

VAVnets: retinal vasculature segmentation in few-shot scenarios

No Author Given

No Institute Given

Abstract. The structure of the retinal vasculature can indicate various health issues. Quantitatively measuring changes in retinal arteries and veins offers significant potential for disease prevention and management. We propose VAVnets: three variants of a deep-learning network to generate Vessels, Arteries and Veins binary segmentations. Training is conducted in a few-shot and cross-dataset manner using images from four different fundus datasets, collectively comprising 137 images: DRIVE [1], DUMO [2], HRF [3], and LESAV [4]. Training occurs with 3-fold cross-validation. Each fold employs a cross-dataset-build of 40 images (10 from each dataset), with testing on the remaining 97 images. In this article, we discuss our experiments involving architectural choices, transfer-learning, and data augmentation. We assess performances using the dice score as we aim to achieve the best possible pixel-wise segmentations. Our dice scores for each dataset are, for vessels: 0.81, 0.83, 0.81, 0.86; for veins: 0.78, 0.81, 0.78, 0.79; and for arteries: 0.73, 0.78, 0.74, 0.77. To the best of our knowledge, VAVnets demonstrate superior performances compared to existing few-shot methods across these datasets.

Keywords: Deep-learning · Segmentation · Retinal vasculature

1 Introduction

Changes in retinal vascular morphology are associated with various ocular and systemic diseases. Quantitative measurements of retinal arteries and veins could thus serve as biomarkers for managing such diseases. Performing measurements requires a segmentation step to differentiate the retinal vasculature from other fundus structures. However, the scarcity of available data for arteries and veins makes it challenging to train accurate segmentation networks. We propose VAVnets: three variants of a deep-learning network to generate vessels, arteries and veins binary segmentations. To train and test the networks, we use images from four different fundus datasets that are available with the retinal arteries and veins segmentations: DRIVE [1], DUMO [2], HRF [3] and LESAV [4]. Having access to a total of solely 137 images, we perform the training using a cross-dataset-build of 40 images, 10 from each dataset (which is a few-shot scenario), and the testing on the remaining 97 images. Training occurs with 3-fold cross-validation to compensate for this small amount of data. In Section 2, we

present the datasets and summarize their characteristics. In Section 3, we survey the state-of-the-art few-shot retinal vasculature segmentation of vessels, veins, and arteries. In Section 4, we discuss our experiments involving architectural choices, transfer-learning, and data augmentation. In Section ??, we analyze our obtained results and compare them to the state of the art. Finally, in Section 7, we outline future applications of the methods in the field.

2 Datasets

Datasets of fundus images with manual segmentation of veins and arteries are scarce due to the labor-intensive and costly nature of the annotation process, which requires specialized expertise. Ethical and privacy concerns limit data sharing, while variability in image quality and lack of standardized protocols further complicate the creation of comprehensive datasets. Additionally, the technical challenges in accurately distinguishing between veins and arteries contribute to the scarcity of such annotated datasets. Still, some datasets including veins and arteries manual segmentation are publicly available: *DRIVE* [1], *DUMO* [2], *HRF* [3] and *LESAV* [4]. Fig. 1 shows a sample of each dataset.

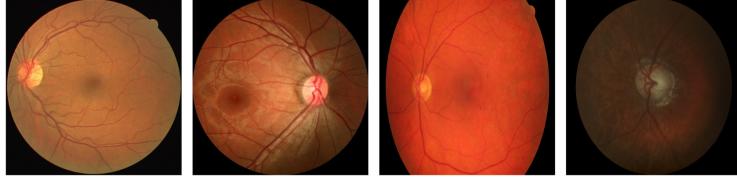


Fig. 1. Samples of the four datasets used in this study.
From left to right: *DRIVE* [1], *DUMO* [2], *HRF* [3], *LESAV* [4].

The datasets characteristics are sum-up in Tab. 1. Collectively, the four datasets comprise 137 fundus images. These fundus images have been captured using different materials (row. 1), with different resolutions (row. 2), and comprise various pathological signs such as: hemorrhages, glaucoma and exudates.

	DRIVE	DUMO	HRF	LESAV
Material	Canon CR5	OT-110M	Canon CR-1	NA
Resolution	565*584	1024*1024	3504*2336	1620*1444
Images	40	30	45	22
Pathologies	7	NA	30	11
Focus	macula	macula	macula	optic disc

Table 1. Datasets characteristics

3 Related works

Numerous methods have been developed for retinal vasculature segmentation in fundus images. For this task, deep-learning-based methods produce the best results [5], and the most effective methods are built by modifying the U-Net [6] architecture. However, few have addressed the challenge of retinal vasculature segmentation in few-shot scenarios. This gap has recently been explored in seven articles, five for vessels segmentation, and two for veins and arteries segmentation. All these methods, published between 2022 and 2023, underscore the growing interest in tackling real-world application problems within this field.

