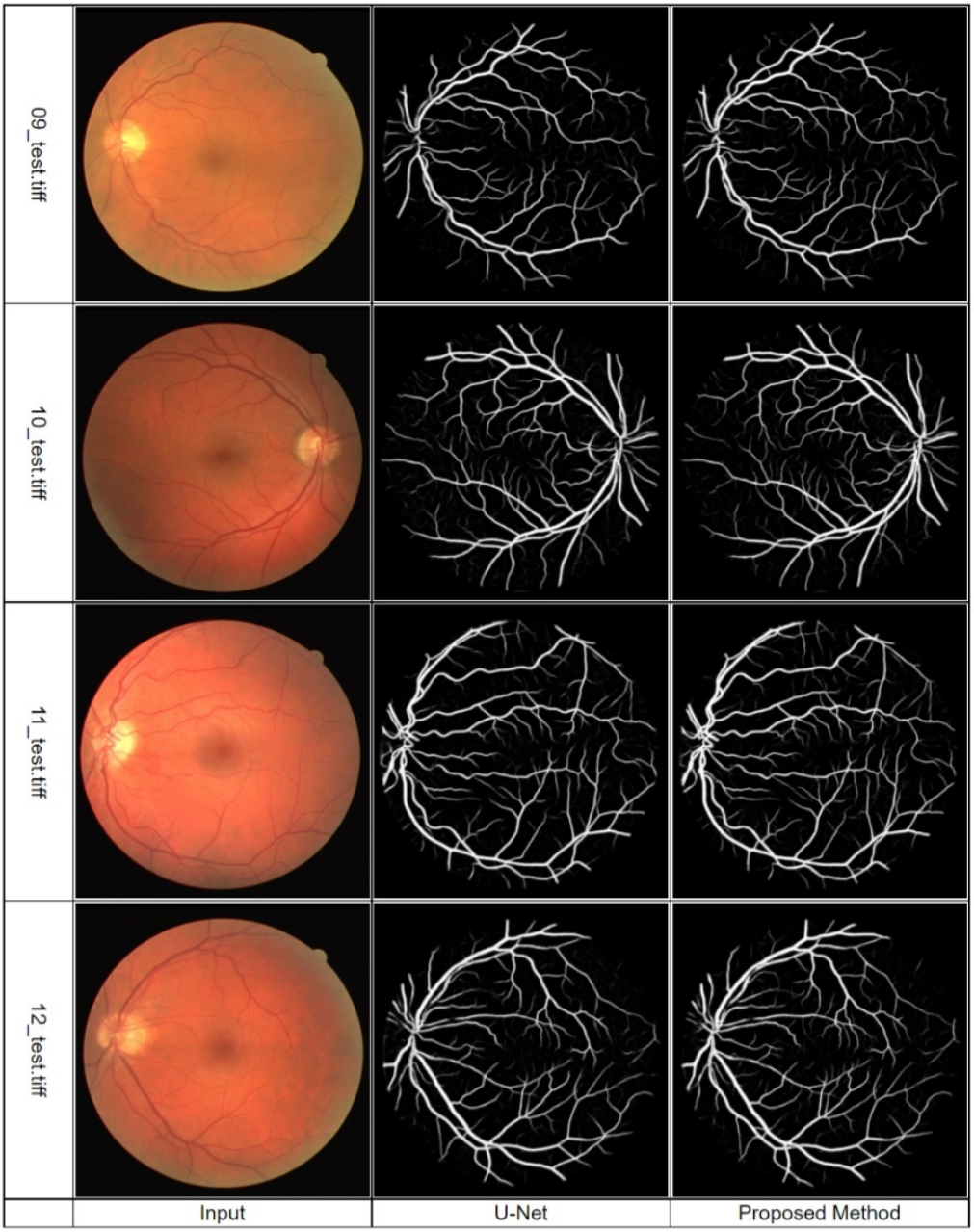


Fig. 4.10 Comparation result between proposed network with and without EC regularizer (Cont.)

