EDGE-AWARE VESSEL SEGMENTATION USING SCRIBBLE SUPERVISION

Zhanqiang Guo[1,2], *Jianjiang Feng*[1,2(✉)], *and Jie Zhou*[1,2]

[1] Department of Automation, Tsinghua University, Beijing, China
[2] Beijing National Research Center for Information Science and Technology, Beijing, China

ABSTRACT

Accurate segmentation of blood vessels is critical for diagnosing various diseases. However, the complexity of manually labeling vessels impedes the practical adoption of fully supervised methods. To alleviate this challenge, we propose a weakly supervised vessel segmentation framework. Our approach leverages scribble annotation to train the Unet and identifies reliable foreground and background regions. Addressing the issue of insufficient boundary information inherent in scribble annotation, we incorporate a conventional approach specifically designed to leverage the innate structural attributes of vessels for edge detection, subsequently ensuring effective edge supervision. In addition, a bilateral filtering module is introduced to improve edge awareness of network. Furthermore, to augment the quantity of annotated pixels, we employ an image mixing strategy for data augmentation, thereby enhancing the network's segmentation capability. The experimental results on three datasets show that our framework outperforms the existing scribble-based methods.

***Index Terms*—** Vessel Segmentation, Weakly-Supervised, Edge Awareness

1. INTRODUCTION

Vascular structures exhibit wide prevalence within the human body, encompassing notable instances such as retinal vessels and coronary arteries. Automatic extraction of blood vessels from vascular imaging assumes a vital role in clinical diagnosis [1, 2]. The fully-supervised segmentation methods [3, 4] based on deep convolutional neural networks (CNNs) necessitate significant annotated images for effective training. Due to the complexity of objects, especially for the vascular structure that is characterized by numerous diminutive branches, the manual annotation process is exceedingly time-consuming and arduous, even for professional experts [5].

The challenge of acquiring labeled data has led to significant interest in weak supervision techniques for image segmentation. Weak labels encompass various categories such as image-level labels [6], bounding boxes [7], scribbles [8, 9] and keypoints [10]. Image-level annotation is not typically employed in segmentation tasks with fixed object classes, such as vessels and background in our task [10]. Vascular structures exhibit extensive spatial extension, rendering bounding boxes insufficient in providing abundant informative cues. Similarly, due to the complex nature of blood vessels, the locating of keypoints becomes more cumbersome. Compared with bounding boxes and points, scribble annotation offers greater flexibility, particularly for segmenting irregularly shaped foregrounds [11]. Therefore, in our study, we adopt graffiti as our weak annotation. Notably, foreground scribble annotation is also time-consuming for the smaller blood vessels located at the periphery. Therefore, we mark N pairs of vessel foreground points ($N = 12$ for retinal vessels and $N = 6$ for coronary arteries), encompassing both major vessels and small branches to the maximum extent feasible. And we employ the Dijkstra's algorithm [12] to obtain annotation lines, followed by manually selecting the correct foreground lines as annotation, as shown in Fig. 1.

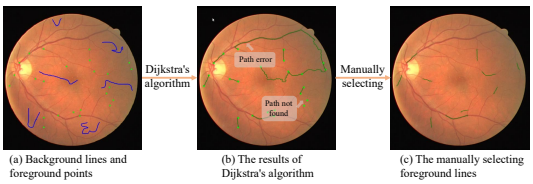

Fig. 1. The process of annotating retinal vessels. The average time for a radiologist to perform weak label is 51.4s, significantly less than the time (exceeding 10min) needed for complete annotation from scratch.

Scribble annotation, commonly applied in weakly supervised segmentation, has garnered significant attention in the segmentation of natural images and large organs [9, 13]. The utilization of a fully supervised network [3] and Cross-entropy (CE) loss is the most intuitive training algorithm. Based on it, Obukhov et al. [8] proposed gated Conditional Random Field (CRF) loss for the unlabeled pixels to supervise the segmentation. Zhang et al. [13] proposed CycleMix to adopt the mixup strategy with a dedicated design of random occlusion to perform increments and decrements of scribbles. And Luo et al. [14] employed a dual-branch network and dynamically mixed the two decoders' predictions to generate pseudo labels for auxiliary supervision. However, these methods primarily focused on the segmentation of natural images and large organs, whereas the identification of

blood vessel boundaries requires specific attention. Although certain approaches integrated regularization functions to optimize boundaries [8], implicit constraints did not effectively contribute to the accurate delineation of vessel edges.

