Connected-Components-based Post-processing for Retinal Vessels Deep-Learning Segmentation

Idris Dulau
LaBRI UMR 5800
France, Bordeaux Univ.
idris.dulau@u-bordeaux.fr

Benoit Recur
INSERM LAMC U1029
France, Bordeaux Univ.
benoit.recur@gmail.com

Catherine Helmer
INSERM U1219
France, Bordeaux Univ.
catherine.helmer@u-bordeaux.fr

Cecile Delcourt
INSERM U1219
France, Bordeaux Univ.
cecile.delcourt@u-bordeaux.fr

Marie Beurton-Aimar
LaBRI UMR 5800
France, Bordeaux Univ.
marie.beurton@u-bordeaux.fr

***Abstract*—The analysis of fundus images may reflect systemic and cerebral vascular status through a non-invasive, rapid, and cost-effective method. Accurate characterization of the retinal vessels is critical for this status assessment. Medical professionals can perform diagnosis on measurements extracted from the retinal vessels, which are identified through segmentation. Supervised-Learning is used to perform this segmentation task and has been shown to produce higher-quality results compared to traditional methods. However, the Supervised-Learning-based binary method leads to segmentations with multiple Connected Components (CC). Amongst these components, some are disconnected retinal vessels (mentioned as branches), others are artifacts. Artifacts are disconnected miss-classified components resulting from the Supervised-Learning segmentation and that should be removed. Conversely, branches should be kept and further re-connected as they are anatomically supposed to be connected. In this study, we propose a Connected-Components-based post-processing procedure to remove artifacts while preserving the most possible amount of branches. Our methodology involves a relative threshold to cluster the CC based on their areas. We also introduce a useful evaluation metric for the segmentations in the case of measurements extractions on retinal vessels. Over 615 predicted segmentations from six datasets, we improved the dice by a substantial 0.062 leading from 0.782 to 0.844. In conclusion, our method has the potential to significantly enhance the usability and reliability of retinal vessels segmentations, making it a valuable tool for medical professionals in the assessment of systemic and cerebral vascular status. Our work also provides useful insights for future research in this area, especially to address the re-connection of the remaining branches.**

***Index Terms*—Retinal Vessels, Deep-Learning Segmentation, Post-processing, Evaluation Metric, Connected Components**

I. Introduction

The analysis of fundus images may reflect an individual's systemic and cerebral vascular status through a non-invasive, rapid, and cost-effective method. The critical aspect of this status assessment is to accurately characterize the retinal vessels structures in the fundus images. A diagnosis is performed by medical professionals using measurements extracted from the retinal vessels. To extract measurements from the fundus images, the retinal vessels have to be identified. Deep-Learning techniques can achieve this task of dividing an image into parts named segmentation, where a label is assigned to each part based on its visual characteristics. Presently, the fundus images are segmented in two parts, retinal vessels in white, background in black. A segmented image resulting of a Deep-Learning segmentation is called a prediction or a predicted segmentation. Supervised-Learning is a subset of Deep-Learning. It consists of training the Deep-Learning networks with labeled data. Each input data point is associated with a corresponding output label. The goal of Supervised-Learning is to enable the network to make accurate predictions on new unseen data based on the patterns it has learned from the labeled training data. In the case of retinal vessels segmentation, it has been shown to produce higher-quality results [1], [2] compared to traditional methods, i.e. improved levels of details and robustness to contrast variations. Retinal vessels are anatomically supposed to be connected, describing a tree-like structure whose root is the optic disk. However, despite its benefits, the presented Supervised-Learning-based binary method leads to segmentations with multiple Connected Components (CC). Thus, the retinal vessels in the predicted segmentations are not connected as a tree-like structure. The predicted segmentations are composed of a bigger retinal vessels structure, and other smaller components that we mention as disconnected components. Amongst these disconnected components, some are disconnected retinal vessels that we mention as branches (of the supposed tree-like structure), others are artifacts. Artifacts are disconnected miss-classified components resulting from the Supervised-Learning segmentation errors. They are classified as retinal vessels in white, but should be classified as background in black. From those disconnected components, the branches should be kept whereas the artifacts should be removed. Retinal Vessels Post-processing techniques for artifacts removal in segmentations aim at addressing this goal. Post-processing techniques are algorithms applied to the output of predicted segmentations.

