

# VessShape: Few-shot 2D blood vessel segmentation by leveraging shape priors from synthetic images

Wesley Nogueira Galvão<sup>1</sup> and Cesar H. Comin<sup>1,\*</sup>

<sup>1</sup>*Department of Computer Science, Federal University of São Carlos, São Carlos, SP, Brazil*

(Dated: September 26, 2025)

??

Keywords: Blood vessel segmentation, Connectivity, post-processing

## I. INTRODUCTION

VessShape is a synthetic image dataset that combines tubular, vessel-like shapes with diverse foreground and background textures. The central idea is to provide a robust set that can be used to pre-train blood vessel segmentation models by keeping geometric priors fixed and drastically changing the texture, encouraging models to learn shape cues (connectivity, tapering, bifurcations) rather than learning the texture.

## II. RELATED WORKS

Many previous works have considered shape priors for medical image segmentation [??]. This is possible when organelles, cells, or organs have a known shape and the segmentation process can assume an optimal shape or additional optimization criteria such as the requirement of smooth borders. Instance segmentation is probably the most common task in which shape priors have been explored. A particular challenge with blood vessel identification is that it usually requires a semantic segmentation of the image.

Before the popularity of deep learning methods, local shape priors were the dominant approach to vessel segmentation. Blood vessels tend to have a tubular structure. Thus, the usual approach was to develop filters aimed at identifying tubular objects. A popular line of research was based on the eigenvalues of the Hessian matrix [1, 2]. Arguably, the most popular Hessian-based method is the Frangi filter [3], which consists of combining the eigenvalues to define a tubularity score for the pixels of the vessels. Another popular line of research was based on the definition of line or Gabor filter templates to identify relevant vessel structures. An important drawback of these methods is that the tubularity assumption is not valid at bifurcation and termination points.

With the emergence of deep learning, some works have explored adding shape priors during network training [4]. Priors have been added on the network input using vesselness filters [5–7], on the network architecture using learnable vesselness or Gabor filters [8–10] and on the network output using topology-aware loss functions [11–13].

For neural networks, it is simpler and likely more flexible to consider a data-based approach to add shape priors to the model. That is, to focus on generating a large and diverse set of images containing a priori knowledge regarding the structure of blood vessels. To our knowledge, no previous work tried to use the same method considered in our work. The closest approach is to generate synthetic images that are as similar as possible to the samples in the dataset. This has been done using two main strategies: i) using an appearance model of the vessels and image background [14–17] and ii) using style transfer or a generative model to create the samples.

Modeling-based approaches usually start by generating a biologically plausible topology of the vasculature, followed by the definition of varying radii for vessel segments and the inclusion of texture for the vessels and the background. The typical noise found in the imaging modality of interest is also modeled. An important recent work in this direction is a foundation model called VesselFM [15] that was trained on a large number of synthetic and real images.

??

The main drawback of the aforementioned works is that the network is trained to identify both the shape and texture of the blood vessels in the samples. Changes in the texture of the vessels due to diseases or modifications of the imaging device can lead to low segmentation performance. In addition, different models must be trained for different imaging modalities, where annotated data can be scarce. Our approach aims at training neural networks to segment any vascular tissue that follows the shape priors acquired from VessShape.

## III. METHODOLOGY

### A. The VessShape Dataset

The geometry of the synthetic images is based on Bézier curves, which allow a flexible and controlled representation of tubular shapes. Each vascular segment is described by an  $n$ th order Bézier curve with control points  $\{\mathbf{p}_i\}_{i=0}^n$ . The tortuosity of the segments is adjusted by small perturbations to the control points, ensuring that vessel geometry is realistic and diverse. The Bézier curve  $\mathbf{c}(t)$  of a vessel segment is given by

---

\* Corresponding author: [comin@ufscar.br](mailto:comin@ufscar.br)

$$\mathbf{c}(t) = \sum_{i=0}^n \binom{n}{i} (1-t)^{n-i} t^i p_i, \quad (1)$$

where  $t$  varies from 0 to 1.