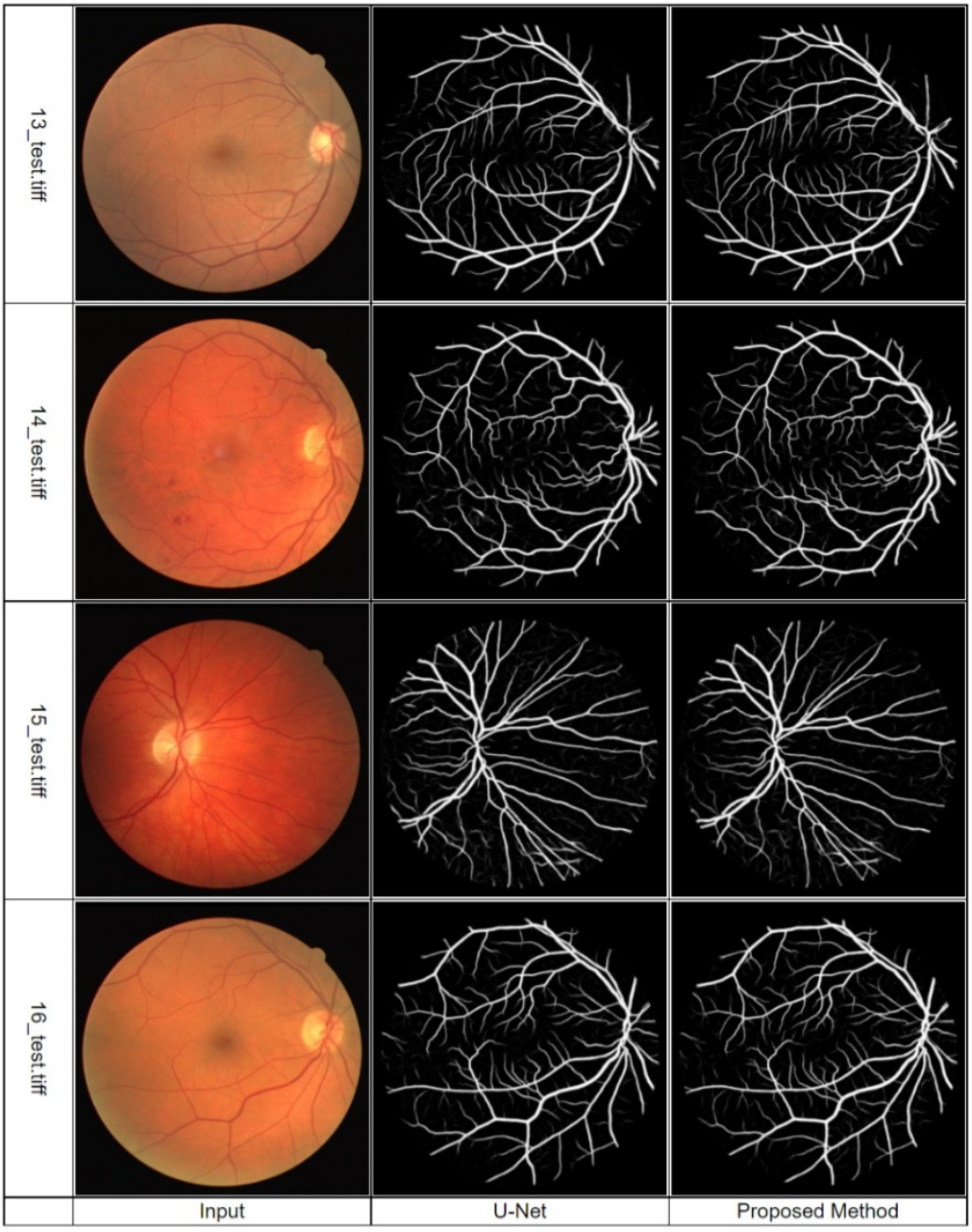


Fig. 4.11 Comparation result between proposed network with and without EC regularizer (Cont.)

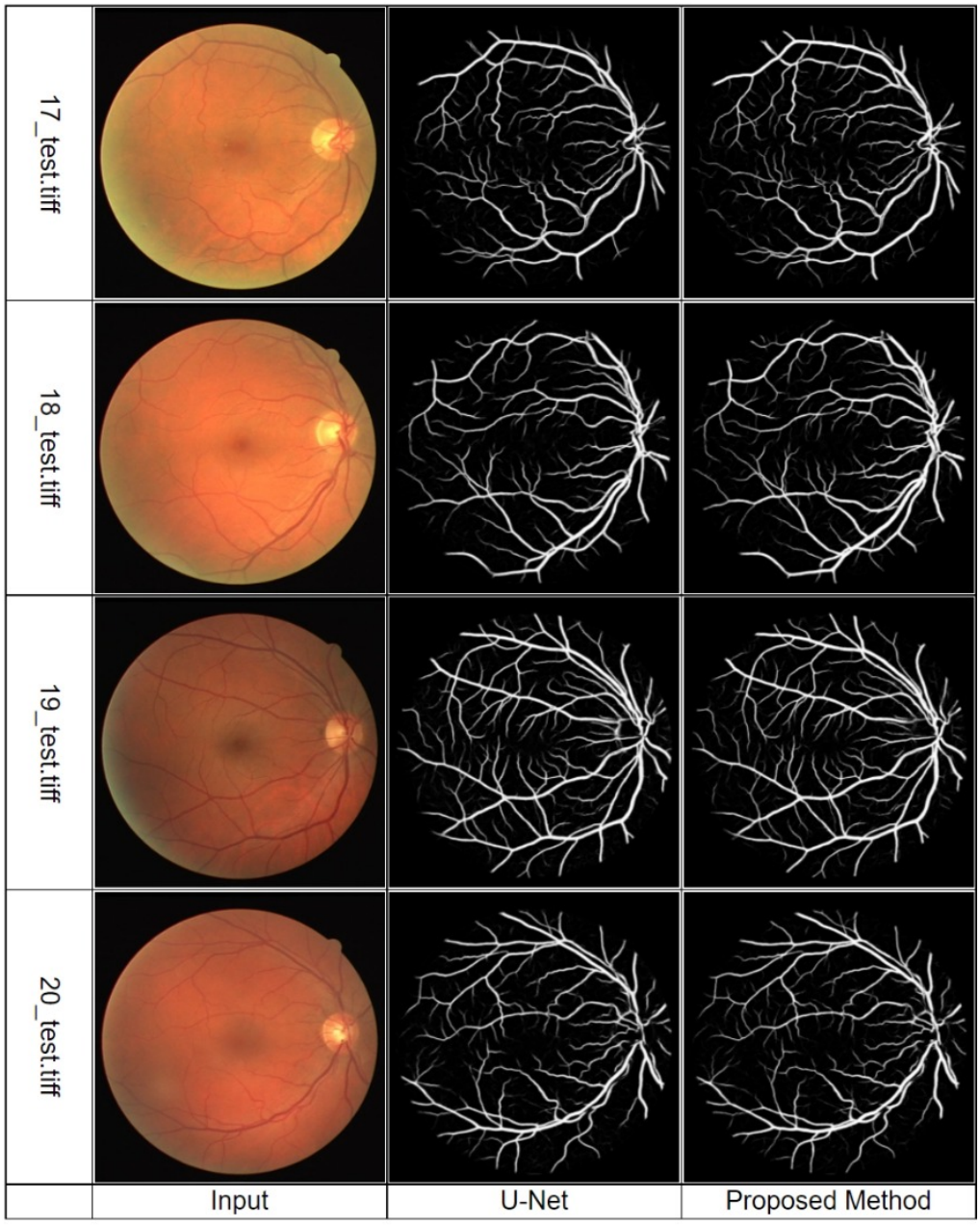


Fig. 4.12 Comparation result between proposed network with and without EC regularizer (Finish.)