For vessels segmentation, Xu Jianguo *et al.* in [7] propose architectural modifications to U-Net, focusing primarily on enhancing the skip-connections. Similarly, Hao-Chiang Shao *et al.* in [8] also make architectural changes but concentrate mainly on the loss function. Junyan Lyu *et al.* in [9] introduce an automatic data augmentation method. Hu Dewei *et al.* in two works [10, 11] employ prior knowledge of the vasculature to modify U-Net, emphasizing the "tubular" shapes of vessels.

In veins and arteries segmentation, Shi Danli *et al.* in [12] leverage 220 fundus images from an in-house dataset to perform cross-training with public datasets. Notably, in contrast to most related work (not limited to few-shot), they generate veins and arteries as two separate binary outputs. A year later, in [13], they adopted a different approach which implements a multi-class output strategy, relying heavily on transfer-learning from fundus fluorescein angiography images.

4 Methods

The primary objective of the proposed VAVnets is to perform the best pixel-wise segmentations of vessels, veins and arteries in few-shot scenarios. To do so, we investigate various aspects of the deep-learning training processes, including architecture, transfer-learning, and data-augmentation.

4.1 Architecture

As mentioned in Sec. 3, the most effective retinal vasculature segmentation methods are built by modifying the U-Net [6] architecture. Thus, the proposed VAVNets perform in a similar fashion, as illustrated in Fig. 2. In addition to the original design, we first include new blocks into the network such as: (i) attention, (ii) normalization, and (iii) dropout; we adjusted: (iv) the feature map sizes and depths of the networks; and we replaced U-Net building blocks by: (v) others blocks that perform similar functions but in different ways. Attention [14] (i) enable more focus on relevant parts of the input data. Normalization [15] (ii) ensure consistent feature distribution across different inputs. Dropout [16] (iii) reduce reliance on specific features. Adjustment of feature maps size and depth

[17] (iv) enable the networks to capture more intricate features and patterns in the down-sampled input. Other blocks (v) and related experiments are discussed in Sec. 5.

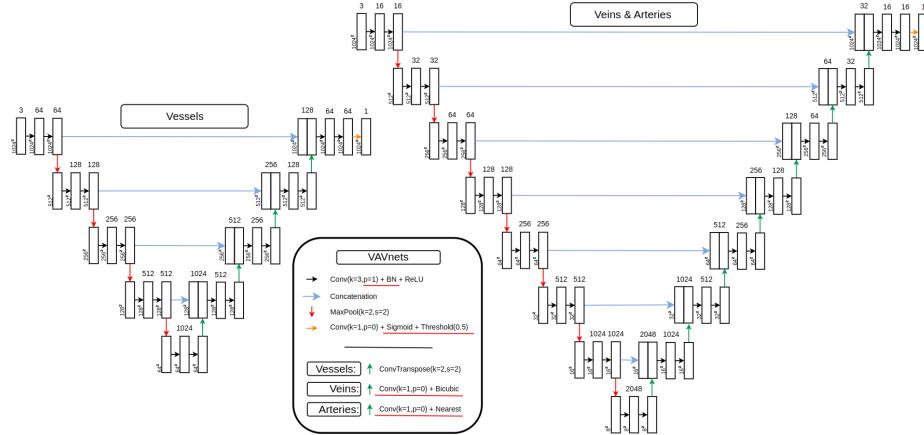


Fig. 2. VAVnets configurations: vessels (left), veins and arteries (right).

The proposed **VAVnets**, share common characteristics between themselves as illustrated in Fig. 2. Batch-normalization layers are added after each convolution and before each ReLU. Feature maps are concatenated between up and down paths Maximum-pooling layers are used for down-sampling. Convolution followed by a sigmoid and a threshold is used to generate a single class output. Vessels, arteries and veins networks have also distinct characteristics between themselves in term of depth, feature map size and up-sampling. The vessels segmentation network is built with transpose-convolution for up-sampling, and a depth of five with feature maps size ranging from 64 to 1024 as in the U-Net. The veins network performs the up-sampling with a convolution followed by a bicubic interpolation, while the arteries network up-sample with a convolution followed by a nearest neighbors interpolation. Both share a depth of eight and a feature map size adjusted from 16 to 2048. Some modifications of the U-Net [6] leading to VAVnets are underlined in red in Fig. 2. First, padding is applied to the convolutions to maintain the size of the convolved input across consecutive layers, eliminating the need for cropping before concatenating skip connections. Also, padding in par with changes in the input size now being 1024*1024 enable to adjust the feature maps size and to go deeper in the architecture. Batch-normalization added after each convolution lead to interesting findings for few-shot scenarios (see in Sec. 5). The output is a single channel processed through a sigmoid function and then thresholded to produce the binary segmentations. Finally, the up-sampling operators are modified for arteries and veins networks.