979-8-3503-3337-4/23/$31.00 ©2023 IEEE

Algorithms are either traditional methods or Deep-Learning networks. This latter case can be defined as a Network Followed Network (NFN) architecture. The network that outputs the predicted segmentation is followed by another network that applies post-processing on it. As we need a tree-like structure to extract measurements from the retinal vessels predicted segmentations, without post-processing we can only use the image part previously referred as *bigger retinal vessels structure*. Segmentations containing only the biggest retinal vessels structure (now mentioned as $CC = 1$) have less information than the corresponding originally predicted segmentation (now mentioned as $pred$). Performing measurements on $CC = 1$ is necessary as it can't be done on $pred$, but induces information loss. A way to perform measurements on $pred$ is to remove artifacts and re-connect the remaining disconnected components, i.e. branches. In this study, we present our Connected-Components-based Post-processing procedure to tackle this problem. It aims at removing all the artifacts in $pred$ while preserving the most possible amount of disconnected branches. This is a preliminary but essential work to further address the re-connection of the remaining disconnected branches. Our Connected-Components-based Post-processing methodology involves a relative threshold to cluster the CC based on their number of pixels (now mentioned as CC's areas). It takes $pred$ as input, and outputs the segmentation where artifacts are removed and most branches are preserved. In this study, we also introduced a metric that gives the number of CC in segmentations. It highlights either a connected tree-structure segmentation or the need to re-connect remaining disconnected branches. It thus provides a more accurate quality assessment of segmentations that will be measured for diagnosis. The sections of this paper are structured as follows: In section II, a literature review of related work. In section III, an explanation of our post-processing method and our evaluation metric. In section IV, the results of the proposed post-processing method performed on six retinal vessels binary segmentation datasets. In section V, the conclusion of our paper and potential directions for future research.

II. RELATED WORK

A. Retinal Vessels Supervised-Learning Segmentation

The analysis of fundus images may reflect an individual's systemic and cerebral vascular status through a non-invasive, rapid, and cost-effective method. Accurately characterizing the retinal vessels structures in fundus images is crucial for systemic and cerebral vascular status assessment, as medical professionals use measurements extracted from these vessels to diagnose patients. In order to extract measurements, the vessels must first be identified, which can be achieved using Deep-Learning segmentation, a subset of computer vision and machine learning techniques for dividing an image into parts. Each part of the image is assigned a label (Fig. 1), retinal vessels in white, background in black. A two-part segmentation is called a binary segmentation. A segmented image resulting of a Deep-Learning segmentation task is called a prediction or a predicted segmentation. Supervised-Learning is a Deep-Learning technique that involves training the networks with labeled data, where each input data point is matched with an output label. The objective of Supervised-Learning is to enable the network to make accurate predictions on previously unseen data by learning patterns from the labeled training data. In the case of retinal vessels segmentation, it has been shown to produce higher-quality results [1], [2] compared to traditional methods, i.e. improved levels of details and robustness to contrast variations. The most efficient architectures in this field yield from *U-Net* [3], itself built upon Convolutional Neural Networks (CNN). *U-Net* architecture is commonly named an encoder-decoder. The network consists of a contracting path (downsampling) to capture high-level to low-level features from the input, and a symmetric expanding path (upsampling) that enables precise localization of the features to reconstruct the output segmentation map. *U-Net* also incorporates skip connections between the down and up sampling paths to combine features from multiple scales. Since then, new architectures for retinal vessels segmentation came up with modifications, whether on encoder as *PCAT-UNet* [6] with the *PCAT blocks*; on decoder as *DR-VNet* [7] with *Squeeze*, *Excitation* and *Dense-net blocks*; on bottleneck as *SA-unet* [8] with Spatial Attention module; or skip connections as *r2u-net* [9] with residual connections.

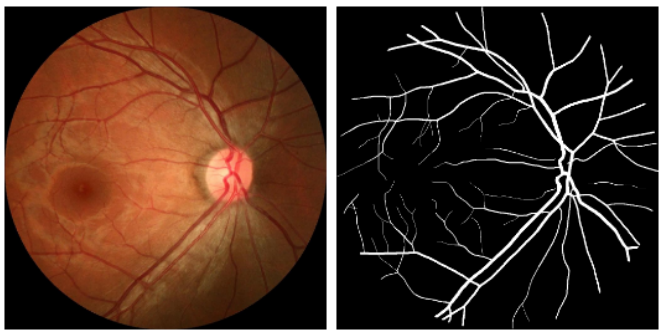

Fig. 1. Fundus image (left), Retinal vessels segmentation (right) [10]