To address the challenges mentioned above, we present a novel weakly-supervised segmentation framework that emphasizes edge-attention awareness (as shown in Fig. 2). Initially, we employ the U-net [3] and CE loss for preliminary training, utilizing predefined thresholds to acquire reliable foreground and background regions. Subsequently, in response to the deficiency of boundary information in scribble annotation, a traditional algorithm incorporating vascular structural properties is employed to identify reliable boundary areas, which are then utilized to retrain the segmentation network. To further enhance edge awareness, we integrate the bilateral filter module into the network architecture. Additionally, drawing inspiration from [15], we adopt an improved image mixup strategy to refine the network's segmentation performance. The effectiveness of our method is verified on three public datasets.

2. METHOD

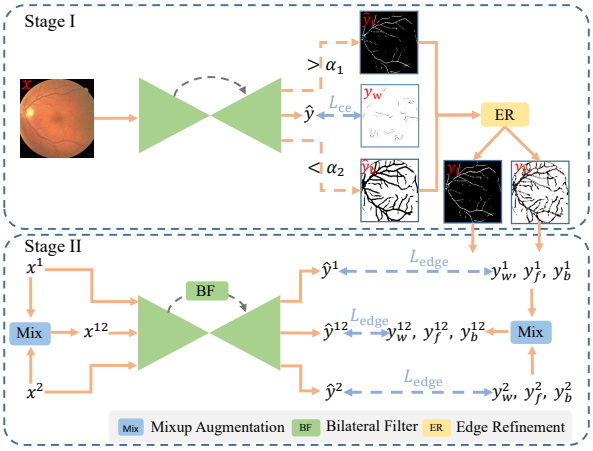

Fig. 2. Illustration of the proposed framework.

2.1. Initial Training and Edge Refinement (ER)

Let $x \in R^{C\times H\times W}$ denote the image, and y_{w} denote the scribble annotation. Initially, the Unet g_1 is trained with y_{w}, under the supervision of CE loss:

$$L_{\mathrm{ce}} = -\frac{1}{|\omega_0|}\sum_{i\in\omega_0}\log(1-\hat{y}_i) - \frac{1}{|\omega_1|}\sum_{i\in\omega_1}\log(\hat{y}_i), \quad (1)$$

where $\hat{y} = g_1(x)$, ω_0 is the set of labeled background pixels and ω_1 is the set of labeled foreground pixels. In this process, the absence of boundary information in scribble annotations poses a challenge, leading to inadequate learned boundary. To address this issue, we incorporate the Edge Refinement (ER) module, which enables the acquisition of precise blood vessel boundaries. Specifically, we first establish a larger threshold to obtain a credible foreground area ($\hat{y}_{\mathrm{f}} = (\hat{y} > \alpha_1)$) and a smaller threshold to derive a reliable background area ($\hat{y}_{\mathrm{b}} = (\hat{y} < \alpha_2)$). And we employ sobel operators in various directions to filter the image, followed by superposition to generate the blood vessel boundary map, as illustrated in Fig. 3(c). Furthermore, $\hat{y}_{\mathrm{f}}$ is refined to generate the skeleton map, where bifurcation points are eliminated, resulting in Fig. 3(b). For each point within the skeleton map, the edge enhancement map patch centered on it is cropped with the size of 15×15. Subsequently, we calculate the tangent and normal directions of the blood vessels based on the center line using curve fitting techniques (Fig. 3(d)). By analyzing the blood vessel enhancement map along the normal direction, we identify the position with the highest gray gradient, the sides of which are considered the foreground ($\tilde{y}_{\mathrm{f}}$) and background ($\tilde{y}_{\mathrm{b}}$) regions, respectively. The final generated foreground and background areas with boundary information are $y_{\mathrm{f}} = \tilde{y}_{\mathrm{f}} \cup \hat{y}_{\mathrm{f}}$ and $y_{\mathrm{b}} = \tilde{y}_{\mathrm{b}} \cup \hat{y}_{\mathrm{b}}$.