4.2 Transfer-learning

To cope with a lack of data, we experiment transfer-learning with additional fine-tuning using several publicly available datasets as well as pre-trained networks. We define various configurations utilizing: common objects, curvilinear structures, and retinal vessels from other datasets. Fig. 3 shows samples of the data involved in the transfer-learning experiments.

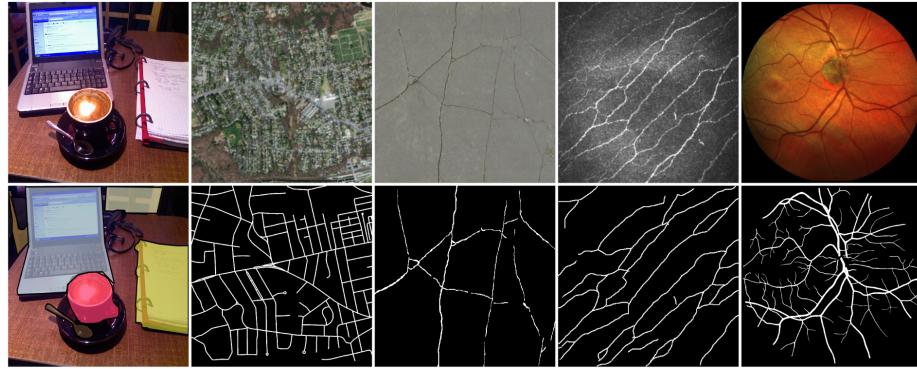


Fig. 3. Samples of the data involved in the transfer-learning experiments.
From left to right: common objects [18], roads [19], cracks [20], corneal nerves [21],
and retinal vessels [22,23,24].

First, fine-tuning a network pre-trained on a huge dataset of common objects enable to extract consequent amount of features. By leveraging ResNet-50 [25] abilities to extract rich features from COCO [18] images, we benefit from fundamental low-level features like edges, textures, and colors. Then, curvilinear structures such as corneal nerves [21], wall cracks [20] and aerial roads [19] are good candidates to benefit from more targeted features like tortuosity and connectivity. Up to 10k images can be collected, making these smaller yet sizable datasets but some more targeted choices of external knowledge. Finally, retinal vessels from other datasets can serve to fine-tune our networks. The 1.5K images collected from public datasets [22,23,24] makes it the less sizable but more targeted option.

4.3 Data augmentation

Lastly, we experiment many data augmentation techniques to compensate for the data scarcity of few-shot configurations. Fig. 4 shows samples of the data augmentation techniques we experiment during the development process of the VAVnets.

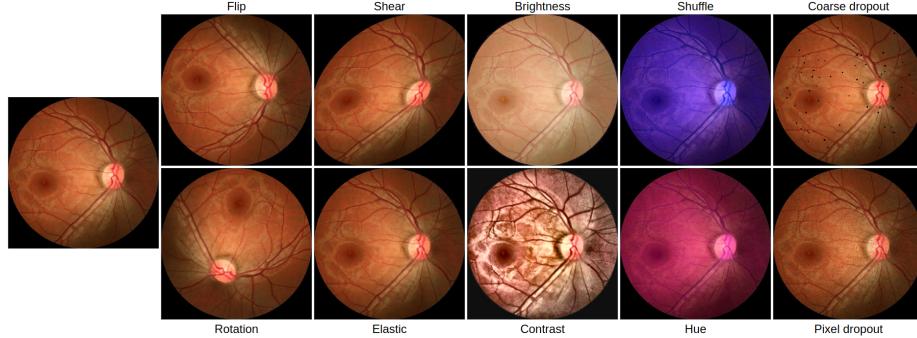


Fig. 4. Samples of the data augmentation techniques

We selected augmentation techniques to mimic the theoretical new unseen data. The different perspectives and orientations that a camera could have taken to obtain the fundus image is reproduced by flips and rotations. The variety of potential retinal vascular shapes that could be present in other fundus is approximated by shearing and elastic transforms. Lighting conditions during the image acquisition are mimicked by non-linear transformations of pixel intensities leading to changes in brightness and contrast. The presence of pathologies or masking luminous artifacts caused during acquisition is represented by masking out small to very small regions of the image, referred to as coarse and pixel dropout.

Also, we observed many manual segmentations performed on identical images by different annotators. Some annotators tend to rely heavily on colors to differentiate arteries and veins. This leads to huge classification differences between them on smaller vasculature, which thus tend to be way less differentiable by solely their color aspect. To steer the deep-learning process away from focusing too much on color, we shuffle the color channels and modify the hue of the image.