B. Retinal Vessels Post-processing for artifacts removal

Retinal vessels are supposed to have a tree-like structure whose root is the optic disk, but Supervised-Learning-based binary method leads to segmentations with multiple Connected Components (CC). Presently, each predicted segmentation is composed of a bigger retinal vessels structure and other smaller components that we mention as disconnected components. Amongst these disconnected components, some are disconnected retinal vessels that we mention as branches (of the supposed tree-like structure), others are artifacts. Artifacts are disconnected miss-classified components resulting from the Supervised-Learning segmentation. They are classified as retinal vessels in white, but should be classified as background in black. From those disconnected components, the branches should be kept and the artifacts removed. Retinal Vessels Post-processing techniques for artifacts removal in segmentations

aim at addressing this goal either with traditional methods or Deep-Learning networks. This latter case can be defined as a Network-Followed-Network (NFN) architecture. The network that outputs the predicted segmentation is followed by another network that applies post-processing on it. In the field of retinal vessels post-processing, few techniques have been presented due to a predominant community research emphasis on network architectures for segmentation. Most of the time, these post-processing techniques are presented as a briefly outlined component towards the end of a network architecture article. Existing post-processing techniques applied to retinal vessels segmentations can be categorized as follows: i) Morphological operations, such as dilation, erosion (opening, closing) [11], and skeleton [12] (combined with convex hull) [13]; ii) Thresholds, such as absolute threshold on connected components' areas [14], or absolute threshold on connected components' elongatedness [15], or feature maps Otsu treshold [16], or edge-based Hysteresis threshold [17]; iii) Deep-Learning Network-Followed-Network, with Conditional Random Fields [18], or with encoder-decoder [19].

C. Retinal Vessels Evaluation Metrics

Evaluation metrics are quantitative measures used to assess the performances of a statistical model for a specific task. The evaluation metrics typically include four basic numbers that define the confusion matrix (CM): True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN). The CM is obtained by a pixel comparison between a predicted segmentation and its associated groundtruth. For retinal vessels, groundtruth is a manually annotated segmentation performed by a medical specialist. In the CM, Positive refer to retinal vessels pixels in white and negative refer to background pixels in black. True refers to well-classified pixels and False refers to miss-classified pixels. Thus, TP is the number of well-classified white pixels. Based on the CM, several evaluation metrics can be used to summarize the overall performances of a model, such as: Precision ($\mathcal{P}$), Recall ($\mathcal{R}$), F1-score and Dice Similarity Coefficient ($\mathcal{D}$) (N.B. F1-score and Dice Similarity Coefficient are numerically equal in the binary case). Each of these metrics, Table I, have in common that they don't use the TN values as part of the quality evaluation. This particularity leads to a more robust evaluation of the segmentations qualities in our case of class imbalance where the retinal vessels represent only 10% of the groundtruth segmentation pixels. Furthermore, F1-score and Dice Similarity Coefficient are particularly interesting as they balanced Precision and Recall. Other evaluation metrics for assessing retinal vessels segmentation performances are also used such as: Accuracy, Balanced Accuracy, Specificity and Area Under the receiver operating characteristic Curve (AUC). These metrics, conversely to Precision, Recall and Dice Similarity Coefficient, compute TN as part of their evaluation which we consider an error (e.g. segmentation performed on a highly padded image will result in better performances as the added padding is always well-classified as a TN value). Our results won't mention these latter metrics.

TABLE I
PRECISION, RECALL & DICE FORMULAS

Precision	Recall	Dice
$\mathcal{P} = \frac{TP}{TP+FP}$	$\mathcal{R} = \frac{TP}{TP+FN}$	$\mathcal{D} = \frac{2*TP}{2*TP+FP+FN}$

III. METHODOLOGY

A. Retinal Vessels Datasets

In order to assess our method's generalizability, we evaluated performances on six public datasets, which collectively comprise 205 image pairs. Each pair contain an RGB fundus image and its binary manual annotation of retinal vessels. Fig. 2 shows fundus images of the datasets, from top to bottom and from left to right: *CHASEDB1* [20], *DRHAGIS* [21], *DRIVE* [22], *HRF* [23], *IOSTAR* [24] and *LESAV* [25]. These fundus images have been captured using different materials (e.g. Nidek NM-200-D, Topcon TRC-NW6s, Canon CR5 non-mydriatic with 3CCD sensor, ...), from different angles (e.g. optic disk centered, macula centered), with different resolutions (e.g. 999*960, 4752*3168, 565*584, ...), and comprise various pathological signs (e.g. glaucoma, drusen, exudate, ...).