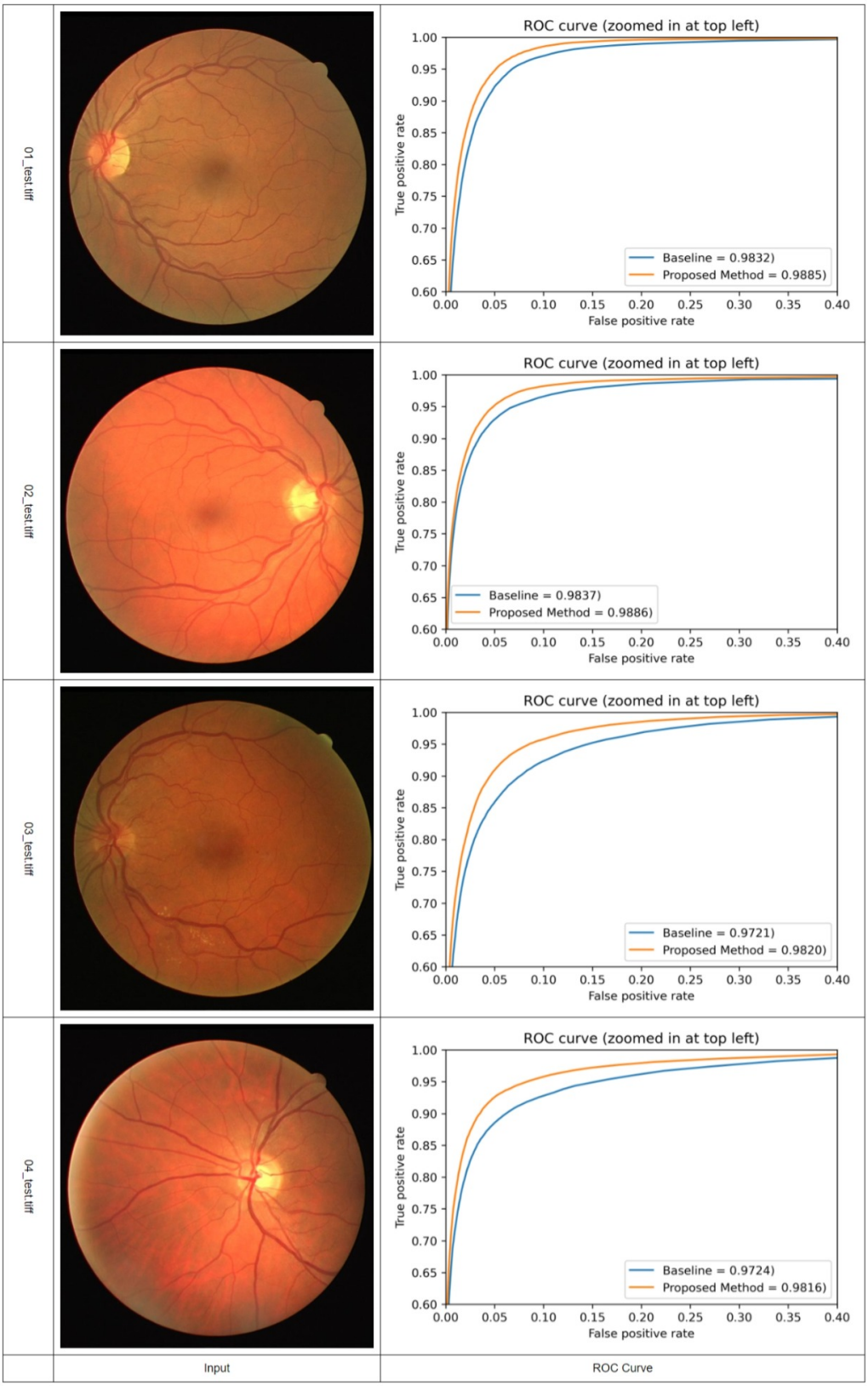


Fig. 4.13 Comparation of ROC Curve between with and without proposed regularizer on DRIVE dataset (Cont.)

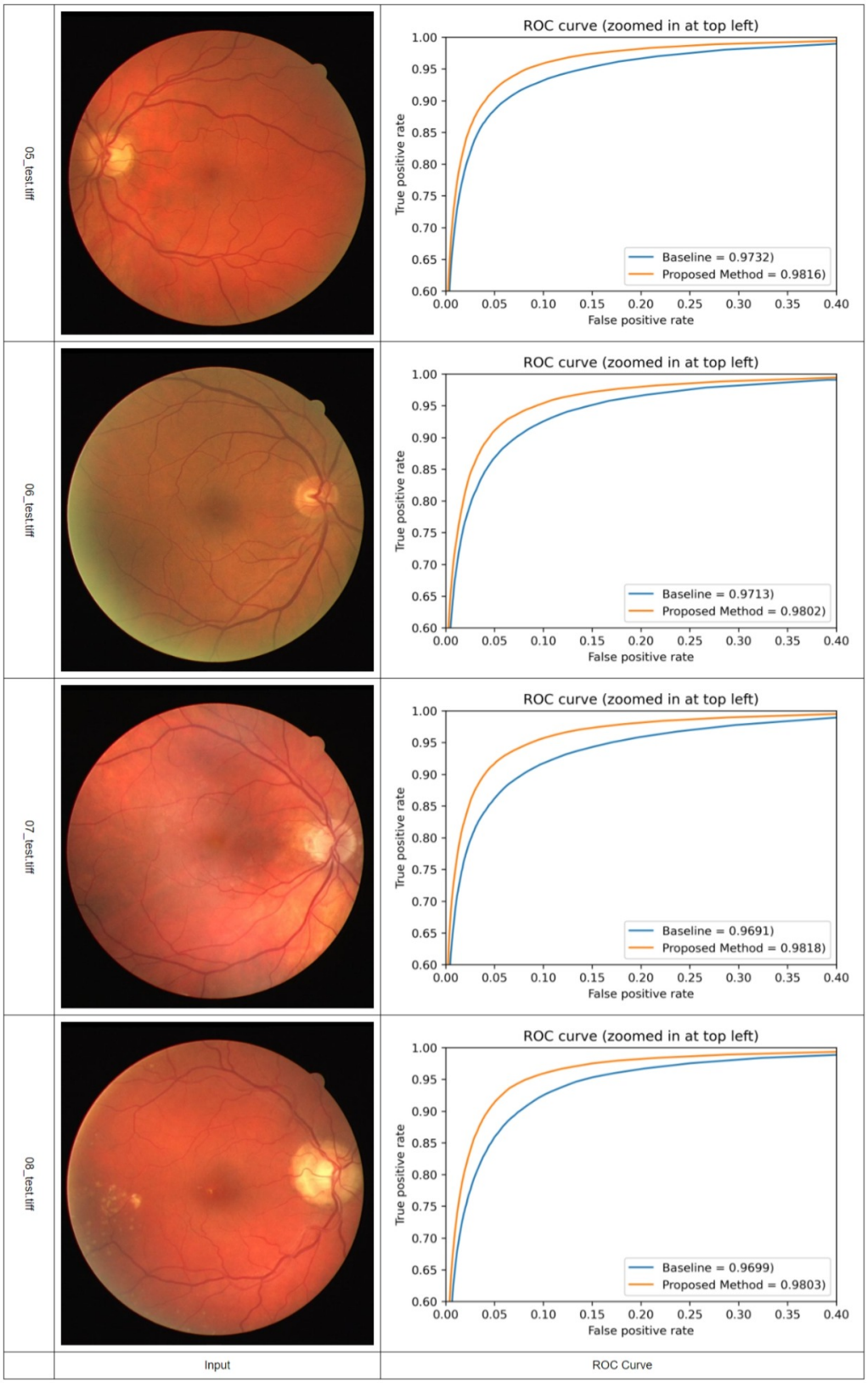


Fig. 4.14 Comparation of ROC Curve between with and without proposed regularizer on DRIVE dataset (Cont.)

