Vessel-Width-Based Metrics and Weight Masks for Retinal Blood Vessel Segmentation

João Paulo M. Linaris and Nina S. T. Hirata
Instituto de Matemática, Estatística e Ciência da Computação
Universidade de São Paulo

***Abstract*—Visual inspection of fundus images is widely used to diagnose various eye diseases, such as diabetic retinopathy and glaucoma. Alterations in the blood vessel structure are known to be related to these and other diseases. Therefore, automatic segmentation of vessels is of great interest. Existing methods, in general, struggle in segmenting thin vessels accurately. Motivated by this, we propose weight masks built based on the thickness of the vessels to be used with the well-known binary-cross entropy (BCE) loss. We also propose segmentation performance metrics conditioned on the vessel thickness. Then, we experiment different compositions of the loss function with the U-Net architecture to better understand the impact of each weight mask, especially on the segmentation of thin vessels. The experimental results indicate that the proposed weight masks present better performance regarding thin vessel segmentation, while, at the same time, maintaining competitive overall performance.**

I. Introduction

The retina is considered one of the least invasive biomarkers, allowing the detection of various diseases such as diabetic retinopathy (DR) through a non-invasive procedure. The growth of abnormal blood vessels in the retina is related to many complications like glaucoma, age-related macular degeneration, blindness and other diseases [1]. Early detection of vessel abnormalities could prevent serious complications such as blindness.

Based on this, in order to aid the detection of diseases such as DR that have impact on the morphology of retina blood vessels, segmentation of blood vessels in fundus images is a topic of great interest in the literature. Many algorithms have been proposed for this task, with varying degrees of success. After vessels are segmented, a subsequent analysis of various morphological features of the vessels such as width, curvature and vascular pattern can be performed. A widely applied model for the vessel segmentation task is the U-Net network [2], whose performance in segmentation has been greatly improved by many works [3]. However, accurately segmenting thin vessels remains a challenge for U-Net-based models, and, to our best knowledge, there is no well-established method to reliably evaluate segmentation of thin vessels.

To address these challenging issues, rather than using very large models, we use U-Net as the segmentation model [2] and focus on how to explore different ways to compute the loss. Closely tied to this, we also focus on defining evaluation metrics that is suited to assess segmentation performance restricted to vessels of a specific thickness range.

In this work, we propose not only a set of metrics for evaluating models performance on thin vessels, but also two weight masks that assign relevance to vessel pixels based on the thickness, which is then used with the BCE loss. The major contributions of our research are:

- A robust algorithm to compute the skeleton of the vessel network, which is fundamental for accurate thickness computation
- A method that computes evaluation metrics such as recall and precision restricted to vessels with thickness up to a given threshold
- Two weight masks based on vessel thickness that aim to enhance the model ability to segment thin vessels
- Extensive experimentation with distinct loss variations, that indicate improvement of thin vessel segmentation when the proposed weight masks are used.

II. Methodology

A. Vessel Thickness Calculation

The idea consists in first computing the distance transform [4] and the skeleton (the medial axis) [4] of the segmentation mask. The distance transform is an image of the same size, where each foreground pixel value is replaced with the distance to its closest background pixel. We then combine both of them to generate what we call the *Distances Skeleton*, an image where each skeleton pixel stores the distance to the background, and all remaining pixels are set to zero. Ultimately, the *Distances Skeleton* stores the local radius of the vessel at each skeleton pixel (and therefore we can recover the local vessel thickness).

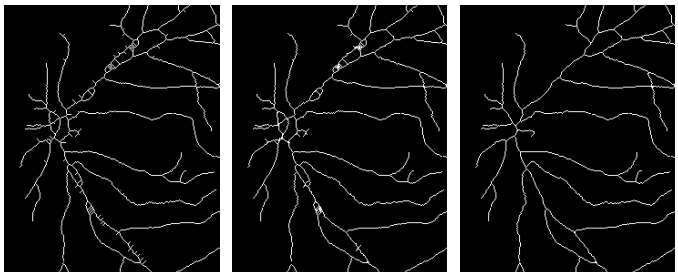

Fig. 1. Zoom in on pruning and closing to generate a "clean" skeleton. Skeleton computed on the original vessel mask (left); pruned skeleton (center); final result applying closing before skeleton computation and then applying pruning.