5 VAVnets experiments

For the experiments on the **architecture**, we found that the use of normalization techniques proves to be most beneficial for our specific tasks. In particular, batch normalization [15] when added after each convolution layer. However, we utilize it in a non-standard manner during inference. After training on input training data, we can perform inference on test data using either (Learn) or (Pred). When using (Learn), the batch normalization layers' expectation and variance are computed over the training data. Conversely, when using (Pred), the batch normalization layers' expectation and variance are computed over the testing data. Applying (Pred) on the generated weights and biases during inference, as mentioned in [15], allows us to recover the original activations by setting $\gamma^{(k)} = \sqrt{\text{Var}[x^{(k)}]}$ and $\beta^{(k)} = \text{E}[x^{(k)}]$. It involves no cost, neither in time nor

in training resources. This utilization of normalization is particularly important in domain shift scenarios, considering the performance gap observed between (Pred) and (Learn), as shown in Tab. 2, where (pred) outperform (Learn) at all tasks. The performance gap is observed at the very first stage of the network development (First results) as well as in the very last stage (Last results). The first modification to U-Net is the addition of batch-normalization layers after each convolution. The last development step is the use of data augmentation describe afterwards. Results are reported as the average of each segmentation dice score over all datasets.

	Vessels	Veins	Arteries
First results			
Pred	0.813 ± 0.001	0.716 ± 0.002	0.689 ± 0.006
Learn	0.754 ± 0.005	0.591 ± 0.017	0.512 ± 0.027
Last results			
Pred	0.821 ± 0.001	0.788 ± 0.003	0.746 ± 0.005
Learn	0.765 ± 0.003	0.677 ± 0.012	0.626 ± 0.017

Table 2. Inference using the learned data statistics (Learn) compared to inference using the testing data statistics (Pred). **Bold green** is the best dice score.

Less impactful but still valuable insights are utilized to construct the current configurations. For the vessels segmentation task, U-Net depth proves to be the optimal choice. Deeper or shallower layers compared to U-Net result in a performance decrease. Such cases may introduce either too little or unnecessary complexity for vessels segmentation. In contrast, for veins and arteries, deeper layers yield significantly better results. This suggests that deeper layers are advantageous for capturing the intricate patterns that help differentiate between veins and arteries in the vascular tree. Experiments involving down-sampling layers and skip connections do not lead to any improvement compared to U-Net. Regarding up-sampling layers, while variations in performance are observed among the interpolation methods tested, the reasons behind these differences are not entirely clear. The distinctions are subtle, particularly for veins, where the superiority of one method over another is weakly discerned. However, the chosen changes still result in improved results.

For the experiments on **transfer-learning**, performances gradually increase with the proximity of the used data to the vascular structure: common objects, curvilinear structures, and retinal vessels. Initially, the use of COCO-trained weights probably under-performed because vessels, veins, and arteries are too thin and precise structures to benefit from common objects' gathered information. Subsequently, for the use of curvilinear structures, the results were higher than with common objects, but still insufficient. Finally, the experiments exploring the utilization of new vessels datasets to enhance performances lead to the most interesting findings. We conducted three approaches: transfer-learning with

no fine-tuning (referred to as: *TL*), transfer-learning with complete fine-tuning (referred to as: *FT*), and single-pass cross-dataset training (referred to as: *CD*) with 1493 images from new vessels datasets combined with the 40 images in the cross-dataset-build training subset. *TL* results in performance reduction. Examination of the predicted segmentations reveals that without fine-tuning, the network only generates larger structures. While this approach creates very few false positives, it also misses the details of the retinal structures, explaining the diminished performance. In contrast, both *FT* and *CD* demonstrate comparable results with the 40 images cross-dataset-build training from scratch. This indicates that using the pre-trained network with complete fine-tuning or incorporating cross-dataset training put the networks performances to a level on par with the network trained from scratch. Despite the multiplication of training data by 37, it does not yield additional performance improvements.

In contrast to the use of external datasets, **data augmentation** significantly improves performance, particularly for veins and arteries segmentation. Initially, the increase in data quantity through data replication led to a slight decrease in performance, indicating slight over-fitting to the training set. The increase in performances that we observed afterwards is thus solely related to the data augmentation techniques employed. For all three tasks, as shown in Tab. 3 the most significant performance enhancements are consistently achieved through the flips & rotations, showing growth that correlates with the data replication scaling. Shearing & elastic transforms augmentations exhibit a more ambiguous impact on performance, with slight enhancements for veins and arteries while showing a performance decrease for vessels. In opposition, brightness & contrast, coarse dropout & pixel dropout, and channel shuffle & hue modifications augmentations consistently result in performances inferior to those without any augmentation.