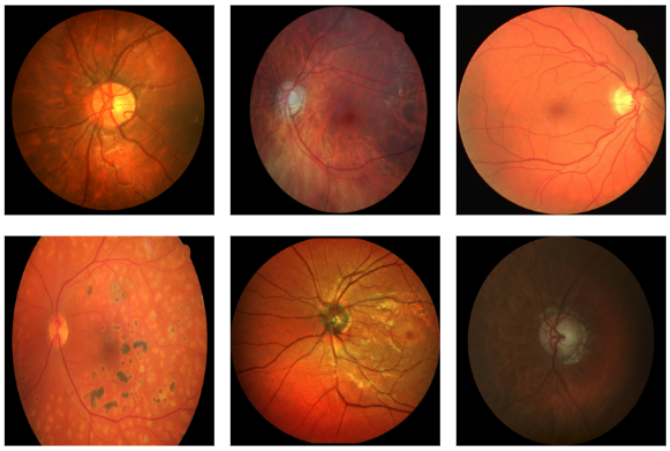

Fig. 2. CHASEDB1, DRHAGIS, DRIVE, HRF, IOSTAR, LESAV

B. Data Preparation

Post-processing techniques are algorithms applied to the output of predicted segmentations. We first need to obtain these predicted segmentations. We perform an independent Supervised-Learning network training for each of the six datasets. As shown in Fig. 3, input training data are the RGB fundus and binary groundtruth resized to 1024*1024 pixels. To reflect the state of the art [1], [2], our training architecture takes *U-Net* [3] as a baseline, where we incorporated modifications on encoder and decoder parts: i) Convolutions in the encoder are performed with padding; ii) Batch normalization [4] is applied before ReLU activation; iii) We minimize a Dice Loss [5] instead of a Cross Entropy Loss, as we are dealing with imbalanced classes (10% of white pixels for 90% of black pixels). Output data are the predicted segmentations

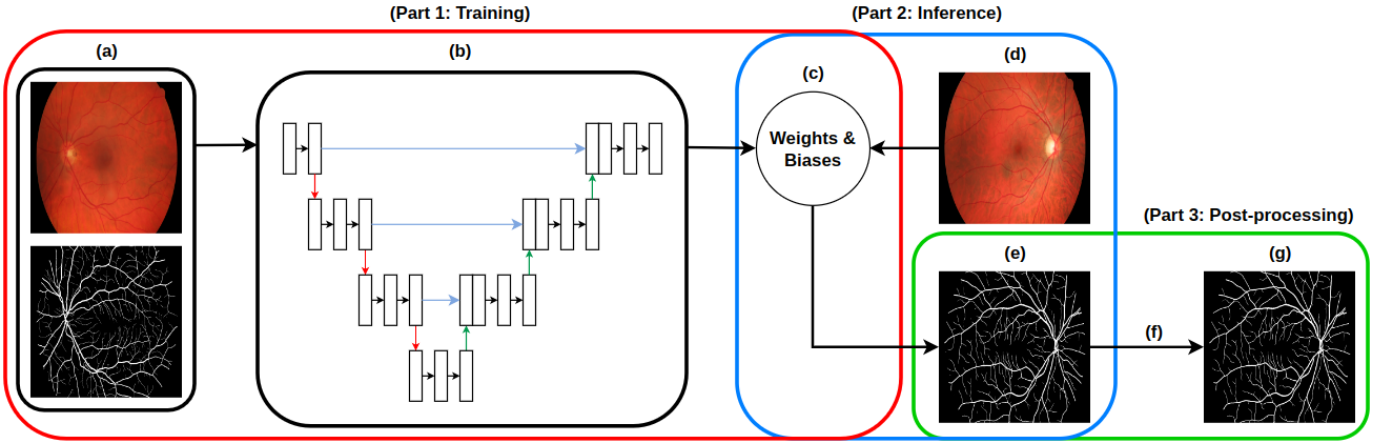

Fig. 3. Data processing pipeline illustration from training inputs to post-processing outputs: (a) Training fundus and associated retinal vessels groundtruth pairs, (b) Neural Network, (c) Generated weights and biases, (d) Fundus to segment, (e) Retinal vessels predicted segmentations, (f) Post-processing procedure, (g) Retinal vessels post-processed segmentations

of size 1024*1024 pixels. We will not compare to others Deep-Learning segmentation methods. As mentioned earlier, we do not want to highlight our predicted segmentations results, but we want to highlight our post-processing method results applied on these predicted segmentations. As mentioned earlier, predicted segmentations are characterized by a bigger retinal vessels structure and other smaller components. A repeated training procedure with the exact same parameters leads to variations in results. The bigger retinal vessels structure is mainly still but the smaller components positioning is more different. We thus repeated three times each training on the six datasets to obtain a variability in the predictions. This leads to eighteen sets of predicted segmentations (615 images) for which inference takes in mean 0.5s per image. Each of the eighteen sets of predictions has a mean Dice over 0.80 and up to 0.88 which match or improve state of the art performances. It thus ensures that: i) the artifacts removal procedure will be able to improve current and future high quality predicted segmentations; ii) the predicted segmentations reflect most of the spatial information of a tree-like structure.

