

An Analysis of the Influence of Transfer Learning When Measuring the Tortuosity of Blood Vessels

Matheus V. da Silva^a, Julie Ouellette^{b,c}, Baptiste Lacoste^b and Cesar H. Comin^{a,*}

^aDepartment of Computer Science, Federal University of São Carlos, São Carlos, SP, Brazil

^bDepartment of Cellular and Molecular Medicine, Faculty of Medicine, University of Ottawa, Ottawa, ON, Canada

^cNeuroscience Program, Ottawa Hospital Research Institute, Ottawa, ON, Canada

ARTICLE INFO

Keywords:

Blood Vessel
Confocal Microscopy
Morphometry Performance
Transfer Learning
Tortuosity

ABSTRACT

Background and Objective: Convolutional Neural Networks (CNNs) can provide excellent results regarding the segmentation of blood vessels. One important aspect of CNNs is that they can be trained on large amounts of data and then be made available, for instance, in image processing software. The pre-trained CNNs can then be easily applied in downstream blood vessel characterization tasks, such as the calculation of the length, tortuosity, or caliber of the blood vessels. Yet, it is still unclear if pre-trained CNNs can provide robust, unbiased, results in downstream tasks involving the morphological analysis of blood vessels. Here, we focus on measuring the tortuosity of blood vessels and investigate to which extent CNNs may provide biased tortuosity values even after fine-tuning the network to a new dataset under study.

Methods: We develop a procedure for quantifying the influence of CNN pre-training in downstream analyses involving the measurement of morphological properties of blood vessels. Using the methodology, we compare the performance of CNNs that were trained on images containing blood vessels having high tortuosity with CNNs that were trained on blood vessels with low tortuosity and fine-tuned on blood vessels with high tortuosity. The opposite situation is also investigated.

Results: We show that the tortuosity values obtained by a CNN trained from scratch on a dataset may not agree with those obtained by a fine-tuned network that was pre-trained on a dataset having different tortuosity statistics. In addition, we show that improving the segmentation accuracy does not necessarily lead to better tortuosity estimation. To mitigate the aforementioned issues, we propose the application of data augmentation techniques even in situations where they do not improve segmentation performance. For instance, we found that the application of elastic transformations was enough to prevent an underestimation of 8% of blood vessel tortuosity when applying CNNs to different datasets.

Conclusions: The results highlight the importance of developing new methodologies for training CNNs with the specific goal of reducing the error of morphological measurements, as opposed to the traditional approach of using segmentation accuracy as a proxy metric for performance evaluation.

1. Introduction

Blood vessels take part in many physiological processes in humans and animals and can be found almost anywhere in an organism. Therefore, many ailments and developmental disorders may be caused by abnormal blood vessels [33]. Thus, characterizing blood vessels is an important matter not only for diagnosis but also to help answer important research questions regarding angiogenesis [12, 7], blood vessel related ailments [5] and the blood-brain barrier [7, 8]. Common metrics for characterizing blood vessels are density and tortuosity since they have been shown to influence neuronal activation [20] and blood flow [13]. Acquiring precise values for those metrics usually requires annotating blood vessels in images, which is time-consuming and error-prone. Consequently, many image processing techniques have been developed for automatically characterizing blood vessels [25, 46]. Usually, the most challenging step of the developed techniques lies in segmenting the blood vessels [21].

More recently, owing to the recent advancements of Deep Learning methods on many computer vision tasks [9], Convolutional Neural Networks (CNN) have been applied with great success for segmenting blood vessels. Among

the most common types of images where CNNs have been applied for blood vessel analysis are retina fundus and magnetic resonance images [22, 28]. Recent works have also explored the use of CNNs for blood vessel microscopy images [44, 19, 45]. Many CNN architectures have been defined for blood vessel segmentation, U-Net-based architectures [22] being the most common. Given that annotating blood vessel images, which can contain thousands of blood vessel segments, is a demanding task, most approaches for training CNNs involve transfer learning methods [44, 19, 45]. This is especially true for 3D images, since tracing blood vessels in volumetric images is a particularly difficult task. Most commonly, transfer learning consists in obtaining some off-the-shelf network, that may or may not have been trained on blood vessel images, and fine-tuning the network for segmenting blood vessels for the type of image under study.

In this work, we investigate a possible pitfall when using transfer learning for blood vessel segmentation. The following situation, illustrated in Figure 1, is considered. Suppose that a CNN was trained based on a set of images that were obtained and manually annotated from some clinical trial or biological experiment. Most likely, the objective of the training was to maximize some segmentation performance metric such as the Intersection over Union (IoU) or the Dice coefficient. The trained network was then made available

*Corresponding author

 chcomin@gmail.com (C.H. Comin)

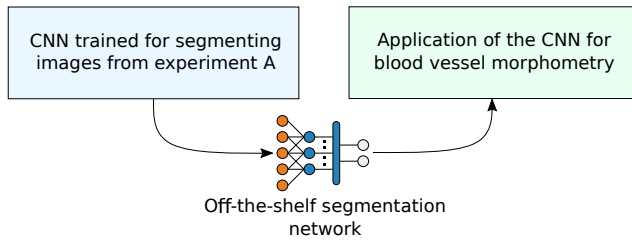


Figure 1: Illustration of a transfer learning approach commonly used for training CNNs for image segmentation. A CNN is trained on a dataset and made available to other researchers. The off-the-shelf CNN is then used as is, or fine-tuned on another dataset and included in a pipeline for blood vessel morphometry.

on some repository or software. Then, a new set of images obtained under distinct experimental conditions was acquired during some other experiment. Notice that the experiment might even be carried out by a different group of researchers. Suppose that the overall objective of this new experiment is to study the tortuosity of the blood vessels. For this task, the off-the-shelf network trained using the data from the previous experiment is applied for segmenting the blood vessels, followed by a calculation of the tortuosity.

The above situation has two important possible drawbacks that need to be taken into account. First, the objective of the original experiment where the network was initially trained was solely the segmentation of the blood vessels, there is no guarantee that the values obtained for the tortuosity are correct. Second, the data used in the two situations might have distinct characteristics. Here we do not focus on possible changes on the image characteristics, such as brightness or contrast, or the type of blood vessel under study, but on more subtle differences that might affect the blood vessels. Specifically, we consider two sets of images that have similar statistics but contain blood vessels having distinct degrees of tortuosity. The idea being that a CNN trained on blood vessels having low tortuosity, when used for segmenting highly-tortuous blood vessels, might lead to an overestimation or underestimation of the tortuosity values. The same situation is studied for the opposite problem, that is, CNNs trained on highly-tortuous blood vessels being used for segmenting blood vessels with low tortuosity.

It is important to point out that the possible drawback of transfer learning considered in this work happens mainly on segmentation tasks. For instance, for classification tasks, labeling the data is usually the end goal of digital image processing and computer vision methodologies. Contrariwise, for many analyses in biology and medicine involving digital images, the segmentation of the samples in a dataset tends to be the first step of an overall procedure aimed at characterizing the objects in the images (e.g., cells, organs and blood vessels). Thus, the results of the segmentation are used in downstream tasks involving, for instance, measuring different properties regarding the shape of the objects in the images.

In the case of blood vessels, the tortuosity is a commonly analyzed property.

The problem considered in this study is particularly relevant for current researches involving blood vessels, where the impact of an experimental condition might lead to subtle changes in the morphology of the blood vessels, which need to be quantified using robust and systematic approaches. We specifically focus on the tortuosity in this study because it is a difficult metric to be measured and compared by visual inspection, and thus requires particularly reliable means for automatic quantification.

Two main investigations are considered. The first is to assess if there are indeed possible biases that may appear on the calculated tortuosity when using off-the-shelf CNNs. Provided this is true, we verify to which extent such biases can be avoided by fine-tuning the network on the new dataset. This requires the manual annotation of additional images, and thus we also verify how many images need to be annotated in order to reduce the observed biases. In addition, we also investigate a possible data augmentation technique that can be used when training the networks in order to better generalize them for segmenting blood vessels possessing different degrees of tortuosity. All in all, our analysis involves training a CNN over 7000 times in different experimental conditions.

The analysis is focused on a large dataset of confocal microscopy images, but the results should be general for other types of images having blood vessels with a similar appearance.

1.1. Literature Review

Recently, many Deep Learning methodologies have been developed for blood vessel segmentation. Much of the effort has been focused on the segmentation of 2D retinographies. This is likely due to the availability of many annotated public datasets. For instance, Liskowski and Krawiec [23] evaluated six methodologies for retinal vessel segmentation and found that correcting class imbalances tends to be more critical for performance than applying data augmentation techniques. In the last few years, several retinal vessel segmentation analyses involving different adaptations of the U-Net architecture have been published [48, 49, 16, 11].

One disadvantage of public retinography datasets is that they are usually small. For instance, the DRIVE dataset [42], one of the most popular retinography datasets in the literature, contains only 40 images. Thus, several works in the literature use transfer learning for improving the convergence of the networks, which is done by either pre-training with extensive datasets such as the ImageNet¹ [15, 27, 26], or by generating additional synthetic data [50, 51, 2].

Besides retinography analysis, there have been some important recent developments regarding the segmentation of blood vessels in the brain, such as the DeepVesselNet pipeline [44]. This pipeline addresses the issue of collecting and annotating structures in large volumetric data. The authors suggested applying a pre-training step using a database composed of artificially generated blood vessels. This transfer

¹<https://image-net.org/>

learning approach allowed a faster network convergence while maintaining a high-quality segmentation. Todorov et al. [45] presented a methodology for analyzing important characteristics of whole mouse brain vasculature. The network architecture used was similar to DeepVesselNet, except for the input layer, which used two-channel images. A transfer learning approach involving artificially generated data was also used. With only 0.02% of the collected mouse brain being annotated, the authors were able to segment the vasculature with high accuracy. Kirst et al. [19] also defined a methodology for the segmentation of whole mouse brain vasculature using a combination of traditional methods and CNNs. While capillaries could be segmented by image filtering, thicker vessels appeared as hollow tubes. Thus, a CNN trained with synthetic data was used to fill hollow regions.