r	Vessels	Veins	Arteries
No DA			
1	0.813 ± 0.001	0.732 ± 0.002	0.696 ± 0.007
Flip & Rotation			
1	0.813 ± 0.004	0.736 ± 0.001	0.697 ± 0.014
2	0.819 ± 0.003	0.758 ± 0.008	0.720 ± 0.008
3	0.821 ± 0.001	0.776 ± 0.003	0.737 ± 0.005
Shear & Elastic			
1	0.810 ± 0.003	0.719 ± 0.012	0.696 ± 0.007
2	0.808 ± 0.004	0.734 ± 0.006	0.705 ± 0.005
3	0.805 ± 0.004	0.742 ± 0.002	0.696 ± 0.002
Brightness & Contrast			
1	0.809 ± 0.001	0.721 ± 0.002	0.680 ± 0.002
2	0.799 ± 0.002	0.721 ± 0.004	0.669 ± 0.001
3	0.799 ± 0.002	0.710 ± 0.003	0.649 ± 0.004
Channel Shuffle & Hue			
1	0.804 ± 0.004	0.719 ± 0.001	0.661 ± 0.007
2	0.800 ± 0.001	0.719 ± 0.007	0.659 ± 0.012
3	0.793 ± 0.002	0.713 ± 0.001	0.652 ± 0.008
Coarse dropout & Pixel dropout			
1	0.811 ± 0.002	0.730 ± 0.001	0.686 ± 0.009
2	0.802 ± 0.001	0.732 ± 0.002	0.688 ± 0.004
3	0.801 ± 0.002	0.724 ± 0.002	0.684 ± 0.004

Table 3. Data augmentation (DA) scaling through replication factor (r). **Bold green** is the best dice score. **Bold blue** signals that the dice score increased compared to (No DA). **Bold red** signals that the dice score decreased compared to (No DA). **Bold black** signals equivalent performances to (No DA).

Finally, as shown in Tab. 4 we scaled further the data replication factor (r) during the experiments. For vessels, the replication was stopped at $r = 6$ due to consecutive performance decreases from $r = 3$. For arteries and veins, replication was extended to $r = 15$ to ensure a thorough understanding of the behavior, as performance continued to improve with increasing replication factors. The vessels segmentation performances are already high without any augmentation which limits the impact of scaling to provide additional enhancements. The arteries and veins, as for them, benefits a lot more than the vessels. Results are reported as the average of each segmentation dice score over all datasets.

r	Vessels	Veins	Arteries
1	0.813 ± 0.004	0.736 ± 0.001	0.697 ± 0.014
2	0.819 ± 0.003	0.758 ± 0.008	0.720 ± 0.008
3	0.821 ± 0.001	0.776 ± 0.003	0.737 ± 0.005
4	0.819 ± 0.002	0.781 ± 0.002	0.733 ± 0.008
5	0.818 ± 0.001	0.778 ± 0.007	0.743 ± 0.001
6	0.815 ± 0.002	0.782 ± 0.001	0.745 ± 0.002
7	-	0.783 ± 0.006	0.743 ± 0.006
8	-	0.782 ± 0.004	0.746 ± 0.005
9	-	0.784 ± 0.007	0.737 ± 0.001
10	-	0.780 ± 0.007	0.742 ± 0.003
12	-	0.780 ± 0.005	0.742 ± 0.006
15	-	0.781 ± 0.006	0.730 ± 0.009

Table 4. Flips and rotations augmentations response to scaling. **Bold green** is the best dice score. **Bold blue** is the second best.

In conclusion, retinal vessels in fundus images are unique structures that vary significantly between datasets, influenced by acquisition material, illumination conditions, and the presence of diseases. A well-suited architecture is essential to maximize performances in each sub-task of vasculature segmentation: vessels, veins, and arteries. Counterintuitively, external data do not enhance performance in this cross-dataset and few-shot scenario. This is also true for most data augmentation techniques that attempt to replicate the appearance of new unseen data. However, some data augmentation techniques, as flips and rotations, work efficiently and demonstrate a notable ability to scale with the quantity of data. Fig. 5 shows the results of VAVnets segmentations alongside corresponding groundtruth.