To generate a curve, the first ( $\mathbf{p}_0$ ) and last ( $\mathbf{p}_n$ ) control points are randomly drawn with uniform probability. The remaining control points are generated by defining  $n - 2$  equally spaced points between a line connecting  $\mathbf{p}_0$  and  $\mathbf{p}_1$ . Each of these points is then displaced by a random amount along a normal vector  $\mathbf{n}_l$ . This normal vector is a unit vector perpendicular to the line between  $\mathbf{p}_0$  and  $\mathbf{p}_1$ . The displacement amount is drawn from a uniform distribution in  $[-\delta, \delta]$ . Lower values of  $\delta$  lead to more straight curves.

For the generation of the binary mask  $M$ , each curve is discretized by sampling points at a sufficient resolution to capture its curvature, which are then sequentially connected to form a 1-pixel-thick polyline on the image grid. Subsequently, a binary morphological dilation with a disk-shaped structuring element of radius  $r_0$  is applied, assigning a constant tubular thickness to the segments.

To generate each binary mask, the number of segments  $K$ , the order  $n$  of the Bézier curves, the displacement scale  $\delta$  and the radius  $r_0$  are all randomly sampled from an interval to ensure a wide variety of shapes. Table I summarizes the parameters used in the VessShape dataset generation, along with their sampling ranges and descriptions.

To compose the final image  $I$  from a binary mask  $M$ , a foreground texture  $F$  and a background texture  $B$  are inserted into, respectively, the generated vessel segments and the background of the image. The textures are randomly selected from ImageNet [18]. Specifically, for each mask  $M$ , two images are randomly drawn from two distinct classes of the ImageNet dataset. The images are then randomly cropped and resized to the target dimensions ( $H \times W$ ). An alpha matte  $A$  is then generated by smoothing  $M$  with a Gaussian filter of standard deviation  $\sigma$  and normalizing its values to the  $[0, 1]$  range. The textures are subsequently blended using this matte according to

$$I = A F + (1 - A) B, \quad (2)$$

which ensures that vessel regions ( $A \approx 1$ ) preserve the foreground while non-vessel regions ( $A \approx 0$ ) retain the background.

Parameter  $\sigma$  controls the smoothness of the vessel boundaries. After composition, the image  $I$  undergoes channel-wise normalization using the ImageNet statistics for compatibility with pre-trained models. Examples of generated masks and images are shown in Figure 1.

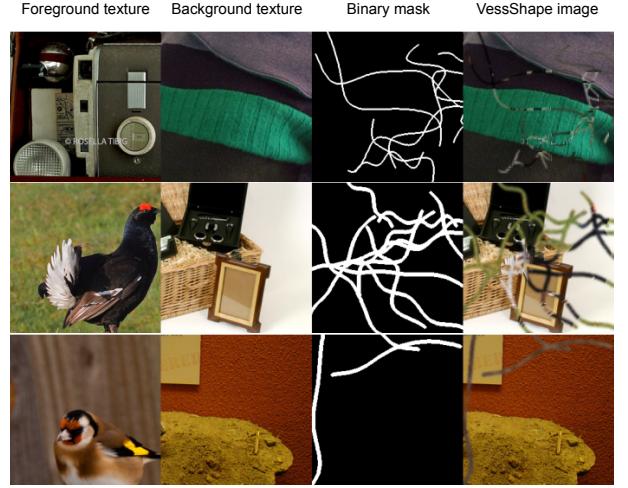


FIG. 1. Examples of ImageNet samples, synthetic geometry and respective VessShape images. The ImageNet samples are used as textures for the synthetic binary masks, defining the respective VessShape image.

## B. Real-world data for validation

To quantify the usefulness of the shape bias introduced by the VessShape dataset, we consider two blood vessel datasets: DRIVE and VessMAP. The DRIVE dataset [9] serves as a popular standard for benchmarking retinal vessel segmentation algorithms and is composed of 40 fundus photographs split into 20 for training and 20 for testing, each measuring  $584 \times 565$  pixels. The VessMAP dataset consists of 100 images,  $256 \times 256$  pixels each, acquired via fluorescence microscopy of the mouse cortex. It was curated to include a variety of challenging vascular features, such as inconsistent noise and contrast levels, different vessel sizes, prominent imaging artifacts, and intensity fluctuations within vessel structures.