C. Artifacts removal procedure

To extract measurements from predicted retinal vessels segmentations, a tree-like structure is required. Without post-processing, we can only use the image part previously referred as *bigger retinal vessels structure*. However, such segmentation (denoted as $CC = 1$) provides less information than the originally predicted segmentations (denoted as $pred$). Measurements must be taken from $CC = 1$ due to the impossibility of doing so on $pred$, but this induces information loss. To overcome this limitation, artifacts must be eliminated, and branches must be reconnected. In this section, we present our Connected-Components-based Post-processing procedure to tackle this problem. Our Connected-Components-based Post-processing is related to the threshold post-processing family mentioned in section II-B. It involves a relative threshold to cluster the CC based on their number of pixels (now mentioned as CC's areas). It takes $pred$ as input, and outputs the segmentation where artifacts are removed and most branches are preserved. It aims at removing all the artifacts in $pred$ while preserving the most possible amount of disconnected branches. This is a preliminary but essential work to further address the re-connection of the remaining disconnected branches. The procedure can be defined in four steps. First, on the Deep-Learning predicted segmentations we find and store every CC's areas in an 8-connected pixel neighborhood by performing a Connected-Component-Labeling (CCL) algorithm [26]. Second, we sort the CC's areas by decreasing order. Third, we define an area-related threshold of *"X times the biggest CC's area"*. Fourth, we erase each CC whose areas are under the threshold. Performances of this technique rely on threshold definition, which can be chosen through the CC's areas distribution understanding. This CC's areas distribution sorted by decreasing order can be seen in three parts (even if there is no distinct threshold value that can cluster CC's areas in three distinct parts). These three parts are: i) High areas, exclusively disconnected branches; ii) Low areas, overwhelmingly artifacts and few very small parts of disconnected branches; iii) Between high and low areas, no real possibility to determine whether it's artifacts or disconnected branches. Understanding that, threshold settings can lead to three different behaviours: i) High areas threshold, to erase all artifacts and exclusively keep disconnected branches (but not all); ii) Low areas threshold, to keep all disconnected branches but also some artifacts (but not all); iii) Between high and low areas threshold, to perform an operation with an undetermined and thus not valuable behaviour. As mentioned earlier, the procedure aims at removing all the artifacts in $pred$ while preserving the most possible amount of disconnected branches. This is a preliminary but essential work to further address the re-connection of the remaining disconnected branches. This goal can be fulfilled using our Connected-Components-based Post-processing procedure and setting a threshold in the high CC's areas. Results are discussed in section IV.

D. Evaluation Metric

In this section, we introduced a metric to provide a more accurate assessment of retinal vessels segmentations. The metric will give the number of CC in segmentations. If CC=1, the retinal vessels in the segmentations can be measured. Else, there are still post-processing procedures to perform. First, our Connected-Components-based Post-processing procedure to remove artifacts. Second, a branches re-connection procedure. The metric highlights either a connected tree-structure segmentation or the need to re-connect remaining disconnected branches. It thus provides a more accurate quality assessment of retinal vessels segmentations for measurements extractions.

IV. RESULTS

As mentioned in section III-C, retinal vessels in predicted segmentations (now mentioned as $pred$) have to be connected to extract measurements. Else, without post-processing we can only use the image part containing only the biggest retinal vessels structure (now mentioned as $CC = 1$). This latter has less information than $pred$. Performing measurements on $CC = 1$ is necessary as it can't be done on $pred$, but induces information loss. A way to perform measurements on $pred$ is to remove artifacts and re-connect the remaining disconnected components, i.e. branches. In this section, we present the performances of our Connected-Components-based Post-processing procedure that tackles the artifact removal problem. It aims at removing all the artifacts in $pred$ while preserving the most possible amount of disconnected branches. As mentioned in section III-A six datasets were used to generate predicted segmentations. As mentioned in section III-B we trained each network thrice leading to three different sets of predicted segmentations. Results will thus be assessed on eighteen different sets of predicted segmentations for a total of 615 images. Mean performances of the three sets of predictions for each dataset are shown in Table II. In each row, we compared $pred$, $CC = 1$ and $BestDice$. This latter illustrates the procedure applied on known data. In our experimentations, for every predicted segmentation of every dataset, we compared $pred$ to its groundtruth, then we recursively removed its smallest CC and compared it again until it is equal to $CC = 1$ in order to explore the Dice values of every case. From there, we know the best Dice value for each set of predictions. We also know the other associated metrics values as performances are assessed using TP, KTP(%), Dice, CC & THLD. THLD refers to the CC's area-related threshold mentioned in section III-C. KTP(%) defines the amount of Kept True Positives. It highlights the amount of branches kept by the procedure. The difference of TP between $pred$ and $CC = 1$ corresponds to the maximum number of re-connectable pixels. The more TP remain at the end of the procedure, the more branches are kept. The kept TP are the sum of the TP in the kept CC. The number of kept CC is selected using the maximum Dice through the exploration the Dice values we mentioned earlier. Table II first row can thus be read as: "In average for the dataset *CHASEBD1*, the procedure kept 96% of $pred$'s TP. It increased the Dice score of both $CC = 1$ and $pred$. It is done by keeping 20 CC out of 48. It corresponds to keep the CC whose areas are over 0.00094*74541 (70 pixels). This avoid only increasing TP while both losing precision, i.e. keeping artifacts, and losing recall, i.e. removing re-connectable branches. Precision alone is not sufficient to assess our performances. Branches are also composed of FP as their diameters are not exactly the same as their groundtruth. As smaller disconnected components have proportionally much more FP than bigger one, precision is necessarily decreasing when the number of kept CC is increasing. Additionally to the mean performances per dataset in Table II, we summed up the results in Table III. We also provided the mean number of removed artifacts, the gain of Dice generated by the procedure and the maximum obtained THLD. These results are discussed in section IV-A.