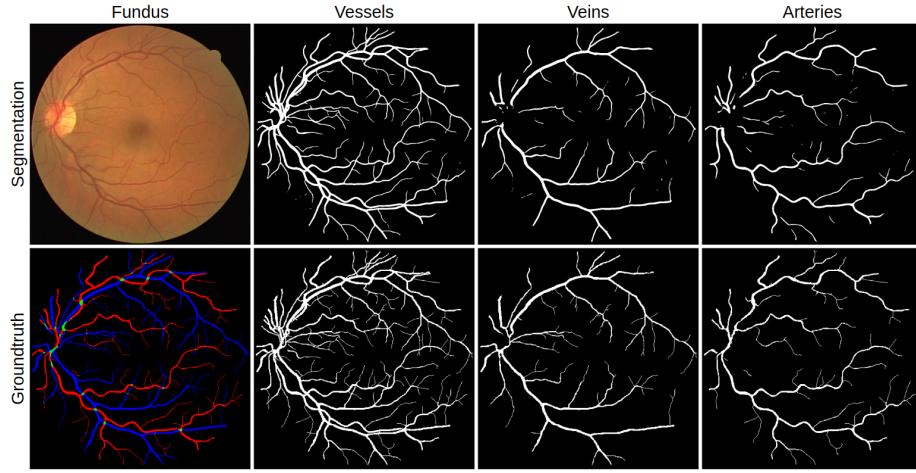


Fig. 5. Segmentation and corresponding groundtruth for vessels, veins and arteries.

6 Comparison to related works

The performances of VAVnets are compared to the state-of-the-art methods for vasculature segmentation in few-shot scenarios. Tab. 5 shows vessels segmentation and Tab. 6 shows veins and arteries segmentation. The values are reported from the referenced articles and rounded to two decimal places when more were given. If the third decimal is five or higher, the value is rounded up.

date	method	DRIVE	DUMO	HRF	LESAV
2022	[11]	-	-	0.64	-
2022	[7]	< 0.80	-	-	-
2022	[9]	0.79	-	0.74	-
2023	[10]	-	-	0.68	-
2023	[8]	0.77	-	0.67	-
2024	VAVnets	0.81	0.83	0.81	0.86

Table 5. Comparison between vessels segmentation methods in few-shot scenarios.

Bold green is the best dice score. **Bold blue** is the second best.

Vessels segmentation

We have shown in Sec. 3 that [7,8] propose architectural modifications, [10,11] employ prior knowledge, and [9] leverage on data augmentation. For the development of VAVnets, we leverage on architecture, transfer-learning, and data

augmentation all together. Tab. 5 shows the comparisons of VAVnets performances with the state-of-the-art on vessels segmentation. On the DRIVE dataset composed of only seven pathological images out of 40 and a small resolution of 565*584, the performances of VAVnets are superior, but not by a large margin. For HRF, the performances of VAVnets are consequently superior. HRF is composed of 30 pathological images out of 45, and an high resolution of 3504*2336 thus each image present more details due to the visibility of more depth in the vessels trees. We attribute the performances gap to the leverage on architecture, transfer-learning, and data augmentation all together. The performances of VAVnets on DUMO and LESAV are referenced for comparison with other methods that may be developed in the future.

In [7], the F1 scores are not explicitly provided in the text and can only be roughly estimated from a chart. From the chart, it appears that the average F1 score is slightly below 0.8. For the purposes of reported tables, we will assume an average F1 score of approximately 0.8. We want to mention that dice and F_B scores are numerically equal in binary cases.

Veins & Arteries segmentation

The most difficult but also most interesting tasks are the arteries and the veins segmentations. Tab. 6 shows the results of VAVnets compared to related works for the segmentation of veins and arteries in few-shot scenarios.

date	method	Veins				Arteries			
		DRIVE	DUMO	HRF	LESAV	DRIVE	DUMO	HRF	LESAV
2022	[12]	0.63	-	0.50	0.61	0.57	-	0.46	0.58
2023	[13]	0.71	-	0.70	0.67	0.72	-	0.69	0.72
2024	VAVnets	0.78	0.81	0.78	0.79	0.73	0.78	0.74	0.77

Table 6. Comparison between veins and arteries segmentation methods in few-shot scenarios. **Bold green** is the best dice score. **Bold blue** is the second best.

As mentioned in Sec. 3, Shi Danli *et al.* propose a binary-output approach in [12] and a multi-class output strategy in [13]. The performance of the binary-output approach in [12] is consistently lower than that of the multi-class approach in [13]. However, VAVnets, despite using a binary-output strategy, demonstrate superior performances. The only exception is for artery segmentation on the DRIVE dataset, where VAVnets perform almost equivalently.

First, comparing VAVnets to [12] (binary-output approach), we attribute the performance gap to our use of batch-normalization layers during inference, our data augmentation choices, and proper scaling application. Tab. 2 presents the results of VAVnets at a developmental stage without data augmentation. The

(Pred) describe our use of the normalization where (Learn) describe the absence of it. The veins and arteries segmentations performances reported in this table are on par with those of [12]. Remaining performances differences may be attributed to architectural differences.