The majority of the developed methodologies were aimed mainly at improving the segmentation accuracy of blood vessels. Despite the development of additional performance metrics such as cDice [39], the influence of segmentation on downstream tasks, such as blood vessel morphometry, has rarely been addressed.

2. Methods

2.1. Training and Fine-Tuning Off-The-Shelf Segmentation CNNs

As stated above, transfer learning is an important technique for training CNNs. This is particularly true when the size of the dataset is small or it is costly to annotate the images. Figure 2 illustrates in more detail the common approach for transfer learning depicted in Figure 1. First, a group of researchers interested in segmenting blood vessels obtains a set of training images and manually annotates them to optimize a neural network (Figure 2A). After choosing a proper network architecture and optimizing it, a trained neural network is obtained. This neural network is usually evaluated through pixelwise accuracy metrics, such as IoU and Dice. Next, the researchers might make the trained network available on a public source-code repository or an image processing software.

The pre-trained network can then be used by other research groups in future studies aimed at measuring the morphology of blood vessels. The network might be used as is, or an additional fine-tuning step can be applied to adjust the network to the new data (Figure 2B). For fine-tuning the network, a subset of the new data must be manually annotated. Here, it is interesting to highlight two main points. First, the fine-tuning applied to the pre-trained network requires an amount of manual work proportional to the number of additional annotations used. Second, it is usually not possible to directly optimize the network using the morphological metrics being studied since their calculation is not fully differentiable. Thus, the metric that is optimized is usually the cross-entropy of the class probabilities or the segmentation accuracy, and the morphological metrics are calculated based on the resulting segmentations. As a consequence, there is no

guarantee that the original or fine-tuned network will provide segmentations that are unbiased with respect to the desired morphometric analyses.

To investigate the situation described above, we conducted two main analyses: (i) the identification of possible biases when performing the morphometry of a blood vessel dataset using an off-the-shelf network, initially optimized to segment a different dataset; (ii) if biases are identified, to which extent they can be mitigated by fine-tuning the off-the-shelf network to the new dataset. We divided our experiments into two steps. First, we optimized two neural networks, each specialized in segmenting blood vessels with either high or low tortuosity. Next, we measured the reliability of each specialized network when applied to blood vessels having distinct tortuosity than the blood vessels used during training.

2.2. Quantifying the Influence of Transfer Learning on Blood Vessel Tortuosity

2.2.1. Optimizing a CNN for Blood Vessel Segmentation

The first part of our experiment (illustrated in Figure 2A) consists in training a CNN capable of segmenting blood vessels with state-of-the-art accuracy. For the sake of generality, we should ideally use methodologies that are well-adopted in the literature. Therefore, our chosen network architecture was the U-Net [37], as it is widely used for segmentation tasks based on small datasets and it performs well on blood vessels [49, 16, 24].

The U-Net is composed of an encoder and a decoder. The encoder maps the input image into feature maps having progressively lower spatial resolutions, while the decoder progressively upsamples the output of the encoder and incorporates feature maps from intermediate stages of the encoder. The idea behind this methodology is to generate pixelwise probabilities for image segmentation while taking into account image features at different spatial resolutions. For our experiments, we use a U-Net with a ResNet-34 encoder [14].

Recall that the objective of our analysis is to investigate possible biases when segmenting blood vessels with high or low tortuosity. Therefore, we optimized two distinct networks. One network is trained on a dataset composed predominantly of blood vessels with high tortuosity, while the other is trained on a dataset where the blood vessels typically have low tortuosity. Section 2.5 describes the procedure used for generating these two datasets. This initial training step is depicted in Figure 3A for the case of blood vessels with low tortuosity. This training is performed for 30 epochs. 80% of the images containing blood vessels with low tortuosity were used in the training set, 10% in the validation set, and 10% in the test set. The training is carried out using the Adam optimizer [18], with the cross-entropy as a loss function. The learning rate is scheduled using the 1cycle policy [41], with a maximum value of 0.0005. The procedure used for training a network on the dataset containing blood vessels with high tortuosity is identical. Provided the training is successful, the resulting networks will be able to segment

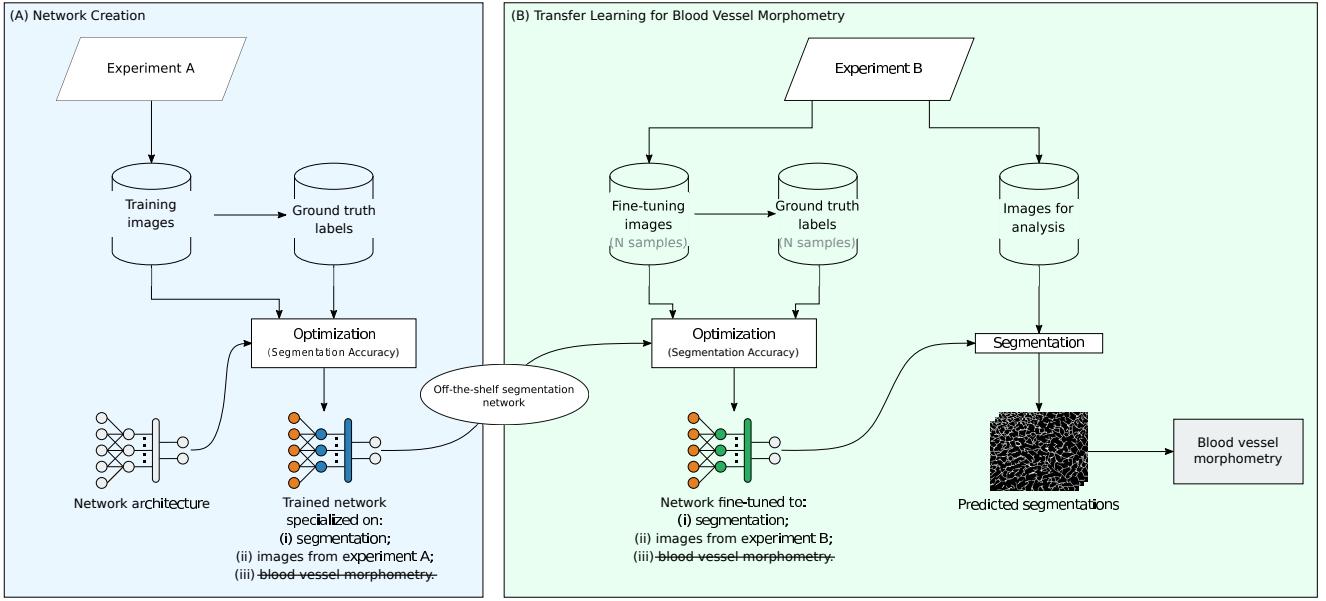


Figure 2: Detailed illustration of the transfer learning approach depicted in Figure 1. (A) Images obtained from a specific experimental condition are annotated and used for training a neural network. This network is then made publicly available. The training focus on a segmentation task. (B) The network is used for processing images obtained from another experiment. Additional images are annotated and used for fine-tuning the network with the new data. Morphological properties of the blood vessels identified in the new data are then obtained.

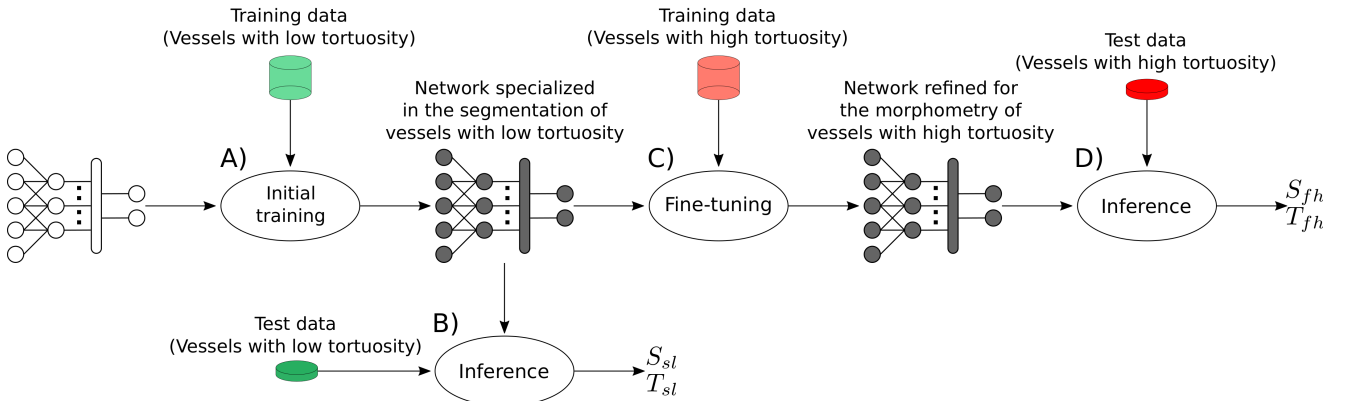


Figure 3: The methodology used for training and fine-tuning the CNNs considered in the analysis. The case where the network is initially trained on vessels with low tortuosity and fine-tuned on vessels with high tortuosity is depicted. The same procedure is used when training on vessels with high tortuosity and fine-tuning on vessels with low tortuosity.

blood vessels having similar characteristics as those in the respective training sets. Thus, we henceforth refer to those networks as being *specialized* on blood vessels with high or low tortuosity.

As illustrated in Figure 3B, after training, the network is applied to the test set in order to generate a set of segmented images S_{sl} , where *sl* stands for *specialized in low tortuosity vessels*. From S_{sl} , we calculate the average tortuosity value of each image using the methodology described in Section 2.4. The set of calculated tortuosity values is represented as T_{sl} . For the network specialized on blood vessels with high

tortuosity, we represent the obtained segmented images as S_{sh} and average tortuosity values as T_{sh} , where *sh* stands for *specialized in high tortuosity vessels*.