However, when we carelessly apply the skeleton algorithm on the segmentation mask, some spurious short ramifications

and segments within closed circular vessel shapes appear (see the first image in Fig. 1). To remove the spurious ramifications, we apply the morphological pruning operator [4] on the skeleton image (center image in Fig. 1) and, to avoid segments within closed loops, we employ morphological closing [4] before computing the skeleton. The final skeleton image, after applying these two additional transformations, is shown in the rightmost image in Fig. 1.

B. Performance metrics based on the thickness

We would like to be able to compute standard metrics such as recall and precision conditioned on vessel thickness. Given a certain thickness value (radius R), we use the *Distances Skeleton* image previously described and select the skeleton pixels with radius R' less than or equal to R. Then, we mark as foreground (white) all pixels within a square neighborhood of side $2\lceil R' \rceil$ for all pixels of the *Distances Skeleton*. Subsequently, a logical AND operation is performed between this squares-based mask and the original vessels mask, as illustrated in Fig 2.

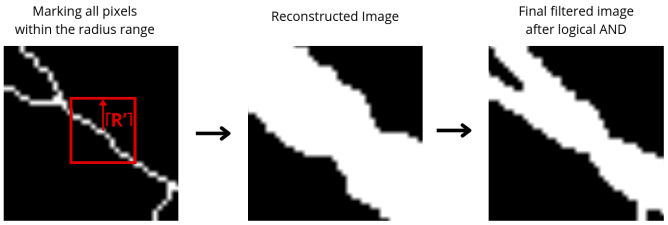

Fig. 2. Scheme of Vessels Thickness Filtering Process

The result of this process, illustrated in Fig. 3, is a filtered vessel mask containing only vessels within the desired thickness range. We note that, due to the pruning operations previously applied to the skeleton, certain vessel pixels may have been erroneously removed. To mitigate this effect and ensure a more faithful representation, these pruned pixels are reintroduced into the filtered mask.

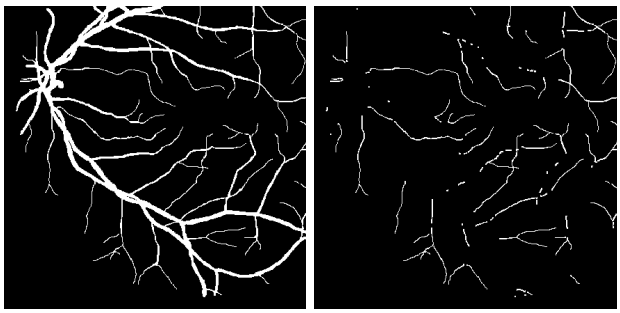

Fig. 3. Zoom in on thin vessel filtering. Original vessel mask (left) and filtered thin (radius=1) vessels only.

For recall computation, the filtering is applied to the ground truth vessel mask. Each foreground pixel (i.e., a white pixel) in the filtered ground truth is compared against the corresponding pixel in the predicted segmentation mask, enabling the identification of true positives and false negatives.

For precision, the filtering process is applied to the predicted segmentation mask. Each white pixel in the filtered prediction is then assessed with respect to the corresponding pixel in the ground truth mask, allowing the determination of true positives and false positives.

C. Weight mask for BCE loss

The core idea is that the weight assigned to each pixel should be inversely proportional to the thickness of the vessel segment it belongs to, thereby prioritizing the contribution of thinner structures during training. We do so by iterating over all foreground pixels p in the *Distances Skeleton*, which encodes the vessel's radius at each point - as previously discussed - and assigning weights to neighboring pixels within a square neighborhood of side $2\lceil R \rceil$.

More specifically, for a pixel p in the *Distances Skeleton*, let R represent the estimated radius. We propose two distinct weighting formulations: for all pixel i belonging to a square of side $2\lceil R \rceil$ centered on p, (a) weight mask $W0$ assigns a constant weight given by $w_i = \dfrac{2}{R^2}$, and (b) weight mask $W1$, defined as $w_i = \dfrac{D_i + 1}{R^2}$, where D_i denotes the distance of pixel i to the background, provided by the distance transform (as explained in Section II-A), modulates the weight according to the pixel distance to the background.