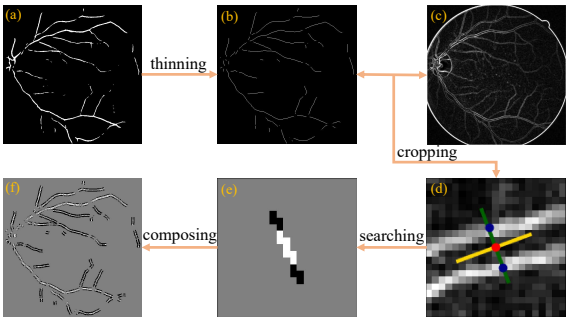

Fig. 3. The process of Edge Refinement. In (e) and (f), white and black pixels represent the foreground and background pixels, respectively, and gray pixels are unlabeled areas.

2.2. Bilateral Filter (BF) Module

In stage II, we retrain the network (g_2) under the supervision of graffiti annotations and generated boundary pseudo-labels:

$$L_{\mathrm{edge}} = L_{\mathrm{ce}} + \alpha\cdot(-\frac{\sum y_{\mathrm{b}}\log(1-\hat{y})}{\sum y_{\mathrm{b}}} - \frac{\sum y_{\mathrm{f}}\log(\hat{y})}{\sum y_{\mathrm{f}}}), \quad (2)$$

where $\hat{y} = g_2(x)$. Furthermore, within the domain of image edge perception, the bilateral filter [16] stands as a widely employed technique known for its ability to preserve edge sharpness while concurrently smoothing the image. To augment the network's edge awareness, we introduce the Bilateral Filtering module into the jumper section of the Unet. Specifically, for the feature $f \in R^{C\times H\times W}$ within the encoding section, it undergoes conversion into a single-channel feature map $\tilde{f} \in R^{H\times W}$ through the utilization of the 1×1 convolution module. Subsequently, we apply bilateral filters to obtain

smoothed features while retaining edge-awareness:

$$f_{\text{new}} = \frac{\sum_{(k,l)\in S(i,j)} \tilde{f}(i,j)\omega(i,j,k,l)}{\sum_{(k,l)\in S(i,j)} \omega(i,j,k,l)}, \tag{3}$$

where

$$\begin{aligned}\omega(i,j,k,l) =& \exp(-\frac{(i-k)^2+(j-l)^2}{2\sigma_{\text{s}}^2})\cdot \\ & \exp(-\frac{||\tilde{f}(i,j)-\tilde{f}(k,l)||^2}{2\sigma_{\text{r}}^2}),\end{aligned} \tag{4}$$

and $S(i,j)$ is a $m \times m$ area centered on (i,j), $\sigma_{\text{s}} = \sigma_{\text{r}} = 0.15 \times m + 0.35$. Within each encode layer, the value of m varies due to disparities in feature size. In our experiments, the values of m in the first to fourth layers are 9, 7, 5, and 3.

The final feature obtained through BF module is calculated as follows:

$$f_{\text{fin}}(c) = f(c) \cdot [1 + \text{Sigmoid}(f_{\text{new}})], c = 1, 2...C, \tag{5}$$

which is seamlessly integrated into the decoder's feature at the corresponding position using skip connection.

2.3. Mixup Strategy

Image mixup augmentation is a strategy employed to merge two images and corresponding labels, thereby augmenting the pixel count for scribble annotations. Taking inspiration from previous studies [13, 17], we employ an enhanced variant of the Cutmix strategy [15] to mix images and labels. Given two images and their corresponding labels $(x^1, y^1 = \{y^1_{\text{w}}, y^1_{\text{b}}, y^1_{\text{f}}\})$ and $(x^2, y^2 = \{y^2_{\text{w}}, y^2_{\text{b}}, y^2_{\text{f}}\})$, the resulting image obtained through the mix strategy is denoted as $(x^{12}, y^{12} = \{y^{12}_{\text{w}}, y^{12}_{\text{b}}, y^{12}_{\text{f}}\})$, and x^{12} is calculated as:

$$x^{12} = (1 - \mathbb{M}) \odot x^1 + \mathbb{M} \odot x^2, \tag{6}$$

where $\odot$ is the element-wise multiplication; $\mathbb{M} \in R^{H\times W}$ denotes a binary mask. Different from the random uniform strategy in [15], to increase the number of labeled pixels, we randomly select a point from the pre-existing scribble annotation and utilize it as the center for $\mathbb{M}$. The height (h) and width (w) of $\mathbb{M}$ are generated using random uniform sampling: $h \sim \text{Unif}(\frac{H}{8}, \frac{H}{2})$, $w \sim \text{Unif}(\frac{W}{8}, \frac{W}{2})$. The calculation of y^{12} is consistent with x^{12}, with the same $\mathbb{M}$.