2.2.2. Fine-Tuning an Off-The-Shelf Network for Blood Vessel Morphometry

The second part of our analysis consists in implementing the process illustrated in Figure 2B. Thus, the networks trained using the procedure described in the previous section are fine-tuned to a new dataset. The approach used for fine-tuning the networks is illustrated in Figure 3C. The figure shows the methodology for the network specialized in low

tortuosity blood vessels, but the procedure is identical for the other network. As depicted in the figure, the network is fine-tuned using a subset of the training set containing images of blood vessels with high tortuosity. The fine-tuning is done for 15 epochs. Moreover, we use the same methodology and hyperparameters employed when creating the specialized networks.

The fine-tuned network is then applied to the test data of the dataset containing vessels with high tortuosity (Figure 3D). A new set of segmented images S_{fh} and average tortuosity values T_{fh} are generated. The subscript fh stands for *fine-tuned with high tortuosity vessels*. As mentioned above, this fine-tuning procedure is also performed for the network specialized in blood vessels with high tortuosity. In this case, we represent the new set of segmented images as S_{fl} and average tortuosity as T_{fl} , where fl stands for *fine-tuned with low tortuosity vessels*.

2.2.3. Quantifying Morphometry Biases

The segmented images and average tortuosity values obtained from the specialized and fine-tuned networks can be used to assess the segmentation quality and identify possible biases generated by the training procedures regarding the tortuosity of the blood vessels. We define the segmentation quality as the IoU between the segmentations obtained by the network fine-tuned on a dataset and those obtained by the network specialized on the same dataset. So, for the blood vessels with low tortuosity:

$$IoU_l = \frac{1}{N} \sum_{i=1}^N IoU(S_{fl}(i), S_{sl}(i)), \quad (1)$$

where $S_{fl}(i)$ and $S_{sl}(i)$ represent, respectively, the i -th image from the sets S_{fl} and S_{sl} , and N is the number of images. The term $IoU(S_{fl}(i), S_{sl}(i))$ in Equation 1 quantifies the IoU between the segmentations obtained by a network that was initially trained on blood vessels with high tortuosity and then fine-tuned on blood vessels with low tortuosity ($S_{fl}(i)$), and the segmentations obtained by a network that was specifically trained for segmenting blood vessels with low tortuosity ($S_{sl}(i)$). Ideally, IoU_l should be close to 1. Notice that this can happen even if the networks cannot segment the blood vessels with good accuracy.

Equivalently, for blood vessels with high tortuosity:

$$IoU_h = \frac{1}{N} \sum_{i=1}^N IoU(S_{fh}(i), S_{sh}(i)). \quad (2)$$

The quality of the tortuosity values obtained by each network can also be measured. For the low tortuosity dataset, we define

$$R_l = \frac{1}{N} \sum_{i=1}^N \frac{T_{fl}(i)}{T_{sl}(i)}. \quad (3)$$

The term $T_{fl}(i)/T_{sl}(i)$ quantifies the average tortuosity obtained using the network that was initially trained on blood vessels with high tortuosity and then fine-tuned on low tortuosity vessels ($T_{fl}(i)$) with respect to the tortuosity value

obtained using the network specialized on low tortuosity blood vessels ($T_{sl}(i)$).

Similarly, for blood vessels with high tortuosity:

$$R_h = \frac{1}{N} \sum_{i=1}^N \frac{T_{fh}(i)}{T_{sh}(i)}. \quad (4)$$

Values R_l and R_h quantify the changes in tortuosity obtained when using an off-the-shelf network instead of training a neural network from scratch. It is important to notice that the off-the-shelf network was not trained on completely unrelated data. The high and low tortuosity datasets have similar characteristics, with their main distinction being the tortuosity of the blood vessels contained in the images. Also, given the natural sinuosity of the blood vessels in our images, we observed in our experiments that the values in T_{sl} and T_{sh} , being averages calculated over the entire images, did not reach values close to zero.

Having defined these quality metrics, we search for biases of segmentation and tortuosity values by calculating the values of relative tortuosity and IoU as we increase the number of annotated images used in the fine-tuning step. The number of images was varied from 0 (no fine-tuning) to 40. Since the annotated images are randomly sampled, the fine-tuning is repeated K times, and each quality metric is calculated as the average for all repetitions. Since using less annotated images leads to large fluctuations in the results, we define K as

$$K = \max\left(5, \frac{t}{2n}\right), \quad (5)$$

where t is the total size of the training set and n the number of images used in the refinement. Thus, the smaller the number of images used in the refinement, the larger is the number of repetitions used in the evaluation. The average value of a quality metric X calculated over K repetitions is henceforth referred to as $\bar{X}$.

A summary of the metrics defined above is presented in Table A1 of the Appendix.

2.3. Data Augmentation and Transfer Learning

Manually labeling blood vessels is a burdensome task. Thus, in many situations, annotating a large number of blood vessels may be prohibitive. Another difficulty with the fine-tuning approach happens when the off-the-shelf network needs to generalize better toward vessels with high tortuosity. In this regard, high tortuosity vessels are usually disease markers. Therefore, it becomes harder to optimize a network to increase its generalization capability since more effort is needed to develop, image, and label samples containing vessels with high tortuosity. Thus, it is beneficial to develop approaches for increasing the generalizability of a network using *as few manual labelings as possible*.

Training CNNs usually requires large amounts of labeled data. This problem motivated the development of several data augmentation techniques over the years. Data augmentation techniques allow a network to learn patterns that are not present in the original data and, consequently, improve its generalizability. We propose a methodology for increasing

the generalization capability of a network specialized in segmenting blood vessels with low tortuosity. This is done by applying data augmentation to the set of low tortuosity vessels using elastic transformations [40], a type of transformation capable of distorting the blood vessels and giving them a more tortuous appearance.

The elastic transformation deforms an image by drawing random displacement fields. Thus, for each image, horizontal (Δ_x) and vertical (Δ_y) displacement fields are defined by drawing random numbers between -1 and +1, generated from a uniform distribution. Then, the fields Δ_x and Δ_y are convolved with a gaussian with standard deviation σ . In this case, σ works as an elasticity coefficient. When $\sigma \approx 0$, the result is an uncorrelated displacement field. For non-zero σ , an elastic deformation effect is obtained. Lastly, the displacement fields Δ_x and Δ_y are multiplied by a scale factor α , which controls the transformation intensity. The elastic transformation can generate realistic-looking highly-tortuous vessels based on images containing vessels with low tortuosity. An example of the application of elastic transformations in our dataset is depicted in Figure 13 of the Supplementary Material. Also, Figure 14 of the Supplementary Material shows typical tortuosity values obtained for different values of parameter α .

It is important to notice that data augmentation can be applied in two distinct situations. The first involves applying data augmentation when training the network from scratch. The second situation concerns applying data augmentation for fine-tuning a pre-trained network.

2.3.1. Data Augmentation When Training From Scratch

To evaluate the influence of data augmentation when training a network from scratch, we define a new set of images based on the application of the elastic transformation to the set of vessels with low tortuosity, which we henceforth refer to as *false highly-tortuous vessels*. Next, we optimize a neural network using the false highly-tortuous vessels and assess the capability of this network in performing the morphometry of the real set of high tortuosity vessels. The training methodology employed is the same as the one used in the optimization of the specialized networks (see Section 2.2.1), including the architecture, hyperparameters, and the number of epochs. After training, the set of blood vessels with high tortuosity is segmented and the set of tortuosity values T_{sf} is calculated. In this case, sf stands for *specialized in false highly-tortuous vessels*. Similarly to Equations 3 and 4, the relative tortuosity in this experiment (R_{sf}) is calculated as the average ratio between each element of T_{sf} and the respective element of T_{sh} obtained by the network specialized in high tortuosity vessels (NSHV), that is,

$$R_{sf} = \frac{1}{N} \sum_{i=1}^N \frac{T_{sf}(i)}{T_{sh}(i)}. \quad (6)$$

Similarly, the relative IoU, IoU_{sf} , is calculated as

$$IoU_{sf} = \frac{1}{N} \sum_{i=1}^N IoU(S_{sf}(i), S_{sh}(i)). \quad (7)$$

2.3.2. Data Augmentation During Fine-Tuning

As mentioned above, data augmentation may also be used when fine-tuning a CNN. Thus, it is interesting to verify if the generalizability of a pre-trained network can be increased by fine-tuning the network using false highly-tortuous vessels. This can be investigated by fine-tuning the network specialized in low tortuosity vessels (NSLV) on false highly-tortuous vessels generated using the elastic transformation with $\alpha = 64$ and $\sigma = 4$. These parameters were chosen because they provided the best performance in our experiments. All other parameters were kept the same as those used in the fine-tuning applied to the NSLV (Section 2.2.2).

2.4. Measuring the Tortuosity of Blood Vessels

An approach for quantifying the tortuosity of blood vessels is required for implementing our experiments. Many different metrics have been defined in the literature [35, 10, 47]. Ramos et al. [35] compared the tortuosity scores provided by five experts regarding blood vessels in retina fundus images. They also compared the scores of the experts with those obtained from different automated approaches for measuring tortuosity. They found a high inter-expert variability as well as different degrees of agreement between the experts and the considered automated measurements. Their results demonstrate that there is no optimal approach for defining tortuosity. Indeed, as discussed in [3], different types of tortuosity can be considered. Therefore, any given tortuosity measurement will have advantages and drawbacks. In this work, we calculate the tortuosity of blood vessels using linear regression residuals².