These formulations aim to increase error penalization as vessel's width decreases, as shown in Fig. 4, encouraging the model to better segment delicate vascular structures.

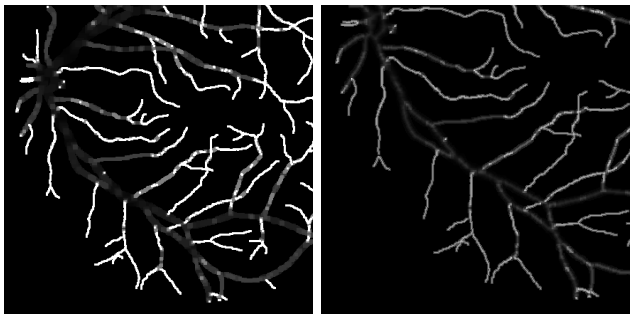

Fig. 4. Zoom in on weight masks. W0 weight mask (left) and W1 weight mask (right). Light gray corresponds to higher weights and dark gray to lower weights.

III. Experimental Results

The dataset employed in the experiments is the publicly available Digital Retinal Images for Vessel Extraction (DRIVE) dataset[1]. It comprises 40 color fundus images, divided into 20 training and 20 testing samples. Each image is accompanied by a manually annotated ground truth segmentation mask, serving as a reference standard for the evaluation of vessel segmentation accuracy.

To implement the U-Net model, train the models, evaluate their performance and import the data, we used the *PyTorch*

[1]https://drive.grand-challenge.org/

TABLE I
PERFORMANCE COMPARISON OF LOSS FUNCTIONS WITH DATA AUGMENTATION. BEST RESULTS USING CLASS-BALANCING WEIGHT MASKS ARE UNDERLINED. BEST RESULTS USING ONE OF THE PROPOSED WEIGHT MASKS ARE IN **BOLD**

Metrics	**Dice**	**Dice+BCE with class-balancing weights**	**Dice+BCE with W0 weight mask**	**Dice+BCE with W1 weight mask**	**BCE with class-balancing weights**	**BCE with W0 weight masks**	**BCE with W1 weight masks**
Accuracy	96.91%	96.26%	**96.92%**	96.84%	94.35%	96.89%	96.78%
Recall	80.39%	88.01%	80.45%	82.12%	94.41%	77.75%	**82.47%**
Precision	84.05%	74.36%	84.10%	82.23%	61.70%	**85.82%**	81.35%
F1-Score	81.89%	80.41%	**81.93%**	81.87%	74.46%	81.30%	81.65%
Recall on Thin Vessels	58.94%	72.90%	61.33%	65.19%	85.15%	59.04%	**68.01%**
Precision on Thin Vessels	67.91%	49.44%	67.35%	63.12%	21.37%	**69.48%**	59.60%
F1-Score on Thin Vessels	62.02%	58.35%	**63.20%**	63.15%	33.81%	62.90%	62.85%

library [5]. Moreover, for the training process, we used the Adam optimizer algorithm, a starting learning rate of 0.001 with PyTorch's *ReduceLROnPlateau* [6] learning rate scheduler and 200 epochs. Furthermore, we used *Numpy* [7], Scikit libraries [8] and an implementation [9] of the discrete skeleton evolution (DSE) algorithm [10] for creating the weight masks and the necessary filtered masks.

A. Comparing different loss functions

We experiment the following variations of loss functions:

- Dice loss
- BCE loss with class-balancing weights [8]
- BCE loss with $W0$ weights
- BCE loss with $W1$ weights
- Dice + BCE loss with class-balancing weights
- Dice + BCE loss with $W0$ weights
- Dice + BCE loss with $W1$ weights

The reasoning for incorporating Dice loss lies in its region-based nature, which captures the global structure of the vessel network. In contrast, BCE loss operates at the pixel level, treating each pixel independently. Since BCE is highly sensitive to class imbalance, whenever we use it, it is used with a pixel-wise weight mask.

Table I shows the results. We partitioned the training set into 5 images for validation and 15 images for training. The training set was augmented to 60 images by performing a horizontal, a vertical, and a composition of both horizontal and vertical flips on each image. Although we also trained the models without data augmentation, due to space limitations, we report only results with data augmentation, which were superior.