TABLE II
MEAN PERFORMANCES ON EACH DATASET

Datasets		TP	KTP(%)[a]	Dice	CC	THLD[b]
CHASEDB1	$pred$	79210	100	0.872	48	0
	$BestDice$[c]	79079	96	0.873	20	0.00094
	$CC = 1$	74541	0	0.851	1	1
DRHAGIS	$pred$	39516	100	0.809	72	0
	$BestDice$	39367	93	0.812	20	0.0103
	$CC = 1$	37047	0	0.787	1	1
DRIVE	$pred$	82389	100	0.835	87	0
	$BestDice$	82234	94	0.836	31	0.00067
	$CC = 1$	79078	0	0.822	1	1
HRF	$pred$	75101	100	0.839	97	0
	$BestDice$	75026	97	0.84	68	0.00034
	$CC = 1$	71599	0	0.821	1	1
IOSTAR	$pred$	73297	100	0.857	56	0
	$BestDice$	73147	99	0.86	12	0.0077
	$CC = 1$	34871	0	0.553	1	1
LESAV	$pred$	69925	100	0.884	25	0
	$BestDice$	69854	98	0.885	13	0.00164
	$CC = 1$	65229	0	0.858	1	1

[a] Amount of Kept TP
[b] CC's area-related threshold
[c] Procedure applied on known data

A. Discussion

Before analyzing the mean statistics of the eighteen predictions, we will discuss the specificities of the six datasets that lead to these performances. As mentioned in section III-A these datasets images have been captured using different materials, from different angles, with different (original) resolutions and comprise various pathological signs. Their groundtruth annotations also transcript this variety. TP and CC distributions are thus different between datasets. In particular, as shown in Table II, the number of TP in *DRHAGIS* is only half of that in the other datasets. Also, the $CC = 1$ Dice of *IOSTAR* is way below the others Dice in the others $CC = 1$ rows as the dataset did not have annotations in the optic disk zone, leading to groundtruth retinal vessels disconnected in two main parts. We can consider three levels that lead to more or less predicted CC (as the predicted details are more

likely to be disconnected). *HRF* is considered highly detailed and shows a mean of 97 CC in its predictions. Conversely, *LESAV* does not present much retinal vessels in fundus images nor in groundtruth segmentations, thus shows a mean of 25 CC in its predictions. *CHASEDB1*, *DRHAGIS*, *DRIVE* and *IOSTAR* belong to the intermediate category. Mean statistics of the procedure experiments on the eighteen sets of predicted segmentations are showed in Table III. Using the procedure over 615 predicted segmentations from six datasets, we are able to remove 57% of the CC (37/64) while keeping 96% of the TP. The procedure improved the $CC = 1$ mean dice by a substantial 0.069 leading from 0.782 to 0.851 (also being higher than the original $pred$ with no removal). We performed the procedure on known data to demonstrate its performances, but we also willed to find the best threshold value to perform the procedure on unknown data. Considering the strong distribution differences between the datasets, the average THLD value in Table III is not relevant to perform the procedure on unknown data. As the most important idea of the procedure is to remove all the artifacts before keeping the most possible amount of branches, we suggest setting a high THLD value. To illustrate our artifacts removal procedure on new unknown data, we provided in Table IV the results using a THLD of 0.0103. This is the maximum THLD value amongst the $BestDice$. It simulates a sub-optimal but safe threshold.