Second, comparing VAVnets to [13], we mainly attribute the performance gap to the use the binary-output strategy instead of the multi-class strategy. The binary-output strategy is better than the multi-class strategy because of the complexity of the segmentation task. The structures are uncommon and consist of complex patterns. The task involves segmenting two tortuous trees of various sizes, depths, and branching patterns. Moreover, the trees overlap in a 2D view and can be obscured by pathologies, making some parts invisible. Nonetheless, overlapping areas are crucial, and should not be classified as only veins or arteries. Multi-class segmentation in the literature typically uses only three classes: arteries, veins, and background. It is the case here in [13]. Methods that attempt to detect overlaps as a separate class often fail because overlaps represent a small and disparate percentage of the data. However, combining binary segmentations of arteries and veins leads to more accurate overlap segmentations.

Another issue with the multi-class approach is that it introduces discontinuities in the structures. We referred to them as intra-vascular and inter-vascular miss-classification. Inter-vascular miss-classification is the close alternation between arteries and veins along a single vascular path. Intra-vascular miss-classification is the presence of veins pixels within a portion that is dominantly classified as arteries, and vice versa. However, both inter-vascular and intra-vascular miss-classifications are less common when performing binary segmentations.

Additionally, the veins are typically better predicted than the arteries in most state-of-the-art methods, as they are more easily differentiable by their appearance. In contrast, in [13], the opposite is true. This may seem unusual, but the explanation lies in both: i) the anatomical specificities of the eye, and ii) the multi-class output strategy without considering the overlap class.

i) In fundus images, at overlapping positions, arteries tend to be over veins, meaning that veins are not visible in the 2D view. In [13], as mentioned in Sec. 3, the method rely heavily on transfer-learning from fundus fluorescein angiography images (referred to as *FFA*). The *FFA* groundtruth are acquired by fluorescein dye. Such *FFA* do not have overlap labels, one pixel correspond to only one class. Thus mostly arteries are classified at overlap points. Prioritizing arteries classification at overlap positions makes the network learn this behavior.

ii) The choice of a multi-class output strategy without considering the overlap class leads to the detection of only arteries at overlap points. On the other hand, DRIVE, HRF, and LESAV datasets utilized for performances assessment have overlap labels. Therefore, veins are never detected at overlaps points, but are compared to a groundtruth that has veins labels at overlap points, resulting in

decreased performances. Conversely, overlaps are easily detected at a convolutional level, thus all the crossings are well predicted as arteries, enhancing the performances. This could justify the better performances for arteries than veins.

Finally, across all datasets, VAVnets demonstrate significantly superior performances for both veins and arteries, with one exception: arteries segmentation on the DRIVE dataset, where results are on par. This can be attributed to two main factors. The aforementioned impact of the multi-class strategy and the utilization of FFA groundtruth. The nature of the DRIVE dataset, which consists of only seven pathological images out of 40 and has a relatively small resolution of 565*584, making the segmentation process easier. Comparatively, HRF comprises 30 pathological images out of 45, with a higher-than-DRIVE resolution of 3504*2336, providing more detailed information. Similarly, LESAV includes 11 pathological images out of 22 and has a higher resolution of 1620*1444 compared to DRIVE, also providing more details.

7 Conclusion

In this article, we propose VAVnets: three variants of a deep-learning network designed to generate binary segmentations of vessels, veins, and arteries. We focus our experiments on few-shot scenarios and cross-dataset training. Through architectural changes, transfer learning, and data augmentation experiments, we optimize VAVnets. Our results demonstrate that VAVnets outperform state-of-the-art methods for vessels, veins, and arteries segmentation across multiple datasets: DRIVE, DUMO, HRF, and LESAV. In conclusion, VAVnets significantly enhance the reliability of retinal vasculature segmentation in few-shot scenarios.

Segmentation alone cannot indicate health issues, so quantitatively measuring changes in retinal veins and arteries is still necessary. Such measurements need to be performed on coherent structures. In the future, we plan to further enhance the segmentation of veins and arteries at both pixel and structural levels. Additionally, we aim to ensure a fully connected structure through reconnection, ultimately enabling quantitative measurements of the retinal vasculature.

References

1. J. Staal, M. D. Abràmoff, M. Niemeijer, M. A. Viergever, and B. van Ginneken, “Ridge-based vessel segmentation in color images of the retina,” *IEEE Transactions on Medical Imaging*, vol. 23, pp. 501–509, 2004.
2. S. Zhang, R. Zheng, Y. Luo, X. Wang, J. Mao, C. J. Roberts, and M. Sun, “Simultaneous arteriole and venule segmentation of dual-modal fundus images using a multi-task cascade network,” *IEEE Access*, vol. 7, pp. 57561–57573, 2019.
3. J. Odstrcilík, R. Kolář, A. Budai, J. Hornegger, J. Jan, J. Gazárek, T. Kubena, P. Černosek, O. Svoboda, and E. Angelopoulou, “Retinal vessel segmentation by improved matched filtering: evaluation on a new high-resolution fundus image database,” *IET Image Process.*, vol. 7, pp. 373–383, 2013.