The two datasets originate from fundamentally different imaging modalities, resulting in distinct characteristics. The fundus images in DRIVE, which capture the entire retina, possess a clear global structure that includes landmarks like the optic disk. The samples also contain many very thin vessels which are challenging to segment. In contrast, the VessMAP images are highly magnified views of small cortical areas and have no discernible global organization. The borders of the vessels are generally less defined than the vessels from DRIVE. Another key difference is that, without any processing, the vessels in VessMAP are bright with dark backgrounds while the vessels in DRIVE are dark with bright backgrounds. Figure 2 shows one sample from each dataset.

TABLE I. Main parameters used for generating the VessShape dataset.

| Parameter                        | Range         | Description  |
|----------------------------------|---------------|--|
| Number of curves $K$             | [1, 20]       | Number of branches/vessels generated per sample.                                       |
| Control points $n+1$             | [2, 20]       | Bézier curve complexity (order $n$ ).  |
| Displacement scale $\delta$ (px) | [50.0, 150.0] | Controls curvature/tortuosity via the typical amplitude of control-point displacement. |
| Initial radius $r_0$ (px)        | [1, 5]        | Basal vessel thickness; a smooth taper is applied along the branch.                    |
| Matting blur $\sigma$            | [1, 2]        | Standard deviation of the Gaussian used for $A = G_\sigma * M$ .                       |

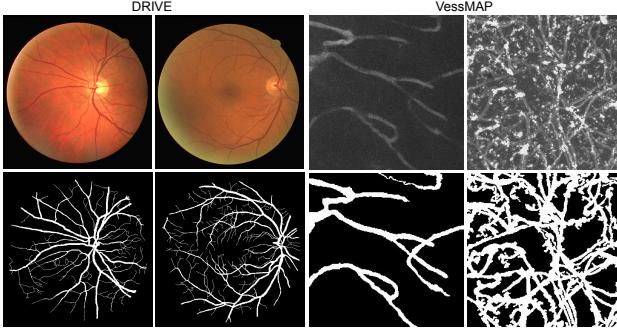


FIG. 2. Samples from the DRIVE and VessMAP datasets and their respective ground-truth masks.

### C. Model architectures and training strategies

We adopt an encoder-decoder U-Net design, that is, the decoder is symmetrical to the encoder and the model contains skip connections between different encoder-decoder stages. Two models are compared, one with a ResNet18 and the other with a ResNet50 encoder [19]. The models were instantiated from the *Segmentation Models Pytorch* Python package<sup>1</sup>.

Two training scenarios are considered. In the first, training is done from scratch separately on the DRIVE and VessMAP datasets to establish a baseline. The second scenario consists of pre-training on the VessShape dataset and fine-tuning on DRIVE and VessMAP to measure the transferability and sample efficiency of the learned representations. These two scenarios involve three distinct training procedures: i) training from scratch on natural datasets; ii) pre-training on VessShape and iii) fine-tuning on real-world data. These procedures are described next.

### D. Training from scratch on natural datasets

#### E. Pre-training on VessShape

The pre-training on VessShape aims to expose the model to a wide variety of tubular geometries while keeping texture as a secondary cue, which tends to benefit the

TABLE II. Pre-training hyperparameters on VessShape.

| Hyperparameter      | VSUNet50 | VSUNet18 |
|---------------------|----------|----------|
| Batch size          | 96       | 192      |
| Learning rate       | 1.0e-3   | 1.0e-2   |
| Learning rate decay | 0.0      | 0.0      |
| Weight decay        | 1.0e-4   | 0.0      |
| Img/epoch           | 50,000   | 50,000   |
| Max epochs          | 3000     | 1000     |

TABLE III. Performance of VSUNet variants after pre-training on VessShape.