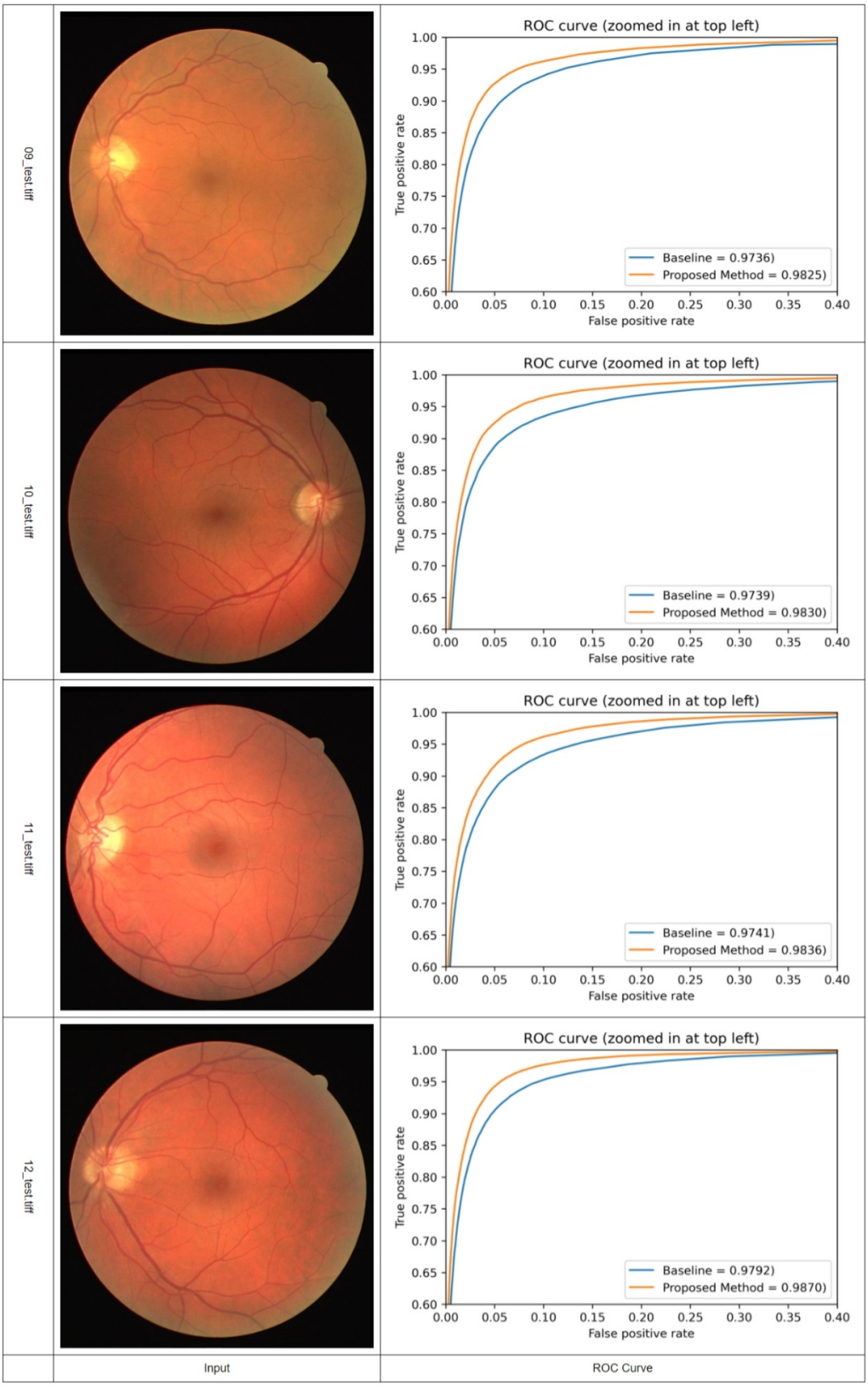


Fig. 4.15 Comparation of ROC Curve between with and without proposed regularizer on DRIVE dataset (Cont.)