While conducting the training of stage II, the loss function of the mixed image is incorporated. The final loss L_{fin} is expressed as:

$$L_{\text{fin}} = L_{\text{edge}}(x^1, y^1) + L_{\text{edge}}(x^2, y^2) + L_{\text{edge}}(x^{12}, y^{12}). \tag{7}$$

3. EXPERIMENTS

3.1. Dataset and Evaluation Metrics

We conduct validation of our method on three datasets, comprising of two retinal datasets and an X-ray coronary dataset. (i) XCAD dataset [5]: This dataset comprises 126 images with annotations. Each image has a resolution of 512×512. The random training/validation/testing case split is 68/16/42. (ii) DRIVE dataset [18]: It encompasses 40 fundus images with size of 565×584 pixels. It has been pre-divided into a training set and a test set, each containing 20 images. We randomly select 2 images from the training set to form the validation set. (iii) CHASEDB1 dataset [19]: This dataset comprises images obtained from 28 eyes of 14 ten-year-old children, with size of 999×960. We allocate the first 18 images for training, the subsequent 4 images for validation, and the remaining 6 images for testing.

Following the setting in [20], we uniformly resize images to 512×512 pixels on two retinal datasets. And we employ random rotating and flipping to augment the images. We evaluate the results with Dice Similarity Coefficient (DSC), Accuracy (ACC) and Area Under ROC (AUC). During training, we employed the adaptive moment estimation (Adam) algorithm with an initial learning rate of 0.001. The batch size was set to 4, and the maximum epoch was 6000. And the threshold values of α_1 and α_2 were 0.995 and 0.05 respectively, while hyperparameter α was 0.25. Our code is available at: https://github.com/gzq17/Scribble-Vessel-Segmentation.

3.2. Comparison with Other Methods

To assess the efficacy of our proposed segmentation algorithm, we replicated several weakly-supervised segmentation approaches, including the Baseline, employing the Unet architecture [3] with the scribble annotation's cross-entropy (CE) loss, as well as GatedCRF [8], Saliency [9], CycleMix [13], and DBDM [14]. Furthermore, we report the results of fully supervised model with the same network structure as proposed method as the upper bound.

Table 1. Comparison with other methods on three datasets, with the best performance highlighted in bold.

Method	XCAD			DRIVE			CHASEDB1		
	DSC(%)	ACC(%)	AUC(%)	DSC(%)	ACC(%)	AUC(%)	DSC(%)	ACC(%)	AUC(%)
Full-sup	75.86	97.66	98.03	77.78	95.97	97.24	75.87	96.15	97.62
Baseline (2016) [3]	65.59	96.06	94.62	63.07	94.05	91.75	64.10	94.15	96.02
GatedCRF (2019) [8]	72.12	96.80	92.69	47.59	87.84	90.97	46.75	88.92	91.76
Saliency (2020) [9]	70.49	96.58	**97.95**	49.32	85.10	91.22	58.09	91.31	94.98
Cyclemix (2022) [13]	60.98	96.24	82.87	63.62	95.05	92.79	67.40	94.78	**96.80**
DBDM (2022) [14]	71.25	96.67	91.15	54.51	88.83	92.81	49.86	86.66	90.76
Ours	**73.36**	**96.91**	97.58	**73.88**	**95.72**	**93.24**	**71.66**	**95.41**	96.65