| Metric | VSUNet50          | VSUNet18          |
|--------|-------------------|-------------------|
| Dice   | $0.861 \pm 0.022$ | $0.859 \pm 0.077$ |
| Acc    | $0.960 \pm 0.008$ | $0.956 \pm 0.037$ |
| IoU    | $0.758 \pm 0.032$ | $0.761 \pm 0.096$ |
| Prec   | $0.780 \pm 0.037$ | $0.774 \pm 0.096$ |
| Rec    | $0.964 \pm 0.012$ | $0.974 \pm 0.018$ |

segmentation of thin structures. The model is trained to minimize the average loss over a large number of samples drawn from the synthetic dataset

Concretely, we pre-train two U-Net models, one with a ResNet18 encoder and another with a ResNet50 encoder. We refer to them as VSUNet18 and VSUNet50, respectively. The training stream intentionally exposes the model to vessel-like geometries, while textures keep changing in the background. This narrative of abundant shapes and shifting appearances nudges the network to rely less on superficial texture cues and more on geometric regularities that matter for connectivity and thin-structure recovery. Below, we detail the VSUNet50 setup and summarize its performance on VessShape.

#### F. Fine-tuning on natural datasets

- Datasets (VessMAP e DRIVE)

- Procedure

<sup>1</sup> [https://github.com/qubvel/segmentation\\_models.pytorch](https://github.com/qubvel/segmentation_models.pytorch)

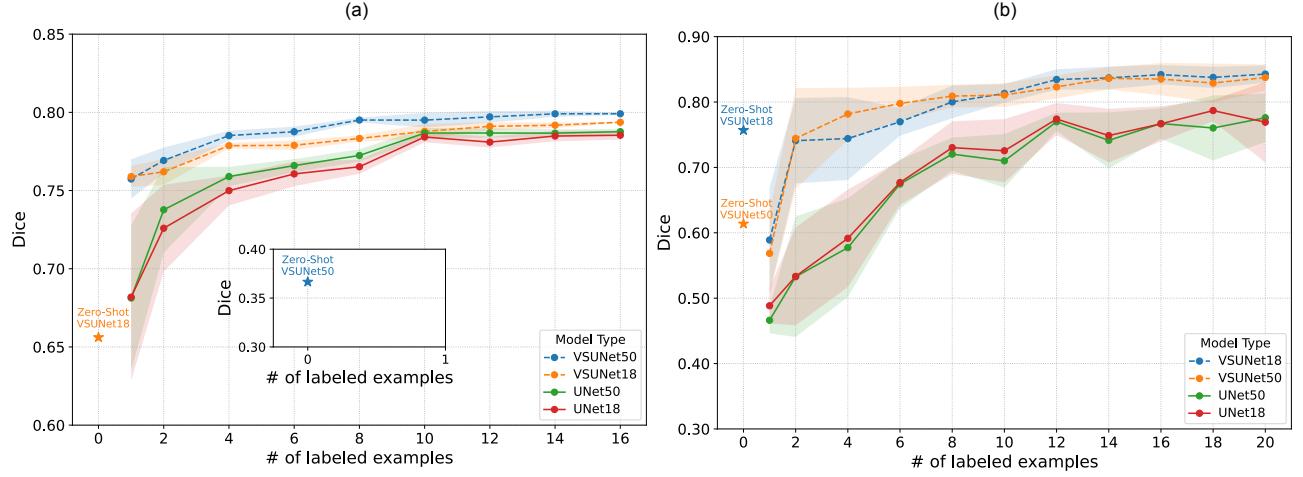


FIG. 3. ??.