The tortuosity is obtained as follows. First, the centerlines of the blood vessels are calculated using the Pal gyi-Kuba topological thinning algorithm [30]. Next, we associate a tortuosity value to each pixel of the generated centerlines. As illustrated in Figure 4, given a reference centerline pixel $p_c(i)$ belonging to a vessel segment, we define a circle of radius r centered at $p_c(i)$. The centerline pixels inside this circle define the neighborhood of $p_c(i)$. Then, a line l is fitted to the set of neighboring pixels using least-squares linear regression. The tortuosity of $p_c(i)$ is then defined as the average of the point-to-line distances between the neighboring pixels of $p_c(i)$ and l (dashed lines in Figure 4). In other words, the tortuosity is given by the root mean squared error of the least-squares regression of the neighborhood of $p_c(i)$. If the blood vessel segment around $p_c(i)$ is relatively straight, the respective tortuosity will be small. Contrariwise, if the segment cannot be well represented by a straight line, the tortuosity will be large. The overall tortuosity of the blood vessels in an image is then defined as the average tortuosity of the centerline pixels, that is

²Code available at <https://github.com/chcomin/pyvane>

$$T = \frac{1}{q} \sum_{i=1}^q p_c(i), \quad (8)$$

where q is the number of centerline pixels. Notice that the radius r controls the size of the detected tortuous structures. For smaller values of r , small sinuous structures will have larger tortuosity values. For larger values of r , longer vessels with smoother curvatures will be recognized as more tortuous. Thus, parameter r controls the scale of the analysis. The computational cost of calculating the tortuosity is dominated by the linear regression for large r , which has an asymptotic time complexity of $O(r)$. Therefore, the calculation of the average tortuosity for a single image has a time complexity of $O(rq)$. For the experiments in this work, we set $r = 10$ since it allows the detection of sharp blood vessel turns in the mouse cortex. The calculation of the average tortuosity of all blood vessels in an image with a resolution of 1376×1104 pixels takes around 9 seconds using the computer described in Section 2.6. An example of tortuosity values obtained for a sample can be found in Figure 12 of the Supplementary Material.

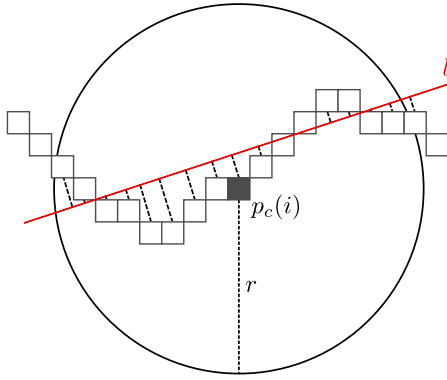


Figure 4: Illustration of the methodology used for calculating the tortuosity of a blood vessel segment. A straight line is adjusted to the points inside a circle of radius r centered at a reference point $p_c(i)$. The tortuosity associated to point $p_c(i)$ is given by the average distance between the line and the points inside the circle.

2.5. Dataset

To apply the methodology proposed in this work, we use a dataset composed of 1838 3D volumes, all obtained from the cerebral cortices of mice. All volumes were collected in 3D using confocal microscopy and have a resolution of $1376 \times 1104 \times 51$ voxels, each voxel representing $0.908\mu\text{m} \times 0.908\mu\text{m} \times 1\mu\text{m}$. The procedure used for obtaining the volumes is described in Ouellette et al. [29].

To obtain the ground truth labels, we used a semi-supervised segmentation approach. This same strategy was previously used in other works [29, 20]. First, a Gaussian filter with unit standard deviation is applied to each 3D volume. Then, vessel regions are identified by an adaptive thresholding algorithm with a window size of $100\mu\text{m} \times 100\mu\text{m}$ applied to each image along the depth of the volume (z-direction). Next,

connected components smaller than $500\mu\text{m}^3$ are removed. Since we are interested in evaluating the performance of 2D CNNs, the maximum intensity projections (MIP) of the volumes (along with their respective segmentations) were used as input to the networks – i.e., 2D images with a resolution of 1376×1104 pixels. Please refer to Ouellette et al. [29] for a more detailed description of the methodology. An example of labeled cortical vasculature is depicted in Figure 5.

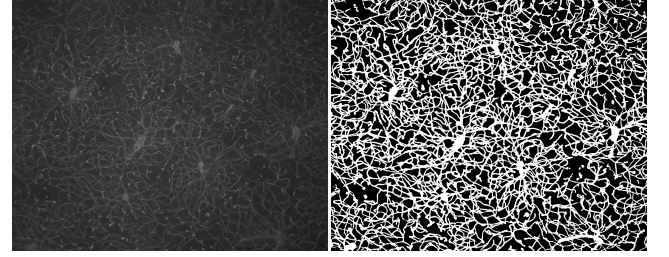


Figure 5: Labeled cortical vasculature obtained by the semi-supervised segmentation approach.

In order to separate the images between sets of high and low tortuosity blood vessels, the average tortuosity value is calculated for each image in the dataset. The 100 images resulting in the highest average tortuosity were labeled as samples containing blood vessels with high tortuosity. Similarly, the 100 images with the lowest average tortuosity were labeled as low tortuosity vessel samples. For performance reasons, each image was divided into 16 windows resulting in 1600 samples for each class with a resolution of 344×276 pixels. We were careful not to include windows from the same image in both training and validation or test sets. All analyses presented in Section 3 were done using these 3200 samples.

2.6. Implementation Details

All experiments were implemented using PyTorch³. The experiments ran on a desktop computer equipped with an Intel i5-10400f 6 core and 12 threads CPU, 16 GB of RAM, and a Nvidia RTX 2060 6GB GPU. The total processing time of all experiments was approximately 34 days.

3. Results and Discussion

3.1. Generation of the Specialized Networks

Two networks were optimized using the methodology described in Section 2.2.1, each specialized in segmenting blood vessels with either low or high tortuosity. After training, the NSLV obtained an average IoU of 0.8882⁴. The NSHV had an average IoU of 0.9158⁵. These results are comparable to the performance obtained by other works in the literature focused on the segmentation of blood vessels imaged by confocal microscopy [45, 43]. Figure 6 shows examples of

³<https://pytorch.org/>

⁴Average Dice of 0.9405

⁵Average Dice of 0.9553

results obtained by the CNNs when compared to the ground truth labels.

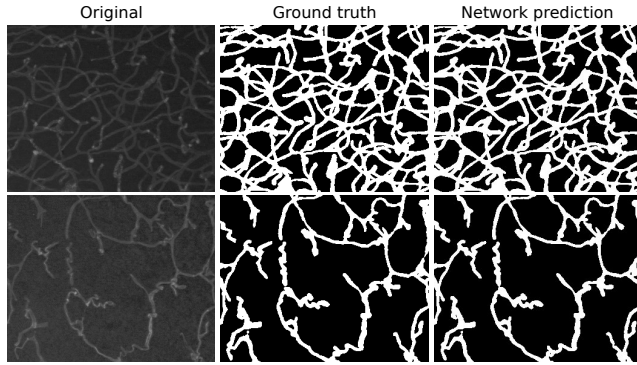


Figure 6: Example of segmentations produced by the specialized networks. Top row: network specialized on low tortuosity vessels applied to an image containing mostly low tortuosity vessels. Bottom row: network specialized on high tortuosity vessels applied to an image containing mostly high tortuosity vessels.

It is worth mentioning that even though our experiments evaluate the segmentation performance using IoU, we optimized our networks using the cross-entropy as loss function. We observed that using the IoU or Dice as loss functions led to a similar performance compared to the cross-entropy, but the cross-entropy led to better convergence.

3.2. Fine-Tuning the Specialized Networks

After optimizing the specialized networks, we examined their ability in segmenting data that they originally were not trained on. When used for segmenting blood vessels with high tortuosity, the NSLV obtained an average IoU of 0.6825. Similarly, when segmenting low tortuosity vessels, the NSHV had an average IoU of 0.6516. Figure 7 shows examples of typical segmentation results. Figure 7B shows the segmentation of a sample containing blood vessels with high tortuosity, obtained using the NSLV. In this case, small connected components that are not observed in the ground truth (Figure 7A) are created. These components represent false positives and thus lead to a high recall and low precision. Such artifacts seem to be associated with the fact that this network tends to classify regions of constant intensity as blood vessels since low tortuosity vessels usually do not present longitudinal intensity discontinuities. Figure 7D shows a typical result obtained by the NSHV when applied to images containing blood vessels that predominantly have low tortuosity. By comparing this segmentation with the ground truth (Figure 7C), a considerable caliber underestimation is observed, that is, the segmented blood vessels are typically thinner than the ground truth. Since the result contains many false negatives, the NSHV leads to high precision and low recall when segmenting images with mostly low tortuosity vessels.

The observed segmentation problems can be considered a typical consequence of applying a pre-trained network to a dataset having different statistics than the dataset used for

training. Nevertheless, we emphasize that in the present study, the dataset differences are not caused by changes in the sample preparation or imaging protocol but by differences in the morphology of the underlying biological system.

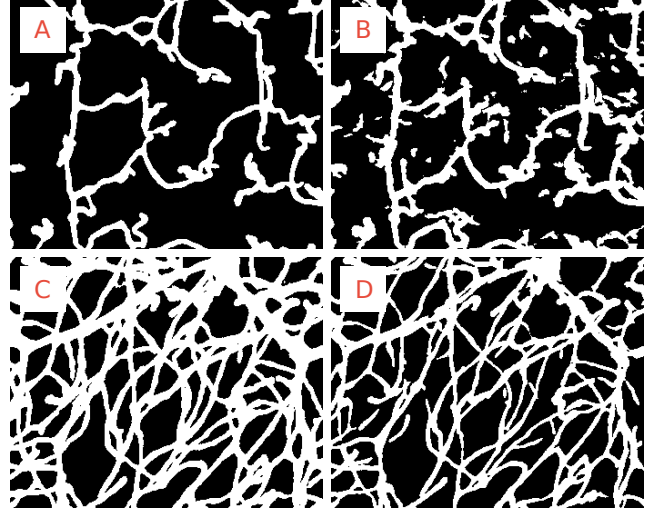


Figure 7: Segmentation artifacts observed when segmenting datasets using inadequate pre-trained networks. (A) Ground truth of a sample containing mostly high tortuosity blood vessels. (B) Segmentation of the sample shown in (A) using a network trained on low tortuosity blood vessels. (C) Ground truth of a low tortuosity blood vessel image. (D) Segmentation of the sample shown in (C) using a network trained on high tortuosity blood vessels.