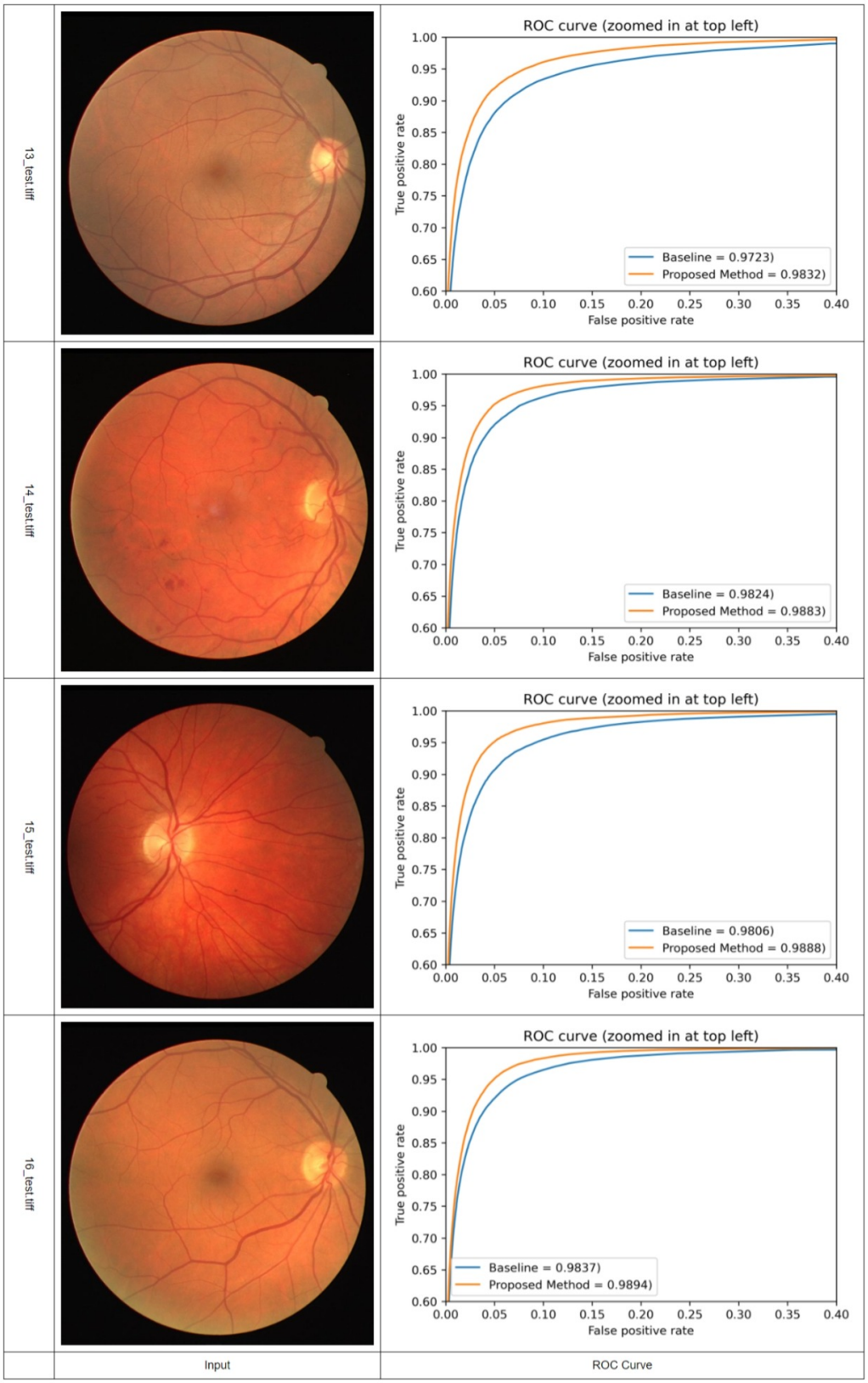


Fig. 4.16 Comparation of ROC Curve between with and without proposed regularizer on DRIVE dataset (Cont.)

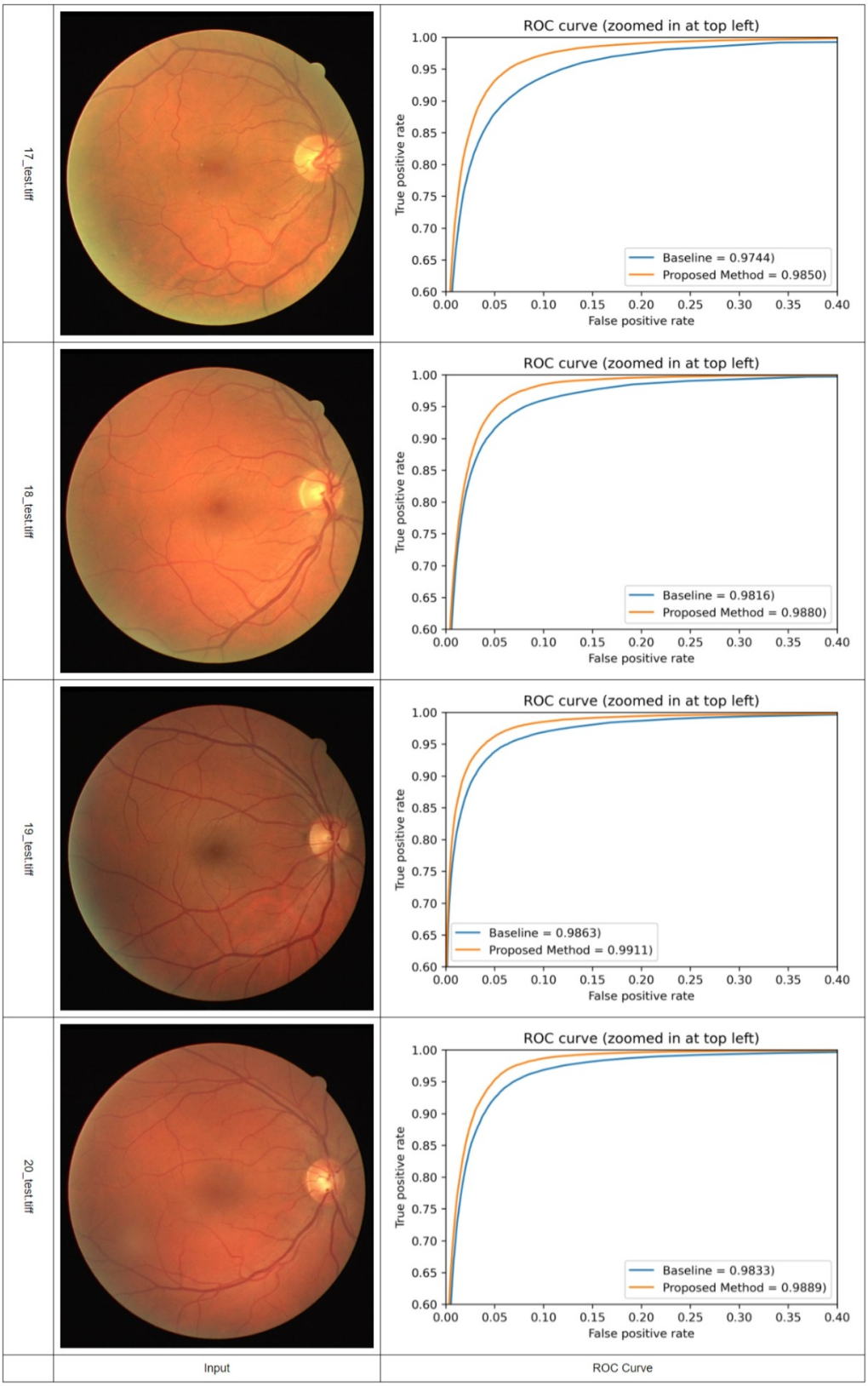


Fig. 4.17 Comparation of ROC Curve between with and without proposed regularizer on DRIVE dataset (End.)

Chapter 5

Conclusion

5.1 Conclusion and Future Works

This study takes advantage of the neighboring pixel information in the image as additional information on the CNN. This information is constructed as a regularization term in the objective function. In this section, we provide a summary, limitations, and future works of our work.

In chapter 2, we newly proposed a graph based smoothing regularization term with the loss function in the U-net framework for the segmentation of small vessels in the retinal image. The proposed regularization term effectively computing the graph laplacians on both vessels and its background regions and thus it significantly reduced the segmentation errors and reconnected small fragmented vessels. Our approach can segment more number of vessels and almost reconnecting all isolated vessels than the baseline U-net without regularizer. Compared to other state-of-the-art methods, our approach demonstarted its improvement in retaining width of the small vessel and disjointed vessel connectivity through its high AUC value. Future work will focus on implementing the proposed regularization term on different retinal image datasets and different segmentation CNN architectures.