TABLE III
MEAN STATISTICS OF THE EIGHTEEN PREDICTIONS

	TP	KTP(%)[a]	Dice	CC	THLD[b]
$pred$	69906	100	0.849	64	0
$BestDice$[c]	69784	96	0.851	27	0.00359
$CC = 1$	60394	0	0.782	1	1
Removed CC			-	37	-
Dice gain over $CC = 1$			0.069	-	-
Maximum THLD			-	-	0.0103

[a] Amount of Kept TP
[b] CC's area-related threshold
[c] Procedure applied on known data

Results in Table IV show the procedure execution using a THLD of 0.0103, i.e. the maximum THLD value amongst the best predictions, simulating a sub-optimal but safe threshold. Over 615 predicted segmentations from six datasets, we are removing 95% of the CC (61/64) while keeping 56% of the TP. The procedure improved the $CC = 1$ mean dice by a substantial 0.062 leading from 0.782 to 0.844. Also, the procedure is fast as it takes only 0.06s per image. You may have noticed that calculating the mean KTP based on the mean TP in Table IV leads to different results (96% $\rightarrow$ 98% & 56% $\rightarrow$ 82%). We have chosen to take the mean of each KTP computed for each dataset to give fairer results. This demonstration of our artifacts removal procedure on unknown data lead to good results considering we still are able to keep 56% of the TP, but obviously lower results compared to that using known data. These differences show the strong importance of the knowledge of the data distribution. To enhance the results on new unknown data, we encourage annotating a small number of images to get a knowledge of the data distribution, to set THLD in a more favorable way.

TABLE IV
PROCEDURE SIMULATION ON NEW UNKNOWN DATA

	TP	KTP(%)[a]	Dice	CC	THLD[b]
$pred$	69906	100	0.849	64	0
$BestDice$[c]	69784	96	0.851	27	0.00359
THLD[b] $>= 0.0103$	68245	56	0.844	3	0.38938
$CC = 1$	60394	0	0.782	1	1

[a] Amount of Kept TP
[b] CC's area-related threshold
[c] Procedure applied on known data

V. CONCLUSION

In this study, we propose a Connected-Components-based post-processing procedure to remove artifacts while preserving the most possible amount of disconnected branches. Our methodology involves a relative threshold to cluster CC based on their areas. We also introduce a useful evaluation metric for the segmentations in the case of measurements extractions on retinal vessels. We performed the procedure over 615 predicted segmentations from six datasets. The procedure removed on average 57% of the CC (37/64) and kept 96% of the TP. The procedure improved the $CC = 1$ mean dice by a substantial 0.069 leading from 0.782 to 0.851 (also being higher than the original predictions mean dice with no removal). We demonstrated the procedure efficiency on unknown data by ignoring our data distribution knowledge to set a threshold value. It led to good results considering that we were able to keep 56% of the TP and to improve the $CC = 1$ mean dice by a substantial 0.062 leading from 0.782 to 0.844. The results on unknown data are lower compared to that using known data but we have to keep in mind that we set a very high threshold value i.e. the maximum threshold value amongst the best predictions: 0.0103. This threshold has removed 95% of the CC (61/64). These differences show the strong importance of the knowledge of the data distribution. To enhance the results on new unknown data, we encourage annotating a small number of images to get a knowledge of the data distribution, thus being able to set a threshold in a more favorable way. Finally, the procedure takes the predicted segmentations as input and produces outputs where artifacts are removed and most disconnected branches are preserved. Our new evaluation metric provides a more accurate assessment of retinal vessels segmentations for measurements by highlighting their number of CC and thus showing information on remaining artifacts or disconnected branches that the dice score alone can't reflect. In conclusion, our method can significantly enhance the usability and reliability of retinal vessels segmentations, making it a valuable tool for medical professionals in the assessment of systemic and cerebral vascular status. Our work also provides valuable insights for future research in this area, especially to address the re-connection of the remaining branches.

REFERENCES

[1] Khandouzi, A., Ariafar, A., Mashayekhpour, Z., Pazira, M., & Baleghi, Y. (2022). Retinal vessel segmentation, a review of classic and deep methods. Annals of Biomedical Engineering, 50(10), 1292-1314.

[2] Chen, C., Chuah, J. H., Ali, R., & Wang, Y. (2021). Retinal vessel segmentation using deep learning: a review. IEEE Access, 9, 111985-112004.

[3] Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18 (pp. 234-241). Springer International Publishing.

[4] Ioffe, S., & Szegedy, C. (2015, June). Batch normalization: Accelerating deep network training by reducing internal covariate shift. In International conference on machine learning (pp. 448-456). pmlr.

[5] Milletari, F., Navab, N., & Ahmadi, S. A. (2016, October). V-net: Fully convolutional neural networks for volumetric medical image segmentation. In 2016 fourth international conference on 3D vision (3DV) (pp. 565-571). Ieee.

[6] Chen, D., Yang, W., Wang, L., Tan, S., Lin, J., & Bu, W. (2022). PCAT-UNet: UNet-like network fused convolution and transformer for retinal vessel segmentation. PloS one, 17(1), e0262689.

[7] Karaali, A., Dahyot, R., & Sexton, D. J. (2022, June). DR-VNet: retinal vessel segmentation via dense residual UNet. In Pattern Recognition and Artificial Intelligence: Third International Conference, ICPRAI 2022, Paris, France, June 1–3, 2022, Proceedings, Part I (pp. 198-210). Cham: Springer International Publishing.