Having made the initial characterization of the segmentation problems, we proceed to try to revert the observed biases by fine-tuning the specialized networks. Thus, as presented in Section 2.2.2, each specialized network undergoes a refinement step using samples from the dataset where the morphometry analysis is to be performed. Figure 8 shows the performance of the NSHV when segmenting vessels with low tortuosity as a function of the number of images used for fine-tuning. The result shows that the values of IoU_l increase with the number of samples. Still, R_l does not display the same behavior. The tortuosity is overestimated when the fine-tuning step is performed with up to approximately 13 images. When more images are used, R_l stabilizes around 1, while IoU_l slowly increases. It is also interesting to note that IoU_l and R_l are not correlated. Therefore, improvements in a pixelwise accuracy metric, such as the IoU, do not necessarily improve the tortuosity estimation.

Figure 9 shows the relationship between the performance of the NSLV when segmenting vessels with high tortuosity and the number of images used for fine-tuning the network. As refinement progresses and more high tortuosity examples are presented to the network, both values of R_h and IoU_h increase. Also, there is a considerable improvement in the segmentation quality (IoU_h) when just a few samples are used for fine-tuning. In this case, it was possible to achieve an average IoU of 0.9 with approximately 10 images, in contrast to the 40 images used for achieving the same

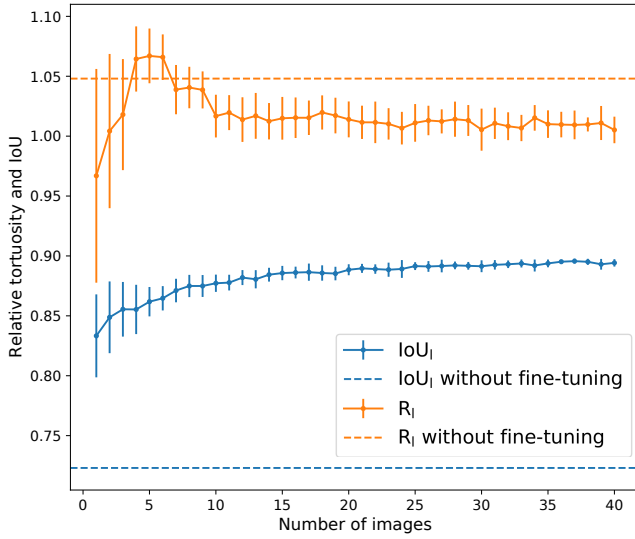


Figure 8: Performance of the network specialized in high tortuosity vessels when segmenting low tortuosity vessels as a function of the number of low tortuosity images used for fine-tuning the network. The plot shows the average values of R_l and IoU_l obtained after repeating the refining experiment K times (defined in Equation 5). Error bars depict the standard deviation of R_l and IoU_l over the K repetitions. The dashed horizontal lines represent the base performance of the NSHV when segmenting low tortuosity vessels without fine-tuning.

performance in the previous experiment (Figure 8). Low tortuosity vessels usually have simple morphologies. By fine-tuning the network to just a few additional samples containing high tortuosity vessels, the network quickly adapted to the new morphological characteristics of the samples. Therefore, it was easier to optimize a network for recognizing more complex vessels, starting from a simpler representation, than to optimize it to simplify a more complex vessel representation.

The results obtained from the two experiments considered in this section show that both networks segmented the out-of-distribution samples with similar accuracy. However, each experiment resulted in different degrees of tortuosity underestimation or overestimation depending on the number of images used for fine-tuning. As observed in Figure 8, without fine-tuning, the NSHV overestimated the average tortuosity of the low tortuosity data by approximately 5% compared to the NSLV performance, given that a value of $R_l = 1.05$ was obtained when no fine-tuning was performed. Similarly, the NSLV underestimated the average tortuosity value of the high tortuosity data by approximately 8% compared to the NSHV performance, as shown in Figure 9. It was observed that these errors could be mitigated by applying a fine-tuning step with 10 to 40 additional images (0.625% to 2.5% of the original training set), depending on the tortuosity of the blood vessels, which led to a more accurate estimation of the tortuosity.

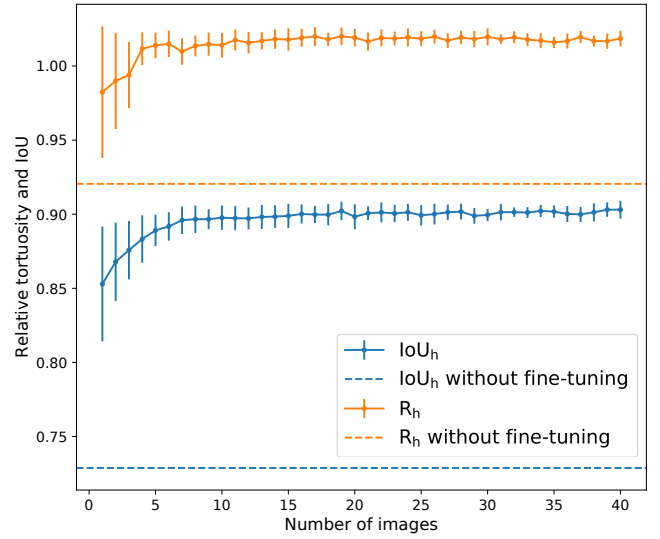


Figure 9: Performance of the network specialized in low tortuosity vessels when segmenting high tortuosity vessels as a function of the number of high tortuosity images used for fine-tuning the network. The plot shows the average values of R_h and IoU_h obtained after repeating the refining experiment K times (defined in Equation 5). Error bars depict the standard deviation of R_h and IoU_h over the K repetitions. The dashed horizontal lines represent the base performance of the NSLV when segmenting high tortuosity vessels without fine-tuning.

3.3. Improving the Tortuosity Estimation Using Data Augmentation

We first verified if data augmentation can improve the estimation of the tortuosity when training the network from scratch. For this experiment, the procedure described in Section 2.3.1 was used. Figure 10 displays R_{sf} and IoU_{sf} as a function of the parameter α used in the elastic transformation. The value of σ was set to 4. We observed that keeping $\sigma = 4$ and varying α between 0 and 70 covered several examples of tortuous blood vessels found in our dataset. There is a considerable drop in R_{sf} when α is between 1 and 10. On the other hand, for $\alpha \geq 12$, training with the set of false highly-tortuous vessels improves the tortuosity estimation. The best performance is observed when α is around 64, where $\bar{R}_{sf} \approx 1$ even though the segmentation performance is relatively low, having an average value of $IoU_{sf} = 0.78$. Another important result is the lack of a strong correlation between R_{sf} and IoU_{sf} for different values of α . This reinforces the idea that improvements in vessel morphometry may be caused by effects that are not properly quantified by pixelwise accuracy metrics.

We also investigate if data augmentation can lead to a better estimation of the tortuosity during fine-tuning. The results of applying the methodology described in Section 2.3.2 are depicted in Figure 11. Contrariwise to what was observed in the results shown in Figures 8, 9 and 10, there was no noticeable improvement on the values of IoU_h and R_h . Even after fine-tuning the network using 40 additional images, the performance of the network was, on average, $\bar{R}_h \approx 0.9$

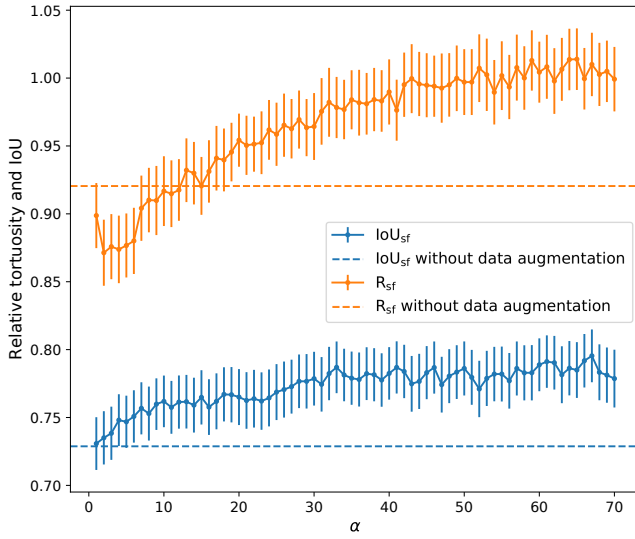


Figure 10: Performance of the network trained with false highly-tortuous vessels when subjected to the morphometry of real blood vessels with high tortuosity. α is the scale factor used in the elastic transformation to obtain the set of false highly-tortuous vessels. Error bars represent the standard errors of the average values of R_{sf} and IoU_{sf} of the test set. The dashed horizontal lines represent the base performance of the NSLV when segmenting high tortuosity vessels.

and $IoU_h \approx 0.8$. Thus, data augmentation was not enough to avoid an underestimation of the tortuosity. This result contrasts with the improvement obtained when applying data augmentation for training the network from scratch.

Some considerations can be made from the obtained result. First, data augmentation was unable to correct the bias of the NSLV after fine-tuning the network. In addition, $\bar{R}_h$ values have large standard deviations, which indicates that the quality of the morphometry obtained by the network is not reliable. Additional approaches could be investigated for fine-tuning the network. Some possibilities include: modifying the learning rate scheduler for finding a solution that better generalizes the data, increasing the number of images used in the fine-tuning step, and using different data augmentation techniques.