4. J. I. Orlando, J. B. Breda, K. van Keer, M. B. Blaschko, P. J. Blanco, and C. A. Bulant, “Towards a glaucoma risk index based on simulated hemodynamics from fundus images,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2018.
5. A. Khandouzi, A. Ariaifar, Z. Mashayekhpour, M. Pazira, and Y. Baleghi, “Retinal vessel segmentation, a review of classic and deep methods,” *Annals of Biomedical Engineering*, vol. 50, no. 10, pp. 1292–1314, 2022.
6. 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*, pp. 234–241, Springer, 2015.
7. J. Xu, J. Shen, C. Wan, Q. Jiang, Z. Yan, and W. Yang, “A few-shot learning-based retinal vessel segmentation method for assisting in the central serous chorioretinopathy laser surgery,” *Frontiers in Medicine*, vol. 9, p. 821565, 2022.
8. H.-C. Shao, C.-Y. Chen, M.-H. Chang, C.-H. Yu, C.-W. Lin, and J.-W. Yang, “Retina-transnet: A gradient-guided few-shot retinal vessel segmentation net,” *IEEE Journal of Biomedical and Health Informatics*, vol. 27, pp. 4902–4913, 2023.
9. J. Lyu, Y. Zhang, Y. Huang, L. Lin, P. Cheng, and X. Tang, “Aadg: Automatic augmentation for domain generalization on retinal image segmentation,” *IEEE Transactions on Medical Imaging*, vol. 41, pp. 3699–3711, 2022.
10. D. Hu, H. Li, H. Liu, X. Yao, J. Wang, and I. Oguz, “Vesselmorph: Domain-generalized retinal vessel segmentation via shape-aware representation,” *arXiv preprint arXiv:2307.00240*, 2023.
11. D. Hu, H. Li, H. Liu, and I. Oguz, “Domain generalization for retinal vessel segmentation with vector field transformer,” in *Proceedings of The 5th International Conference on Medical Imaging with Deep Learning*, 2022.
12. D. Shi, Z. Lin, W. Wang, Z. Tan, X. Shang, X. Zhang, W. Meng, Z. Ge, and M. He, “A deep learning system for fully automated retinal vessel measurement in high throughput image analysis,” *Frontiers in Cardiovascular Medicine*, vol. 9, p. 823436, 2022.
13. D. Shi, S. He, J. Yang, Y. Zheng, and M. He, “One-shot retinal artery and vein segmentation via cross-modality pretraining,” *Ophthalmology Science*, p. 100363, 2023.
14. D. Ruan, D. Wang, Y. Zheng, N. Zheng, and M. Zheng, “Gaussian context transformer,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 15129–15138, 2021.
15. S. Ioffe and C. Szegedy, “Batch normalization: Accelerating deep network training by reducing internal covariate shift,” in *International conference on machine learning*, pp. 448–456, pmlr, 2015.
16. N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, “Dropout: a simple way to prevent neural networks from overfitting,” *The journal of machine learning research*, vol. 15, no. 1, pp. 1929–1958, 2014.
17. R. Eldan and O. Shamir, “The power of depth for feedforward neural networks,” in *Conference on learning theory*, pp. 907–940, PMLR, 2016.
18. T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, “Microsoft coco: Common objects in context,” in *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6–12, 2014, Proceedings, Part V 13*, pp. 740–755, Springer, 2014.
19. V. Mnih, *Machine learning for aerial image labeling*. PhD thesis, 2013.
20. J. König, M. Jenkins, M. Mannion, P. Barrie, and G. Morison, “What’s cracking? a review and analysis of deep learning methods for structural crack segmentation, detection and quantification,” *arXiv preprint arXiv:2202.03714*, 2022.

21. L. Mou, Y. Zhao, H. Fu, Y. Liu, J. Cheng, Y. Zheng, P. Su, J. Yang, L. Chen, A. F. Frangi, *et al.*, “Cs2-net: Deep learning segmentation of curvilinear structures in medical imaging,” *Medical image analysis*, vol. 67, p. 101874, 2021.
22. R. J. Chalakkal, W. H. Abdulla, and S. Sinumol, “Comparative analysis of university of auckland diabetic retinopathy database,” in *International Conference on Signal Processing Systems*, 2017.
23. N. Popovic, S. Vujosevic, M. Radunović, M. Radunović, and T. Popović, “Trend database: Retinal images of healthy young subjects visualized by a portable digital non-mydriatic fundus camera,” *PLoS ONE*, vol. 16, 2021.
24. K. Jin, X. Huang, J. Zhou, Y. Li, Y. Yan, Y. Sun, Q. Zhang, Y. Wang, and J. Ye, “Fives: A fundus image dataset for artificial intelligence based vessel segmentation,” *Scientific Data*, vol. 9, 2022.
25. K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.