Overall, the model trained with Dice loss alone presents good global scores, but struggles with thin vessels. In contrast, the models trained using BCE loss present superior performance on thin vessel recall. However, although the ones with class-balancing weights present top recall scores on thin vessels, their precision is very poor, indicating that simple weighting based on the inverse of class frequency is not sufficient. We highlight that models that used BCE loss with either W0 or W1 weight masks present top f1-scores on thin vessels at the same time that score well with respect to the global metrics.

Thus, we rule out Dice loss alone and BCE with weight-balancing mask, and focus on losses that include either the W0 or W1 masks. Comparing BCE with the proposed weight masks with respect to metrics restricted to thin vessels, W0 achieves higher precision, whereas W1 excels in recall. This characteristic is also observed with respect to the global metrics. We also observe that combining Dice loss with BCE loss tends to balance the model's performance on both the recall and precision metrics on thin vessels, reducing the gap between both in comparison to the models that use BCE loss alone – an expected outcome, given the region-level nature of the Dice loss that emphasizes the preservation of global structures.

Overall, we conclude that using the proposed weight masks is beneficial regarding thin vessel segmentation. Since best recall and best precision on thin vessels are achieved with models that use BCE with either of the masks, examples of the predictions from the models trained with (i) BCE with class-balancing weights, (ii) BCE with W0 weights, and (iii) BCE with W1 weights are shown in Fig. 5.

Prediction seen in Fig. 5 makes it clear that, although the model trained with BCE and class-balancing weights achieved exceptional recall on thin vessels, it compromised the anatomical accuracy of the vessels by producing a large number of false positives. Also, it shows that the model trained using BCE with W1 mask is more permissive in relation to thin vessels segmentation than the one trained using BCE with W0 mask, being consistent with numerical values in Table I.

B. Comparing metrics varying vessels thickness

Up to this point, results and analysis refer to performance assessment on thin vessels (maximum radius of 1) and on the full segmentation mask. For sanity check, we compute the variation of the metrics as we gradually increase the maximum radius value from 1 to 8.5, the smallest multiple of 0.5 greater than the maximum vessel radius observed in the dataset. Figure 6 shows the metrics of the models trained with BCE loss, respectively, with the W1 and W0 weight masks. In both cases, the metrics value increase following an approximately logarithmic trend before converging to the values obtained for the overall metrics, as expected. It clearly shows that W1 emphasizes recall while W0 emphasizes precision.

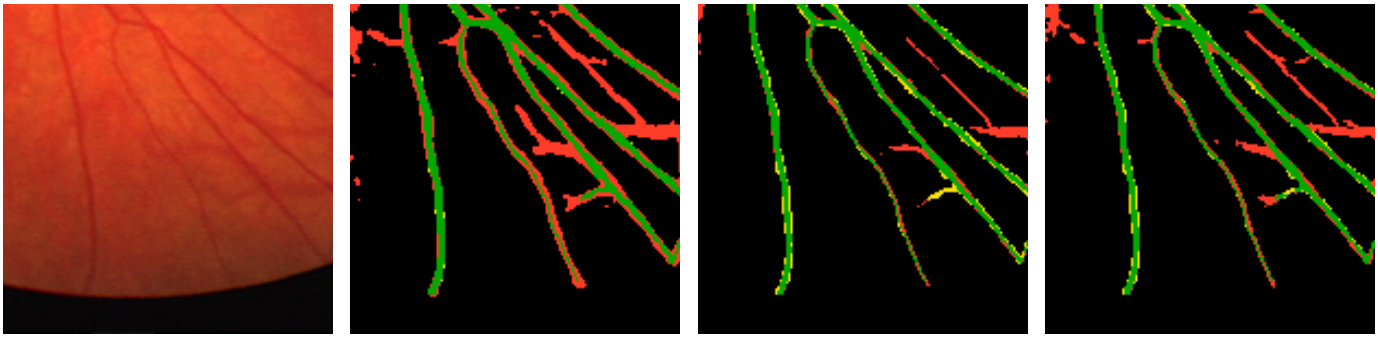

Fig. 5. Zoom in on prediction results. From left to right: original image, BCE with class-balancing weight, BCE with W0 weights, and BCE with W1 weights. Colors: green are true positives, red are false positives and yellow are false negatives.