4. Discussion

The segmentation and characterization of blood vessels is an important task for diagnosing many types of diseases as well as for studying the development of neurovascular systems. Therefore, defining robust and unbiased approaches for identifying blood vessels in digital images is of great importance. Usually, most of the focus is placed on obtaining good segmentation accuracy, commonly measured using the Dice or IoU metrics. We focused on identifying to which extent segmentation accuracy is related to morphometric precision when measuring the tortuosity of blood vessels.

Two main investigations were proposed. The first was to identify if the tortuosity calculated for blood vessels may

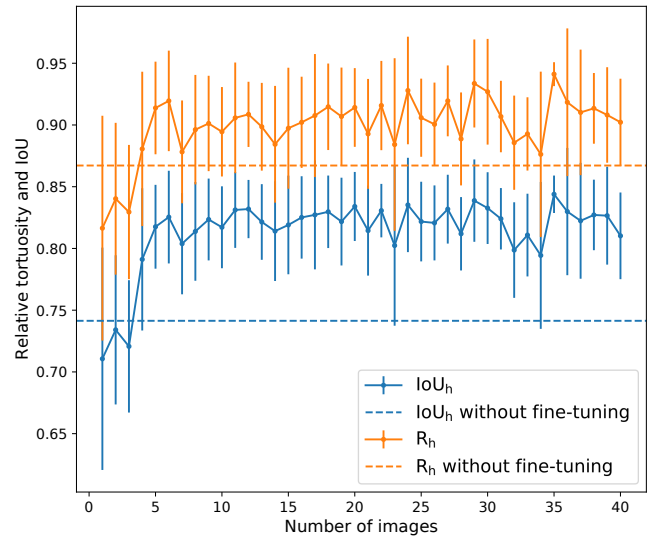


Figure 11: Performance of the network specialized in low tortuosity vessels when segmenting high tortuosity vessels as it goes through a fine-tuning step with increasingly more false highly-tortuous vessels. The plot shows the average values of R_h and IoU_h obtained after repeating the refining experiment K times (defined in Equation 5). Error bars depict the standard deviation of R_h and IoU_h over the K repetitions. The dashed horizontal lines represent the base performance of the NSLV when segmenting high tortuosity vessels without fine-tuning.

change depending on the CNN used for segmenting the blood vessels, even in situations where the dataset used and the segmentation accuracy is similar for the CNNs. A procedure was developed for comparing a CNN that was trained only on images containing mostly high tortuosity blood vessels with a CNN that was trained on low tortuosity vessels and fine-tuned on vessels with high tortuosity. The opposite situation was also investigated. We found that the tortuosity values obtained from a CNN trained from scratch may not agree with those obtained by a CNN that was pre-trained on data acquired from a different experimental condition, and consequently, with slightly different tortuosity statistics. CNNs trained on blood vessels with high tortuosity tended to lead to an overestimation of the tortuosity, while CNNs trained on blood vessels with low tortuosity resulted in an underestimation of the tortuosity.

The second investigation involved the definition of possible procedures for reducing the observed biases. A fine-tuning approach on the dataset used for inference and a data augmentation methodology were considered. We identified that the tortuosity biases could be corrected by applying a fine-tuning step using additional annotated images. On average, around 13 new annotated images were required for removing the biases. Interestingly, the improvement in tortuosity estimation was not correlated with changes in segmentation accuracy when analyzing the low tortuosity dataset.

Regarding data augmentation, we investigated a possible bias-reducing data augmentation approach using elastic

transformations of the training set. This strategy can be helpful for small datasets and also requires fewer manually annotated images. We observed that training the network using the proposed data augmentation approach did not lead to a significant improvement of the IoU metric. However, this approach made the network less biased when used for morphometric quantifications. This means that adding proper data augmentation techniques when training CNNs might be useful even if they do not improve segmentation accuracy. Conversely, when using off-the-shelf networks, it might be useful to give preference to networks that were trained with appropriate data augmentation techniques, even if its accuracy is lower than other state-of-the-art architectures. This result also highlights the importance of correctly optimizing a blood vessel morphometry pipeline for prospective studies regarding vascular systems. The optimization should not consider only pointwise accuracy metrics but also the performance of metrics associated with morphological properties of the blood vessels.

To our knowledge, a similar investigation regarding tortuosity biases has not been performed before in the literature. On the other hand, analyses involving the influence of segmentation on the morphometry of bone trabeculae [31] and brain structures [4, 34, 17] as well as radiotherapy planning [32] have been performed before, where similar results were obtained. For instance, established pipelines for voxel-based morphometry of brain MRIs (Magnetic Resonance Images) can lead to inconsistent results regarding the estimation of age-related gray matter volume decrease [4], brain volume in autism spectrum disorder [17], and gray matter volume atrophy for amyotrophic lateral sclerosis patients [34]. Furthermore, in [32] the authors report that during radiotherapy planning, geometrical similarity scores (such as the Dice coefficient) used to evaluate the segmentation of peripheral (not-targeted) organs, do not correlate well with the changes in the radiation dosage they receive. To mitigate possible morphometry biases, the authors suggested that new evaluation metrics must be developed, in the sense that they should reflect more accurately the underlying clinical problem – in their case, radiation dosage on peripheral organs. In [36] and [6], deep neural networks were trained to directly optimize the morphological features of brain MRIs, rather than the segmentation accuracy. Still, little attention has been given to the influence of network training on morphometry.

For future studies, it would be interesting to consider the influence of segmentation on other morphological quantities, such as blood vessel length and the number of branching points. It would also be interesting to verify if the same results are obtained for other types of blood vessel images such as those acquired by retinal imaging and by angiography. Actually, the same analysis considered in this study can be applied to any system composed of a segmentation step and a downstream quantification task.

5. Conclusion

Going forward, we argue that more focus should be given to the impact of neural network segmentation on downstream tasks related to blood vessel morphometry. Specifically, when developing computerized systems, it is important to take into account that a high, but not perfect, segmentation performance does not necessarily lead to properly segmented images for morphometric analyses. This is especially true for neural networks given that explaining the results provided by such systems is still an active area of research [38, 1]. As shown in our analysis, fine-tuning a pre-trained neural network on new data might not completely remove morphometric biases that may be present in the original network. In such cases, we argue that the network should be retrained from scratch using new data and image augmentation techniques that can directly perturb the morphological measurements that will be calculated.

Funding

Cesar H. Comin thanks FAPESP (grants no. 18/09125-4 and 21/12354-8) for financial support. The authors acknowledge the support of the Government of Canada's New Frontiers in Research Fund (NFRF) (NFRFE-2019-00641) and the Latin America Research Awards (LARA 2021). This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

- [1] J. Adebayo, J. Gilmer, M. Muelly, I. Goodfellow, M. Hardt, B. Kim, Sanity checks for saliency maps, in: S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, R. Garnett (Eds.), *Advances in Neural Information Processing Systems*, volume 31, Curran Associates, Inc., 2018. URL: <https://proceedings.neurips.cc/paper/2018/file/294a8ed24b1ad22ec2e7efea049b8737-Paper.pdf>.
- [2] P. Andreini, G. Ciano, S. Bonechi, C. Graziani, V. Lachi, A. Mecocci, A. Sodi, F. Scarselli, M. Bianchini, A two-stage gan for high-resolution retinal image generation and segmentation, *Electronics* 11 (2022). URL: <https://www.mdpi.com/2079-9292/11/1/60>. doi:10.3390/electronics11010060.
- [3] E. Bullitt, G. Gerig, S.M. Pizer, W. Lin, S.R. Aylward, Measuring Tortuosity of the Intracerebral Vasculature from MRA Images, *IEEE Transactions on Medical Imaging* 22 (2003) 1163–1171. doi:10.1109/TMI.2003.816964.
- [4] D.V. Callaert, A. Ribbens, F. Maes, S.P. Swinnen, N. Wenderoth, Assessing age-related gray matter decline with voxel-based morphometry depends significantly on segmentation and normalization procedures, *Frontiers in Aging Neuroscience* 6 (2014). URL: <https://www.frontiersin.org/articles/10.3389/fnagi.2014.00124>. doi:10.3389/fnagi.2014.00124.
- [5] G. Canton, D.S. Hippe, L. Chen, J.C. Waterton, W. Liu, H. Watase, N. Balu, J. Sun, T.S. Hatsukami, C. Yuan, Atherosclerotic Burden and Remodeling Patterns of the Popliteal Artery as Detected in the