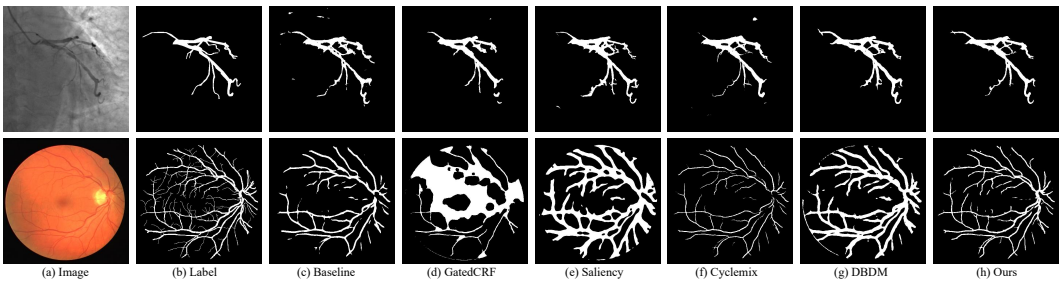

Fig. 4. Segmentation results on two testing images from XCAD and DRIVE dataset.

Table 2. The results of ablation study. We demonstrate the effectiveness of each component.

Method	XCAD			DRIVE			CHASEDB1		
	DSC(%)	ACC(%)	AUC(%)	DSC(%)	ACC(%)	AUC(%)	DSC(%)	ACC(%)	AUC(%)
S2	70.50	96.74	97.33	66.14	94.32	91.81	66.60	94.41	96.17
S2+ER	72.33	96.88	97.50	72.52	94.47	92.31	68.64	94.65	96.34
S2+ER+BF	72.35	96.77	97.44	72.79	95.56	92.45	70.55	95.36	95.07
S2+ER+BF+Mix	**73.36**	**96.91**	**97.58**	**73.88**	**95.72**	**93.24**	**71.66**	**95.41**	**96.65**

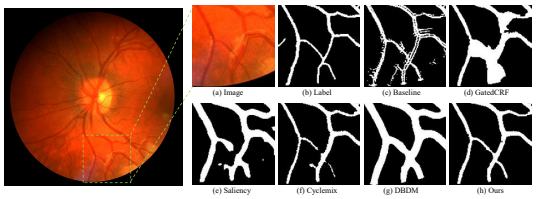

Fig. 5. Magnified view of one testing image from CHASEDB1 dataset.

Fig. 4 illustrates the segmentation results on two testing images from XCAD and DRIVE dataset. Our proposed method yields better connectivity of small vessels when compared to other algorithms. Another observation is that the boundary localization of blood vessels in the results obtained by the comparative methods exhibits significant deficiencies. For instance, on DRIVE dataset, the Baseline, gatedGRF, Saliency, and DBDM yield thicker blood vessels than annotation, whereas Cyclemix produces thinner vessels. In contrast, our method leverages the ER module to precisely locate boundaries and employs the bilateral filtering module to augment the network's perception of these boundaries. Consequently, the blood vessels segmented by our method closely resemble the thickness of authentic vessels. To provide a more comprehensive evaluation, Fig. 5 presents an enlarged view of one testing image derived from the CHASEDB1 dataset, further highlighting the distinct advantages offered by our proposed approach. These findings are consistent with the results in Tab 1, which demonstrates the performance of each method across all indicators.

3.3. Ablation study

We further investigate the effect of each module within our proposed framework, and the results are presented in Tab 2, where $S2$ denotes using the results of stage I for supervision during the training of stage II. The ER module effectively leverages the inherent structural characteristics of blood vessels, leading to precise boundary localization and substantial enhancement in the network's output results (e.g. 66.14% to 72.52% of DCS on DRIVE dataset). And BF module enhances the network's focus on edge awareness, especially on CHASEDB1 dataset, where the central region of the vessels is more similar to the background and the result is more sensitive towards edge detection compared the other two datasets. The observed performance improvement resulting from the Mixup strategy highlights the advantages of augmenting the number of labeled pixels and enhancing dataset diversity. Furthermore, certain compared methods (such as Cyclemix, shown in Tab 1) underperforms $S2$, potentially due to their primary use in segmenting large organs. The unique complexity of vessel structures can lead these methods to misguide network to learn incorrect information.