In chapter 3, we proposed a novel graph approach to construct graph Laplacian regularization term based on differences of neighboring pixels for deep convolutional neural network. Compared with baseline architecture regularizer, the proposed approach demonstrated improvements in enhancing the detail of boundary structure on various types of datasets. Furthermore, its also offers the advantage of smoother line sections, and more consistent sections of the same color. These findings suggest that the proposed system can be very reliable for computer vision tasks. The calculation of GLRDN on a pixel-by-pixel basis is a limitation of this approach. In the future, we plan to expand the work by using sparse graphs to reduce computational costs.

In chapter 4, we proposed EC-based regularizers to estimate the number of isolated objects in U-Net-like deep CNN architecture for delineating small retinal vessel connections on a fundus image. The proposed EC-based MISO and DISO regularizers approach demonstrated improvement in retaining the width of the small and disjointed vessel connectivities through its high AUC values. Our approach can segment more number of vessels and can reconnect all isolated vessels, which is superior than the performance of the baseline architecture not using the proposed regularizer. Furthermore, it also demonstrated that the isolated objects regularizer is beneficial in forcing the architecture to learn the smoothness and isolated pixel level of interpretation of vessel regions. Compared with other state-of-the-art methods, the EC-based regularizer improved the performance in localizing and connectivity between the pixels of the vessel regions with high acceptable value of AUC value. These findings indicate that the proposed system could be a highly reliable detection system for DR.

References

[1] Amin, M. and Yan, H. (2010). High speed detection of retinal blood vessels in fundus image using phase congruency. *Soft Computing*, 15:1217–1230.

[2] Ando, R. K. and Zhang, T. (2007). Learning on graph with laplacian regularization. *In Advances in neural information processing systems, pp*, pages 25–32.

[3] Azzopardi, G., Strisciuglio, N., Vento, M., and Petkov, N. (2015). Trainable cosfire filters for vessel delineation with application to retinal images. *Medical image analysis*, 19(1):46–57.

[4] Badrinarayanan, V., Kendall, A., and Cipolla, R. (2017). Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE transactions on pattern analysis and machine intelligence*, 39(12):2481–2495.

[5] Chakraborti, T., Jha, D., Chowdhury, A., and Jiang, X. A. (2014). self-adaptive matched filter for retinal blood vessel detection. *Machine Vision and Applications*, 26:55–68.

[6] Chen, L. C., Papandreou, G., Kokkinos, I., Murphy, K., and Yuille, A. L. (2017). Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE transactions on pattern analysis and machine intelligence*, 40(4):834–848.

[7] Chen, M. H. and Yan, P. F. (1988). A fast algorithm to calculate the Euler number for binary images. *Pattern Recognition Letters*, 8(5):295–297.

[8] Chen, Y. (2017). A labeling-free approach to supervising deep neural networks for retinal blood vessel segmentation. preprint, arXiv.

[9] Dasgupta, A. and Singh, S. (2017a). A fully convolutional neural network based structured prediction approach towards the retinal vessel segmentation. In *Proceedings - International Symposium on Biomedical Imaging*, pages 248–251.

[10] Dasgupta, A. and Singh, S. A. (2017b). In *fully convolutional neural network based structured prediction approach towards the retinal vessel segmentation*, pages 18–21. 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017.

[11] Diaz De Leon Santiago, J. L. and Sossa-Azuela, J. H. (1996). On the computation of the Euler number of a binary object. *Pattern Recognition*, 29(3):471–476.

[12] Dinesh, C., Cheung, G., and Bajić, I. V. (2020). Point cloud denoising via feature graph laplacian regularization. *IEEE Transactions on Image Processing*, 29:4143–4158.

[13] Dong, C., Loy, C. C., He, K., and Tang, X. (2015). Image super-resolution using deep convolutional networks. *IEEE transactions on pattern analysis and machine intelligence*, 38(2):295–307.

[14] Fraz, M. M., Remagnino, P., Hoppe, A., Uyyanonvara, B., Rudnicka, A. R., Owen, C. G., and Barman, S. A. (2012a). Blood vessel segmentation methodologies in retinal images–a survey. *Computer Methods and Programs in Biomedicine*, 108(1):407–433.

[15] Fraz, M. M., Remagnino, P., Hoppe, A., Uyyanonvara, B., Rudnicka, A. R., Owen, C. G., and Barman, S. A. (2012b). An ensemble classification-based approach applied to retinal blood vessel segmentation. *IEEE Transactions on Biomedical Engineering*, 59(9):2538–2548.

[16] Fu, H., Xu, Y., Lin, S., Wong, D. W. K., and Liu, J. (2016a). Deepvessel: Retinal vessel segmentation via deep learning and conditional random field. In *International conference on medical image computing and computer-assisted intervention*, pages 132–139. Springer.

[17] Fu, H., Xu, Y., Wong, D., and Liu, J. (2016b). In *Retinal vessel segmentation via deep learning network and fully-connected conditional random fields*. 2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI). 698–701.

[18] Hakim, L., Yudistira, N., Kavitha, M. S., and Kurita, T. (2019). U-Net with Graph Based Smoothing Regularizer for Small Vessel Segmentation on Fundus Image. In *International Conference on Neural Information Processing*, pages 515–522. Springer.

[19] Hu, K., Zhang, Z., Niu, X., Zhang, Y., Cao, C., Xiao, F., and Gao, X. (2018a). Retinal vessel segmentation of color fundus images using multiscale convolutional neural network with an improved cross-entropy loss function. *Neurocomputing*, 309:179–191.

[20] Hu, K., Zhang, Z., Niu, X., Zhang, Y., Cao, C., Xiao, F., and Gao, X. (2018b). Retinal vessel segmentation of color fundus images using multiscale convolutional neural network with an improved cross-entropy loss function. *Neurocomputing*, 309:179–191.

[21] Hu, W., Cheung, G., Li, X., and Au, O. C. . (2014). October). In International, I., editor, *Graph-based joint denoising and super-resolution of generalized piecewise smooth images*, pages 2056–2060. Conference on Image Processing (ICIP) . IEEE.

[22] Jin, Q., Meng, Z., Pham, T. D., Chen, Q., Wei, L., and Su, R. (2019). DUNet: A deformable network for retinal vessel segmentation. *Knowledge-Based Systems*, 178:149–162.

[23] Kavitha, M., Lee, C.-H., Shibudas, K., Kurita, T., and Ahn, B.-C. (2020). Deep learning enables automated localization of the metastatic lymph node for thyroid cancer on 131 i post-ablation whole-body planar scans. *Scientific reports*, 10(1):1–12.

[24] Kavitha, M., Yudistira, N., and Kurita, T. (2019). Multi instance learning via deep cnn for multi-class recognition of alzheimer's disease. In *2019 IEEE 11th International Workshop on Computational Intelligence and Applications (IWCIA)*, pages 89–94. IEEE.