- Magnetic Resonance Imaging Osteoarthritis Initiative Data Set, *Journal of the American Heart Association* 10 (2021) e018408. doi:10.1161/JAHA.120.018408.
- [6] R.S. Cruz, L. Lebrat, P. Bourgeat, V. Doré, J. Dowling, J. Frapp, C. Fookes, O. Salvado, Going deeper with brain morphometry using neural networks, in: 2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI), pp. 711–715. doi:10.1109/ISBI48211.2021.9434039.
- [7] F. Fernandez-Klett, L. Brandt, C. Fernández-Zapata, B. Abuelnor, J. Middeldorp, J.A. Sluijs, M. Curtis, R. Faull, L.W. Harris, S. Bahn, E.M. Hol, J. Priller, Denser brain capillary network with preserved pericytes in Alzheimer's disease, *Brain Pathology* 30 (2020) 1071–1086. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1111/bpa.12897>. doi:https://doi.org/10.1111/bpa.12897.
- [8] E. Frías-Anaya, R. Gromnicova, I. Kraev, V. Rogachevsky, D.K. Male, F. Crea, C.A. Hawkes, I.A. Romero, Age-related ultrastructural neurovascular changes in the female mouse cortex and hippocampus, *Neurobiology of Aging* 101 (2021) 273–284. URL: <https://www.sciencedirect.com/science/article/pii/S0197458020304176>. doi:https://doi.org/10.1016/j.neurobiolaging.2020.12.008.
- [9] I. Goodfellow, Y. Bengio, A. Courville, *Deep Learning*, MIT Press, 2016.
- [10] E. Grisan, M. Foracchia, A. Ruggeri, A novel method for the automatic grading of retinal vessel tortuosity, *IEEE Transactions on Medical Imaging* 27 (2008) 310–319. doi:10.1109/TMI.2007.904657.
- [11] Z. Gu, J. Cheng, H. Fu, K. Zhou, H. Hao, Y. Zhao, T. Zhang, S. Gao, J. Liu, CE-Net: Context Encoder Network for 2D Medical Image Segmentation, *IEEE Transactions on Medical Imaging* 38 (2019) 2281–2292. doi:10.1109/TMI.2019.2903562.
- [12] S. Guzel, C.L. Cai, T. Ahmad, M. Quan, G.B. Valencia, J.V. Aranda, K.D. Beharry, Bumetanide Suppression of Angiogenesis in a Rat Model of Oxygen-Induced Retinopathy, *International Journal of Molecular Sciences* 21 (2020). URL: <https://www.mdpi.com/1422-0067/21/3/987>. doi:10.3390/ijms21030987.
- [13] H.C. Han, Twisted blood vessels: Symptoms, etiology and biomechanical mechanisms, *Journal of Vascular Research* 49 (2012) 185–197. doi:10.1159/000335123.
- [14] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778.
- [15] Z. Jiang, H. Zhang, Y. Wang, S.B. Ko, Retinal blood vessel segmentation using fully convolutional network with transfer learning, *Computerized Medical Imaging and Graphics* 68 (2018) 1–15. URL: <https://www.sciencedirect.com/science/article/pii/S0895611118302313>. doi:https://doi.org/10.1016/j.compmedimag.2018.04.005.
- [16] Q. Jin, Z. Meng, T.D. Pham, Q. Chen, L. Wei, R. Su, DUNet: A deformable network for retinal vessel segmentation, *Knowledge-Based Systems* 178 (2019) 149–162. URL: <http://www.sciencedirect.com/science/article/pii/S0950706705119301984>. doi:https://doi.org/10.1016/j.knosys.2019.04.025.
- [17] G.J. Katuwal, S.A. Baum, N.D. Cahill, C.C. Dougherty, E. Evans, D.W. Evans, G.J. Moore, A.M. Michael, Inter-method discrepancies in brain volume estimation may drive inconsistent findings in autism, *Frontiers in Neuroscience* 10 (2016). URL: <https://www.frontiersin.org/articles/10.3389/fnins.2016.00439>. doi:10.3389/fnins.2016.00439.
- [18] D.P. Kingma, J. Ba, Adam: A method for stochastic optimization, 2017. arXiv:1412.6980.
- [19] C. Kirst, S. Skriabine, A. Vieites-Prado, T. Topilko, P. Bertin, G. Gerschenfeld, F. Verny, P. Topilko, N. Michalski, M. Tessier-Lavigne, N. Renier, Mapping the Fine-Scale Organization and Plasticity of the Brain Vasculature, *Cell* 180 (2020) 780 – 795.e25. URL: <http://www.sciencedirect.com/science/article/pii/S0092867420301094>. doi:https://doi.org/10.1016/j.cell.2020.01.028.
- [20] B. Lacoste, C.H. Comin, A. Ben-Zvi, P.S. Kaeser, X. Xu, L. da F. Costa, C. Gu, Sensory-related neural activity regulates the structure of vascular networks in the cerebral cortex, *Neuron* 83 (2014) 1117 – 1130. URL: <http://www.sciencedirect.com/science/article/pii/S089662731400645X>. doi:https://doi.org/10.1016/j.neuron.2014.07.034.
- [21] D. Lesage, E.D. Angelini, I. Bloch, G. Funka-Lea, A review of 3D vessel lumen segmentation techniques: Models, features and extraction schemes, *Medical Image Analysis* 13 (2009) 819–845. URL: <http://dx.doi.org/10.1016/j.media.2009.07.011>. doi:10.1016/j.media.2009.07.011.
- [22] T. Li, W. Bo, C. Hu, H. Kang, H. Liu, K. Wang, H. Fu, Applications of deep learning in fundus images: A review, *Medical Image Analysis* 69 (2021) 101971. URL: <https://www.sciencedirect.com/science/article/pii/S1361841521000177>. doi:https://doi.org/10.1016/j.media.2021.101971.
- [23] P. Liskowski, K. Krawiec, Segmenting Retinal Blood Vessels With Deep Neural Networks, *IEEE Transactions on Medical Imaging* 35 (2016) 2369–2380.
- [24] M. Livne, J. Rieger, O.U. Aydin, A.A. Taha, E.M. Akay, T. Kossen, J. Sobesky, J.D. Kelleher, K. Hildebrand, D. Frey, V.I. Madai, A U-Net Deep Learning Framework for High Performance Vessel Segmentation in Patients With Cerebrovascular Disease, *Frontiers in Neuroscience* 13 (2019) 97. URL: <https://www.frontiersin.org/article/10.3389/fnins.2019.00097>. doi:10.3389/fnins.2019.00097.
- [25] S.J. Ma, M.S. Sarabi, L. Yan, X. Shao, Y. Chen, Q. Yang, K. Jann, A.W. Toga, Y. Shi, D.J. Wang, Characterization of lenticulostriate arteries with high resolution black-blood T1-weighted turbo spin echo with variable flip angles at 3 and 7 Tesla, *NeuroImage* 199 (2019) 184–193. doi:10.1016/j.neuroimage.2019.05.065.
- [26] K.K. Maninis, J. Pont-Tuset, P. Arbeláez, L. Van Gool, Deep retinal image understanding, in: S. Ourselin, L. Joskowicz, M.R. Sabuncu, G. Unal, W. Wells (Eds.), *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016*, Springer International Publishing, Cham, 2016, pp. 140–148.
- [27] F.J. Martinez-Murcia, A. Ortiz, J. Ramirez, J.M. Górriz, R. Cruz, Deep residual transfer learning for automatic diagnosis and grading of diabetic retinopathy, *Neurocomputing* 452 (2021) 424–434. URL: <https://www.sciencedirect.com/science/article/pii/S095231220316520>. doi:https://doi.org/10.1016/j.neucom.2020.04.148.
- [28] S. Moccia, E. De Momi, S. El Hadji, L.S. Mattos, Blood vessel segmentation algorithms — Review of methods, datasets and evaluation metrics, *Computer Methods and Programs in Biomedicine* 158 (2018) 71–91. URL: <https://www.sciencedirect.com/science/article/pii/S0169260717313421>. doi:https://doi.org/10.1016/j.cmpb.2018.02.001.
- [29] J. Ouellette, X. Toussay, C.H. Comin, L.d.F. Costa, M. Ho, M. Lacalle-Aurioles, M. Freitas-Andrade, Q.Y. Liu, S. Leclerc, Y. Pan, Z. Liu, J.F. Thibodeau, M. Yin, M. Carrier, C.J. Morse, P.V. Dyken, C.J. Bergin, S. Baillet, C.R. Kennedy, M.E. Tremblay, Y.D. Benoit, W.L. Stanford, D. Burger, D.J. Stewart, B. Lacoste, Vascular contributions to 16p11.2 deletion autism syndrome modeled in mice, *Nature Neuroscience* 23 (2020) 1090–1101. URL: <https://doi.org/10.1038/s41593-020-0663-1>. doi:10.1038/s41593-020-0663-1.
- [30] K. Palágyi, A. Kuba, A 3D 6-subiteration thinning algorithm for extracting medial lines, *Pattern Recognition Letters* 19 (1998) 613–627. URL: <http://www.sciencedirect.com/science/article/pii/S0167865598000312>. doi:https://doi.org/10.1016/S0167-8655(98)00031-2.
- [31] I.H. Parkinson, A. Badiei, N.L. Fazzalari, Variation in segmentation of bone from micro-ct imaging: implications for quantitative morphometric analysis, *Australasian Physics & Engineering Sciences in Medicine* 31 (2008) 160–164. URL: <https://doi.org/10.1007/BF03178592>. doi:10.1007/BF03178592.
- [32] R. Poel, E. Rüfenacht, E. Hermann, S. Scheib, P. Manser, D.M. Aebersold, M. Reyes, The predictive value of segmentation metrics on dosimetry in organs at risk of the brain, *Medical Image Analysis* 73 (2021) 102161. URL: <https://www.sciencedirect.com/science/article/pii/S1361841521002073>. doi:https://doi.org/10.1016/j.media.2021.102161.