4. CONCLUSION

In this paper, we adopt the method involving annotating point pairs and background lines for acquiring scribble annotations of vessels. And we train the Unet using the weak labels and determine suitable thresholds to obtain reliable foreground and background areas. Recognizing the significance of boundary information, particularly for blood vessels, we propose an edge refinement module to accurately locate the boundaries. Furthermore, a bilateral filtering module is introduced to enhance the network's ability to detect vessel edges in stage II. We also incorporate an image mixup strategy to further enhance the performance of network. The results on three datasets demonstrate the effectiveness of our proposed weakly supervised segmentation framework. In the future, we will extend our method to 3D vessel datasets.

5. COMPLIANCE WITH ETHICAL STANDARDS

This study was conducted using human subject data, available in open access. No ethical approval was required.

6. REFERENCES

[1] Abdolhossein Fathi and Ahmad Reza Naghsh-Nilchi, "Automatic wavelet-based retinal blood vessels segmentation and vessel diameter estimation," *Biomedical Signal Processing and Control*, vol. 8, no. 1, pp. 71–80, 2013.

[2] Paolo Garrone et al., "Quantitative coronary angiography in the current era: principles and applications," *Journal of Interventional Cardiology*, vol. 22, no. 6, pp. 527–536, 2009.

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

[4] Wei Liao, "Progressive minimal path method with embedded CNN," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 4514–4522.

[5] Yuxin Ma et al., "Self-supervised vessel segmentation via adversarial learning," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 7536–7545.

[6] Yi Li, Yiduo Yu, Yiwen Zou, Tianqi Xiang, and Xiaomeng Li, "Online easy example mining for weakly-supervised gland segmentation from histology images," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2022, pp. 578–587.

[7] Reuben Dorent et al., "Inter extreme points geodesics for end-to-end weakly supervised image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2021, pp. 615–624.

[8] Anton Obukhov, Stamatios Georgoulis, Dengxin Dai, and Luc Van Gool, "Gated CRF loss for weakly supervised semantic image segmentation," *arXiv preprint arXiv:1906.04651v1*, 2019.

[9] Jing Zhang, Xin Yu, Aixuan Li, Peipei Song, Bowen Liu, and Yuchao Dai, "Weakly-supervised salient object detection via scribble annotations," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 12546–12555.

[10] Hui Qu et al., "Weakly supervised deep nuclei segmentation using partial points annotation in histopathology images," *IEEE Transactions on Medical Imaging*, vol. 39, no. 11, pp. 3655–3666, 2020.

[11] Qiuhui Chen and Yi Hong, "Scribble2d5: Weakly-supervised volumetric image segmentation via scribble annotations," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2022, pp. 234–243.

[12] Edsger W Dijkstra, "A note on two problems in connexion with graphs," in *Numerische Mathematik*, vol. 1, pp. 269–271. 1959.

[13] Ke Zhang and Xiahai Zhuang, "Cyclemix: A holistic strategy for medical image segmentation from scribble supervision," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 11656–11665.

[14] Xiangde Luo et al., "Scribble-supervised medical image segmentation via dual-branch network and dynamically mixed pseudo labels supervision," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2022, pp. 528–538.

[15] Sangdoo Yun et al., "Cutmix: Regularization strategy to train strong classifiers with localizable features," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 6023–6032.

[16] Carlo Tomasi and Roberto Manduchi, "Bilateral filtering for gray and color images," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*. IEEE, 1998, pp. 839–846.

[17] Krishna Chaitanya et al., "Semi-supervised and task-driven data augmentation," in *International Conference on Information Processing in Medical Imaging*. Springer, 2019, pp. 29–41.

[18] Joes Staal, Michael D Abràmoff, Meindert Niemeijer, Max A Viergever, and Bram Van Ginneken, "Ridge-based vessel segmentation in color images of the retina," *IEEE Transactions on Medical Imaging*, vol. 23, no. 4, pp. 501–509, 2004.

[19] Christopher G Owen et al., "Measuring retinal vessel tortuosity in 10-year-old children: validation of the computer-assisted image analysis of the retina (caiar) program," *Investigative Ophthalmology & Visual Science*, vol. 50, no. 5, pp. 2004–2010, 2009.

[20] Huisi Wu et al., "Scs-net: A scale and context sensitive network for retinal vessel segmentation," *Medical Image Analysis*, vol. 70, pp. 102025, 2021.