[25] Kavitha, M. S., Kurita, T., Park, S. Y., Chien, S. I., Bae, J. S., and Ahn, B. C. (2017). Deep vector-based convolutional neural network approach for automatic recognition of colonies of induced pluripotent stem cells. *PLoS ONE*, 12(12):1–18.

[26] Ke, T. W., Hwang, J. J., Liu, Z., and Yu, S. X. (2018). Adaptive affinity fields for semantic segmentation. In *Proceedings of the European Conference on Computer Vision (ECCV*, pages 587–602.

[27] Khan, M., Soomro, T., Khan, T., Bailey, D., Gao, J., and Mir, N. (2016a). In *Automatic retinal vessel extraction algorithm based on contrast-sensitive schemes*, pages 1–5. 2016 International Conference on Image and Vision Computing New Zealand (IVCNZ.

[28] Khan, M. A., Soomro, T. A., Khan, T. M., Bailey, D. G., Gao, J., and Mir, N. (2016b). Automatic retinal vessel extraction algorithm based on contrast-sensitive schemes. In *2016 International conference on image and vision computing New Zealand (IVCNZ)*, pages 1–5. IEEE.

[29] Kim, J., Lee, J. K., and Lee, K. M. (2016). Accurate image super-resolution using very deep convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1646–1654.

[30] Kirbas, C. and Quek, F. (2004). A review of vessel extraction techniques and algorithms. *ACM Computing Surveys (CSUR)*, 36(2):81–121.

[31] Kobayashi, T. (2018). Trainable co-occurrence activation unit for improving convnet. In *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1273–1277. IEEE.

[32] Kobrin Klein, B. E. (2007). Overview of epidemiologic studies of diabetic retinopathy. *Ophthalmic epidemiology*, 14(4):179–183.

[33] Kurita, T. and Hidaka, A. (2019). *Statistical Pattern Recognition and Discriminant Analysis*. Corona Publishing CO., LTD, Tokyo, JPN, 1st edition.

[34] LeCun, Y. and Cortes, C. (2010). MNIST handwritten digit database.

[35] Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., and Acosta, A. (2017). ... and shi. In *W*, pages 4681–4690. Photo-realistic single image super-resolution using a generative adversarial network. In Proceedings of the IEEE conference on computer vision and pattern recognition.

[36] Li, Q., Feng, B., Xie, L., Liang, P., Zhang, H., and Wang, T. (2016a). A cross-modality learning approach for vessel segmentation in retinal images. *IEEE Transactions on Medical Imaging*, 35(1):109–118.

[37] Li, Q., Feng, B., Xie, L., Liang, P., Zhang, H., and Wang, T. A. (2016b). Cross-modality learning approach for vessel segmentation in retinal images. *IEEE Transactions on Medical Imaging*, 35:109–118.

[38] Li, X., Yang, Y., Zhao, Q., Shen, T., Lin, Z., and Liu, H. (2020). Spatial pyramid based graph reasoning for semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8950–8959.

[39] Lian, S., Li, L., Lian, G., Xiao, X., Luo, Z., and Li, S. (2019). A global and local enhanced residual u-net for accurate retinal vessel segmentation. *IEEE Computer Architecture Letters*, pages 1–1.

[40] Liskowski, P. and Krawiec, K. (2016). Segmenting Retinal Blood Vessels With Deep Neural Networks. *IEEE transactions on medical imaging*, 35(11):2369–2380.

[41] Liu, X., Zhai, D., Chen, R., Ji, X., Zhao, D., and Gao, W. (2018a). Depth super-resolution via joint color-guided internal and external regularizations. *IEEE Transactions on Image Processing*, 28(4):1636–1645.

[42] Liu, Y., Yang, S., Li, B., Zhou, W., Xu, J., Li, H., and Lu., Y. (2018b). Affinity derivation and graph merge for instance segmentation. In *Proceedings of the European Conference on Computer Vision (ECCV*, pages 686–703.

[43] Long, J., Shelhamer, E., and Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440.

[44] Lu, Q., Chen, C., Xie, W., and Luo, Y. (2020). Pointngcnn: Deep convolutional networks on 3d point clouds with neighborhood graph filters. *Computers and Graphics*, 86:42–51.

[45] Luo, L., Chen, D., and Xue, D. (2018). Retinal blood vessels semantic segmentation method based on modified U-Net. In *Proceedings of the 30th Chinese Control and Decision Conference, CCDC 2018*, pages 1892–1895. IEEE.

[46] Luo, Y., Cheng, H., and Yang, L. (2016). Size-invariant fully convolutional neural network for vessel segmentation of digital retinal images. In *2016 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA)*, pages 1–7. IEEE.

[47] Lézoray, O. and Grady, L. (2012). Graph theory concepts and definitions used in image processing and analysis. *Image Processing and Analysing With Graphs: Theory and Practice*, pages 1–24.

[48] Maninis, K.-K., Pont-Tuset, J., Arbelaez, P., and Van Gool, L. (2016). Deep retinal image understanding. In *International conference on medical image computing and computer-assisted intervention*, pages 140–148. Springer.

[49] Marin, D., Aquino, A., Gegundez-Arias, M. E., and Bravo, J. M. A. (2011). New supervised method for blood vessel segmentation in retinal images by using gray-level and moment invariants-based features. In *IEEE Trans*, page 146–158. Med. Imaging. 30.

[50] Matsui, Y., Ito, K., Aramaki, Y., Fujimoto, A., Ogawa, T., Yamasaki, T., and Aizawa, K. (2017). Sketch-based manga retrieval using manga109 dataset. *Multimedia Tools and Applications*, 76(20):21811–21838.

[51] Melinscak, M., Prentasic, P., and Loncaric, S. (2015). Retinal vessel segmentation using deep neural networks. *In VISAPP*, 1:577–582.

[52] Mnih, V. (2013). *Machine learning for aerial image labeling*. University of Toronto (Canada).

[53] Mojoo, J., Zhao, Y., Kavitha, M. S., Miyao, J., and Kurita, T. (2020). Completion of missing labels for multi-label annotation by a unified graph laplacian regularization. *IEICE Transactions on Information and Systems*, 103(10):2154–2161.

[54] Oliveira, A., Pereira, S., and Silva, C. A. (2018). Retinal vessel segmentation based on fully convolutional neural networks. *Expert Systems with Applications*, 112:229–242.