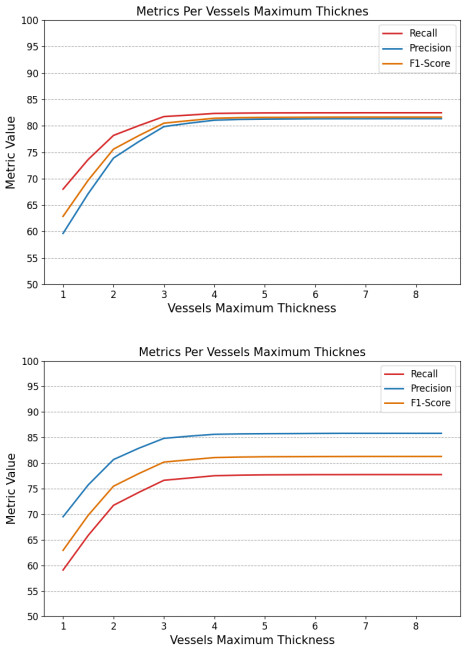

Fig. 6. Performance metrics per vessels maximum thickness using the model trained with BCE loss with W1 (top) and W0 (bottom) weight masks.

IV. Conclusion

In this study, we not only used our proposed metrics for evaluating different models performance on thin vessels segmentation in retinal images, but also introduced vessel-thickness-based weight masks for BCE loss, aimed at improving the U-Net model's performance on this task using the DRIVE dataset. The experimental results provide strong evidence that the proposed masks, combined with data augmentation techniques, enhance the model's ability to segment thin vessels.

The first mask, derived from the W0 formulation, led to models that demonstrated a stronger ability to preserve vessel architecture, as reflected in their relatively high precision on thin vessels. In contrast, models trained with the second mask, based on the W1 formulation, showed greater potential for correctly detecting thin vessels, though at the expense of anatomical fidelity. This is supported by their high recall on thin vessels while still maintaining acceptable but lower precision for these fine structures in comparison to the previous models.

As future work, it would be interesting to investigate new formulas for the mask and evaluate the application of the proposed metrics and weight masks in different fundus images datasets.

Acknowledgment

The authors thank São Paulo Research Foundation (FAPESP) – grants #2024/16942-0 and #2022/15304-4, and MCTI (Ministério da Ciência, Tecnologia e Inovações, Brazil), law 8.248, PPI-Softex - TIC 13 - 01245.010222/2022-44.

References

[1] L. Wei, X. Sun, C. Fan, R. Li, S. Zhou, and H. Yu, "The pathophysiological mechanisms underlying diabetic retinopathy," *Frontiers in Cell and Developmental Biology*, vol. 10, 2022. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fcell.2022.963615

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

[3] C. Chen, J. H. Chuah, R. Ali, and Y. Wang, "Retinal vessel segmentation using deep learning: A review," *IEEE Access*, vol. PP, pp. 1–1, 08 2021.

[4] E. R. Dougherty and R. A. Lotufo, *Hands-on Morphological Image Processing*. SPIE Press, 2003.

[5] PyTorch Contributors, "PyTorch: Tensors and Dynamic Neural Networks in Python with Strong GPU Acceleration," https://github.com/pytorch/pytorch, 2024, accessed: 2025-07-27.

[6] ——, "ReduceLROnPlateau," https://docs.pytorch.org/docs/stable/generated/torch.optim.lr_scheduler.ReduceLROnPlateau.html, 2025, online documentation; accessed 2025-07-27.

[7] NumPy Developers, "NumPy: The fundamental package for scientific computing with Python," https://github.com/numpy/numpy, 2024, accessed: 2025-07-27.

[8] Scikit-learn Developers, "scikit-learn: Machine Learning in Python," https://github.com/scikit-learn/scikit-learn, 2024, accessed: 2025-07-28.

[9] Developers of DSE-Skeleton-Pruning, "DSE," https://github.com/originlake/DSE-skeleton-pruning, 2022, accessed: 2025-07-31.

[10] X. Bai and L. J. Latecki, "Discrete skeleton evolution," in *Proceedings of the 6th International Conference on Energy Minimization Methods in Computer Vision and Pattern Recognition*. Berlin, Heidelberg: Springer-Verlag, 2007, p. 362–374.