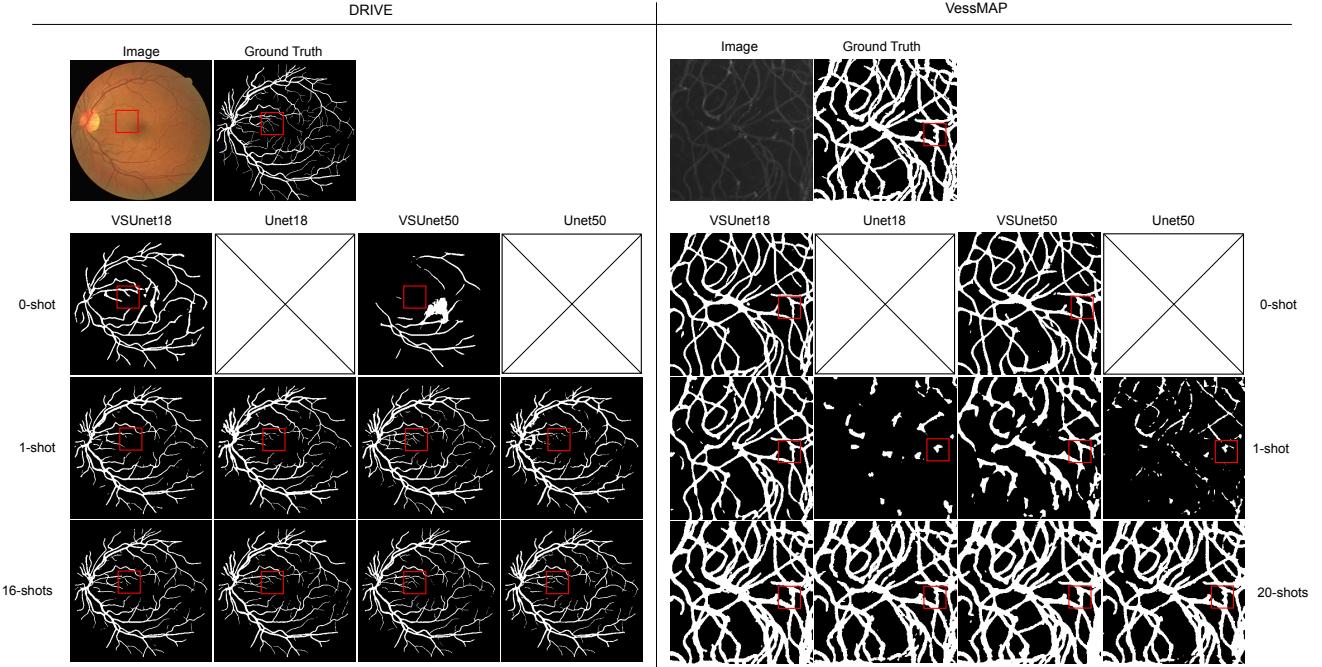


FIG. 4. ??.

## IV. RESULTS

## V. CONCLUSION

## FUNDING

C. H. Comin thanks FAPESP (grant no. 21/12354-8) for financial support.

[1] Muhammad Moazam Fraz, Paolo Remagnino, Andreas Hoppe, Bunyarat Uyyanonvara, Alicja R Rudnicka,

Christopher G Owen, and Sarah A Barman, “Blood



- sel segmentation, centerline prediction, and bifurcation detection in 3-d angiographic volumes,” *Frontiers in Neuroscience* **14**, 592352 (2020).
- [15] Bastian Wittmann, Yannick Wattenberg, Tamaz Amiranashvili, Suprosanna Shit, and Bjoern Menze, “veselfm: A foundation model for universal 3d blood vessel segmentation,” in *Proceedings of the Computer Vision and Pattern Recognition Conference* (2025) pp. 20874–20884.
- [16] Bastian Wittmann, Lukas Glandorf, Johannes C Paetzold, Tamaz Amiranashvili, Thomas Wälchli, Daniel Razansky, and Bjoern Menze, “Simulation-based segmentation of blood vessels in cerebral 3d octa images,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention* (Springer, 2024) pp. 645–655.
- [17] Joël Mathys, Andreas Plesner, Jorel Elmiger, and Roger Wattenhofer, “Synthetic data for blood vessel network extraction,” in *Will Synthetic Data Finally Solve the Data Access Problem?* (2025).
- [18] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei, “ImageNet: A large-scale hierarchical image database,” in *2009 IEEE Conference on Computer Vision and Pattern Recognition* (IEEE, 2009) pp. 248–255.
- [19] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, “Deep residual learning for image recognition,” in *Proceedings of the IEEE conference on computer vision and pattern recognition* (2016) pp. 770–778.