[55] Orlando, J. I., Prokofyeva, E., and Blaschko, M. B. (2016). A discriminatively trained fully connected conditional random field model for blood vessel segmentation in fundus images. *IEEE transactions on Biomedical Engineering*, 64(1):16–27.

[56] Orlando, J. I., Prokofyeva, E., and Blaschko, M. B. A. (2017). Discriminatively trained fully connected conditional random field model for blood vessel segmentation in fundus images. In *IEEE Trans*, page 16–27. Biomed. Eng. 64.

[57] Ortiz, B., Ram, J., and Mart, F. J. (2019). Retinal Blood Vessel Segmentation by Multi-channel Deep Convolutional Autoencoder. In *International Joint Conference SOCO'18-CISIS'18-ICEUTE'18*, volume 771, pages 37–46.

[58] Otsu, N. and Kurita, T. (1988). A New Scheme for Practical Flexible and Intelligent Vision Systems. In *MVA*, pages 431–435.

[59] Pang, J. and Cheung, G. (2017). Graph laplacian regularization for image denoising: Analysis in the continuous domain. *IEEE Transactions on Image Processing*, 26(4):1770–1785.

[60] Primitivo, D., Alma, R., Erik, C., Arturo, V., Edgar, C., Marco, P.-C., and Daniel, Z. (2019). A hybrid method for blood vessel segmentation in images. *Biocybernetics and Biomedical Engineering*, 39(3):814–824.

[61] Ronneberger, O., Fischer, P., and Brox, T. (2015a). U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer.

[62] Ronneberger, O., Fischer, P., and Brox, T. (2015b). U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer.

[63] Roychowdhury, S., Koozekanani, D. D., and Parhi, K. K. (2014). Blood vessel segmentation of fundus images by major vessel extraction and subimage classification. *IEEE journal of biomedical and health informatics*, 19(3):1118–1128.

[Set] Set, C. An Image Dataset of Random Cartoons.

[65] Sharma, A. and Rani, S. (2016). In *An automatic segmentation and detection of blood vessels and optic disc in retinal images*. 2016 International Conference on Communication and Signal Processing (ICCSP.

[66] Solomon, C. and Breckon, T. (2011). *Fundamentals of Digital Image Processing: A practical approach with examples in Matlab*. John Wiley & Sons.

[67] Staal, J., Abramoff, M. D., Niemeijer, M., Viergever, M. A., and Van Ginneken, B. (2004). Ridge-based vessel segmentation in color images of the retina. *IEEE Transactions on Medical Imaging*, 23(4):501–509.

[68] Strisciuglio, N., Azzopardi, G., Vento, M., and Petkov, N. (2016). Supervised vessel delineation in retinal fundus images with the automatic selection of b-cosfire filters. *Machine Vision and Applications*, 27(8):1137–1149.

[69] Tagore, M. R. N., Kande, G. B., Rao, E. K., and Rao, B. P. (2013). *August. Segmentation of retinal vasculature using phase congruency and hierarchical clustering*. In 2013 International Conference on Advances in Computing, Communications and Informatics (ICACCI) (pp. 361-366). IEEE.

[70] Tan, J. H., Acharya, U. R., Bhandary, S. V., Chua, K. C., and Sivaprasad, S. (2017). Segmentation of optic disc, fovea and retinal vasculature using a single convolutional neural network. *Journal of Computational Science*, 20:70–79.

[71] Tang, M., Djelouah, A., Perazzi, F., Boykov, Y., and Schroers, C. (2018). Normalized Cut Loss for Weakly-Supervised CNN Segmentation. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1:1818–1827.

[72] Wang, C., Zhao, Z., Ren, Q., Xu, Y., and Yu, Y. (2019). Dense u-net based on patch-based learning for retinal vessel segmentation. *Entropy*, 21(2):168.

[73] Wang, S., Yin, Y., Cao, G., Wei, B., Zheng, Y., and Yang, G. (2015). Hierarchical retinal blood vessel segmentation based on feature and ensemble learning. *Neurocomputing.*, 149:708–717.

[74] Wilson, C., Horton, M., Cavallerano, J., and Aiello, L. M. (2005). Addition of primary care-based retinal imaging technology to an existing eye care professional referral program increased the rate of surveillance and treatment of diabetic retinopathy. *Diabetes care*, 28(2):318–322.

[75] Xu, Y., Wu, Z., Chanussot, J., and Wei, Z. (2020). Hyperspectral images super-resolution via learning high-order coupled tensor ring representation. *IEEE transactions on neural networks and learning systems*, 31(11):4747–4760.

[76] Yan, Z., Yang, X., and Cheng, K.-T. (2018a). Joint segment-level and pixel-wise losses for deep learning based retinal vessel segmentation. *IEEE Transactions on Biomedical Engineering*, 65(9):1912–1923.

[77] Yan, Z., Yang, X., and Cheng, K.-T. (2018b). A three-stage deep learning model for accurate retinal vessel segmentation. *IEEE journal of biomedical and health informatics*, 23(4):1427–1436.

[78] Yudistira, N., Kavitha, M. S., Itabashi, T., Iwane, A. H., and Kurita, T. (2020). Prediction of Sequential Organelles Localization under Imbalance using A Balanced Deep U-Net. *Scientific Reports*, 10(1):1–11.

[79] Zeng, J., Pang, J., Sun, W., and Cheung, G. (2019). Deep graph laplacian regularization for robust denoising of real images. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1759–1768. IEEE.

[80] Zhang, J., Chen, Y., Bekkers, E., Wang, M., Dashtbozorg, B., and ter Haar Romeny, B. M. (2017). Retinal vessel delineation using a brain-inspired wavelet transform and random forest. *Pattern Recognition*, 69:107–123.

[81] Zhang, L., Li, X., Arnab, A., Yang, K., Tong, Y., and Torr, P. H. S. (2019). Dual graph convolutional network for semantic segmentation. preprint, arXiv.

[82] Zhao, Y., Rada, L., Chen, K., Harding, S., and Zheng, Y. (2015). Automated vessel segmentation using infinite perimeter active contour model with hybrid region information with application to retinal images. *IEEE Transactions on Medical Imaging*, 34:1797–1807.

[83] Zhou, C., Zhang, X., and Chen, H. (2020a). A new robust method for blood vessel segmentation in retinal fundus images based on weighted line detector and hidden markov model. *Computer methods and programs in biomedicine*, 187:105231.

[84] Zhou, W., Wang, Y., Chu, J., Yang, J., Bai, X., and Xu, Y. (2020b). Affinity space adaptation for semantic segmentation across domains. *IEEE Transactions on Image Processing*, 30:2549–2561.