[8] Guo, C., Szemenyei, M., Yi, Y., Wang, W., Chen, B., & Fan, C. (2021, January). Sa-unet: Spatial attention u-net for retinal vessel segmentation. In 2020 25th international conference on pattern recognition (ICPR) (pp. 1236-1242). IEEE.

[9] Alom, M. Z., Hasan, M., Yakopcic, C., Taha, T. M., & Asari, V. K. (2018). Recurrent residual convolutional neural network based on u-net (r2u-net) for medical image segmentation. arXiv preprint arXiv:1802.06955.

[10] Zhang, S., Zheng, R., Luo, Y., Wang, X., Mao, J., Roberts, C. J., & Sun, M. (2019). Simultaneous arteriole and venule segmentation of dual-modal fundus images using a multi-task cascade network. IEEE Access, 7, 57561-57573.

[11] Thangaraj, S., Periyasamy, V., & Balaji, R. (2018). Retinal vessel segmentation using neural network. IET Image Processing, 12(5), 669-678.

[12] Wang, X., & Jiang, X. (2018, August). Post-processing for retinal vessel detection. In Tenth International Conference on Digital Image Processing (ICDIP 2018) (Vol. 10806, pp. 1442-1446). SPIE.

[13] Wang, X., Jiang, X., & Ren, J. (2019). Blood vessel segmentation from fundus image by a cascade classification framework. Pattern Recognition, 88, 331-341.

[14] Zhang, J., Cui, Y., Jiang, W., & Wang, L. (2015). Blood vessel segmentation of retinal images based on neural network. In Image and Graphics: 8th International Conference, ICIG 2015, Tianjin, China, August 13-16, 2015, Proceedings, Part II 8 (pp. 11-17). Springer International Publishing.

[15] Oliveira, W. S., Teixeira, J. V., Ren, T. I., Cavalcanti, G. D., & Sijbers, J. (2016). Unsupervised retinal vessel segmentation using combined filters. PloS one, 11(2), e0149943.

[16] Jiang, Z., Zhang, H., Wang, Y., & Ko, S. B. (2018). Retinal blood vessel segmentation using fully convolutional network with transfer learning. Computerized Medical Imaging and Graphics, 68, 1-15.

[17] Soomro, T. A., Afifi, A. J., Gao, J., Hellwich, O., Khan, M. A., Paul, M., & Zheng, L. (2017, November). Boosting sensitivity of a retinal vessel segmentation algorithm with convolutional neural network. In 2017 International Conference on Digital Image Computing: Techniques and Applications (DICTA) (pp. 1-8). IEEE.

[18] Fu, H., Xu, Y., Wong, D. W. K., & Liu, J. (2016, April). Retinal vessel segmentation via deep learning network and fully-connected conditional random fields. In 2016 IEEE 13th international symposium on biomedical imaging (ISBI) (pp. 698-701). IEEE.

[19] Wu, Y., Xia, Y., Song, Y., Zhang, Y., & Cai, W. (2020). NFN+: A novel network followed network for retinal vessel segmentation. Neural Networks, 126, 153-162.

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

[21] Holm, S., Russell, G., Nourrit, V., & McLoughlin, N. (2017). DR HAGIS—a fundus image database for the automatic extraction of retinal surface vessels from diabetic patients. Journal of Medical Imaging, 4(1), 014503-014503.

[22] Staal, J., Abràmoff, M. D., Niemeijer, M., Viergever, M. A., & Van Ginneken, B. (2004). Ridge-based vessel segmentation in color images of the retina. IEEE transactions on medical imaging, 23(4), 501-509.

[23] Odstrcilik, J., Kolar, R., Budai, A., Hornegger, J., Jan, J., Gazarek, J., ... & Angelopoulou, E. (2013). Retinal vessel segmentation by improved matched filtering: evaluation on a new high-resolution fundus image database. IET Image Processing, 7(4), 373-383.

[24] Zhang, J., Dashtbozorg, B., Bekkers, E., Pluim, J. P., Duits, R., & ter Haar Romeny, B. M. (2016). Robust retinal vessel segmentation via locally adaptive derivative frames in orientation scores. IEEE transactions on medical imaging, 35(12), 2631-2644.

[25] Orlando, J. I., Barbosa Breda, J., Van Keer, K., Blaschko, M. B., Blanco, P. J., & Bulant, C. A. (2018). Towards a glaucoma risk index based on simulated hemodynamics from fundus images. In Medical Image Computing and Computer Assisted Intervention–MICCAI 2018: 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part II 11 (pp. 65-73). Springer International Publishing.

[26] He, L., Chao, Y., Suzuki, K., & Wu, K. (2009). Fast connected-component labeling. Pattern recognition, 42(9), 1977-1987.