- [33] M. Potente, H. Gerhardt, P. Carmeliet, Basic and Therapeutic Aspects of Angiogenesis, *Cell* 146 (2011) 873–887. URL: <https://www.sciencedirect.com/science/article/pii/S0092867411010099>. doi:<https://doi.org/10.1016/j.cell.2011.08.039>.
- [34] V. Rajagopalan, G.H. Yue, E.P. Pioro, Do preprocessing algorithms and statistical models influence voxel-based morphometry (vbm) results in amyotrophic lateral sclerosis patients? a systematic comparison of popular vbm analytical methods, *Journal of Magnetic Resonance Imaging* 40 (2014) 662–667. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/jmri.24415>. doi:<https://doi.org/10.1002/jmri.24415>. arXiv:<https://onlinelibrary.wiley.com/doi/pdf/10.1002/jmri.24415>.
- [35] L. Ramos, J. Novo, J. Rouco, S. Romeo, M.D. Álvarez, M. Ortega, Retinal vascular tortuosity assessment: inter-intra expert analysis and correlation with computational measurements, *BMC medical research methodology* 18 (2018) 1–11.
- [36] M. Rebsamen, Y. Suter, R. Wiest, M. Reyes, C. Rummel, Brain morphometry estimation: From hours to seconds using deep learning, *Frontiers in Neurology* 11 (2020). URL: <https://www.frontiersin.org/articles/10.3389/fneur.2020.00244>. doi:[10.3389/fneur.2020.00244](https://doi.org/10.3389/fneur.2020.00244).
- [37] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in: N. Navab, J. Hornegger, W.M. Wells, A.F. Frangi (Eds.), *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, Springer International Publishing, Cham, 2015, pp. 234–241.
- [38] C. Rudin, Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead, *Nature Machine Intelligence* 1 (2019) 206–215.
- [39] S. Shit, J.C. Paetzold, A. Sekuboyina, I. Ezhov, A. Unger, A. Zhylyka, J.P. Pluim, U. Bauer, B.H. Menze, cldice-a novel topology-preserving loss function for tubular structure segmentation, in: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 16560–16569.
- [40] P. Simard, D. Steinkraus, J. Platt, Best practices for convolutional neural networks applied to visual document analysis, in: *Seventh International Conference on Document Analysis and Recognition*, 2003. *Proceedings.*, pp. 958–963. doi:[10.1109/ICDAR.2003.1227801](https://doi.org/10.1109/ICDAR.2003.1227801).
- [41] L.N. Smith, A disciplined approach to neural network hyperparameters: Part 1 – learning rate, batch size, momentum, and weight decay, 2018. arXiv:1803.09820.
- [42] J. Staal, M.D. Abràmoff, M. Niemeijer, M.A. Viergever, B. Van Ginneken, Ridge-based vessel segmentation in color images of the retina, *IEEE transactions on medical imaging* 23 (2004) 501–509.
- [43] W. Tahir, S. Kura, J. Zhu, X. Cheng, R. Damseh, F. Tadesse, A. Seibel, B.S. Lee, F. Lesage, S. Sakadžić, D.A. Boas, L. Tian, Anatomical Modeling of Brain Vasculature in Two-Photon Microscopy by Generalizable Deep Learning, *BME Frontiers* 2021 (2021) 8620932. URL: <https://doi.org/10.34133/2021/8620932>. doi:[10.34133/2021/8620932](https://doi.org/10.34133/2021/8620932).
- [44] G. Tetteh, V. Efremov, N.D. Forkert, M. Schneider, J. Kirschke, B. Weber, C. Zimmer, M. Piraud, B.H. Menze, DeepVesselNet: Vessel Segmentation, Centerline Prediction, and Bifurcation Detection in 3-D Angiographic Volumes, *Frontiers in Neuroscience* 14 (2020) 1285. URL: <https://www.frontiersin.org/article/10.3389/fnins.2020.592352>. doi:[10.3389/fnins.2020.592352](https://doi.org/10.3389/fnins.2020.592352).
- [45] M.I. Todorov, J.C. Paetzold, O. Schoppe, G. Tetteh, S. Shit, V. Efremov, K. Todorov-Völgyi, M. Düring, M. Dichgans, M. Piraud, B. Menze, A. Ertürk, Machine learning analysis of whole mouse brain vasculature, *Nature Methods* 17 (2020) 442–449. URL: <https://doi.org/10.1038/s41592-020-0792-1>. doi:[10.1038/s41592-020-0792-1](https://doi.org/10.1038/s41592-020-0792-1).
- [46] Y. Tongpob, S. Xia, C. Wyrwoll, A. Mehnert, Quantitative characterization of rodent feto-placental vasculature morphology in micro-computed tomography images, *Computer Methods and Programs in Biomedicine* 179 (2019) 104984. URL: <https://doi.org/10.1016/j.cmpb.2019.104984>. doi:[10.1016/j.cmpb.2019.104984](https://doi.org/10.1016/j.cmpb.2019.104984).
- [47] C.M. Wilson, K.D. Cocker, M.J. Moseley, C. Paterson, S.T. Clay, W.E. Schulenburg, M.D. Mills, A.L. Ells, K.H. Parker, G.E. Quinn, A.R. Fielder, J. Ng, Computerized analysis of retinal vessel width and tortuosity in premature infants, *Investigative Ophthalmology and Visual Science* 49 (2008) 3577–3585. doi:[10.1167/iovs.07-1353](https://doi.org/10.1167/iovs.07-1353).
- [48] X. Xiao, S. Lian, Z. Luo, S. Li, Weighted Res-UNet for High-Quality Retina Vessel Segmentation, in: *2018 9th International Conference on Information Technology in Medicine and Education (ITME)*, pp. 327–331. doi:[10.1109/ITME.2018.00080](https://doi.org/10.1109/ITME.2018.00080).
- [49] Y. Zhang, A.C.S. Chung, Deep Supervision with Additional Labels for Retinal Vessel Segmentation Task, in: A.F. Frangi, J.A. Schnabel, C. Davatzikos, C. Alberola-López, G. Fichtinger (Eds.), *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018*, Springer International Publishing, Cham, 2018, pp. 83–91.
- [50] H. Zhao, H. Li, S. Maurer-Stroh, L. Cheng, Synthesizing retinal and neuronal images with generative adversarial nets, *Medical Image Analysis* 49 (2018) 14–26. URL: <https://www.sciencedirect.com/science/article/pii/S1361841518304596>. doi:<https://doi.org/10.1016/j.media.2018.07.001>.
- [51] H. Zhao, H. Li, S. Maurer-Stroh, Y. Guo, Q. Deng, L. Cheng, Supervised segmentation of un-annotated retinal fundus images by synthesis, *IEEE Transactions on Medical Imaging* 38 (2019) 46–56. doi:[10.1109/TMI.2018.2854886](https://doi.org/10.1109/TMI.2018.2854886).

Appendix: Performance Metrics Used in This Work

Table A1

Symbols and corresponding definitions of the metrics used for quantifying the performance of blood vessel segmentation and tortuosity calculation.

Metric	Description
S_{sl}	Set of segmented images obtained by the application of the network specialized in low tortuosity vessels on images containing low tortuosity vessels.
T_{sl}	Average tortuosity values of S_{sl} .
S_{sh}	Set of segmented images obtained by the application of the network specialized in high tortuosity vessels on images containing high tortuosity vessels.
T_{sh}	Average tortuosity values of S_{sh} .
S_{fl}	Set of segmented images obtained by the application of the network specialized in high tortuosity vessels and fine-tuned on low tortuosity vessels on images containing low tortuosity vessels.
T_{fl}	Average tortuosity values of S_{fl} .
S_{fh}	Set of segmented images obtained by the application of the network specialized in low tortuosity vessels and fine-tuned on high tortuosity vessels on images containing high tortuosity vessels.
T_{fh}	Average tortuosity values of S_{fh} .
IoU_l	Average IoU between the set of segmented images S_{fl} and S_{sl} .
R_l	Average ratio between the set of tortuosities T_{fl} and T_{sl} .
IoU_h	Average IoU between the set of segmented images S_{fh} and S_{sh} .
R_h	Average ratio between the set of tortuosities T_{fh} and T_{sh} .
S_{sf}	Set of segmented images obtained by the application of the network specialized in false highly-tortuous vessels on images containing high tortuosity vessels.
T_{sf}	Average tortuosity values of S_{sf} .
IoU_{sf}	Average IoU between the set of segmented images S_{sf} and S_{sh} .
R_{sf}	Average ratio between the set of tortuosities T_{sf} and T_{sh} .
$\bar{X}$	Average value of metric X over K repetitions of an experiment, where K is defined in Equation 5.

Appendix: Supplementary Material

6. Vessel Tortuosity

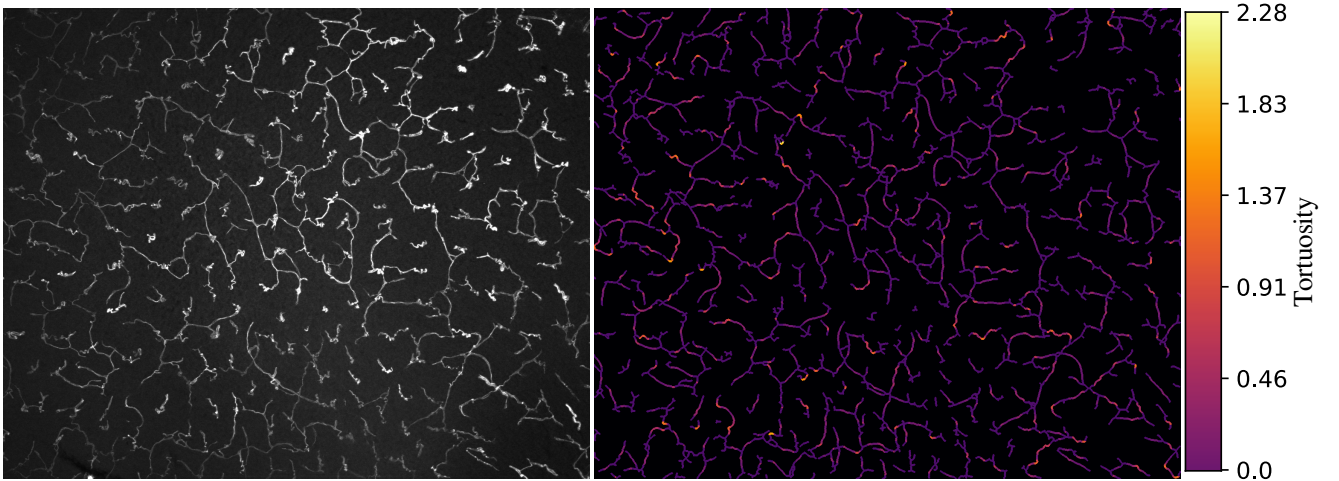


Figure 12: Pixelwise tortuosity obtained for the blood vessels in one of the images of the dataset.

7. Elastic Transformations

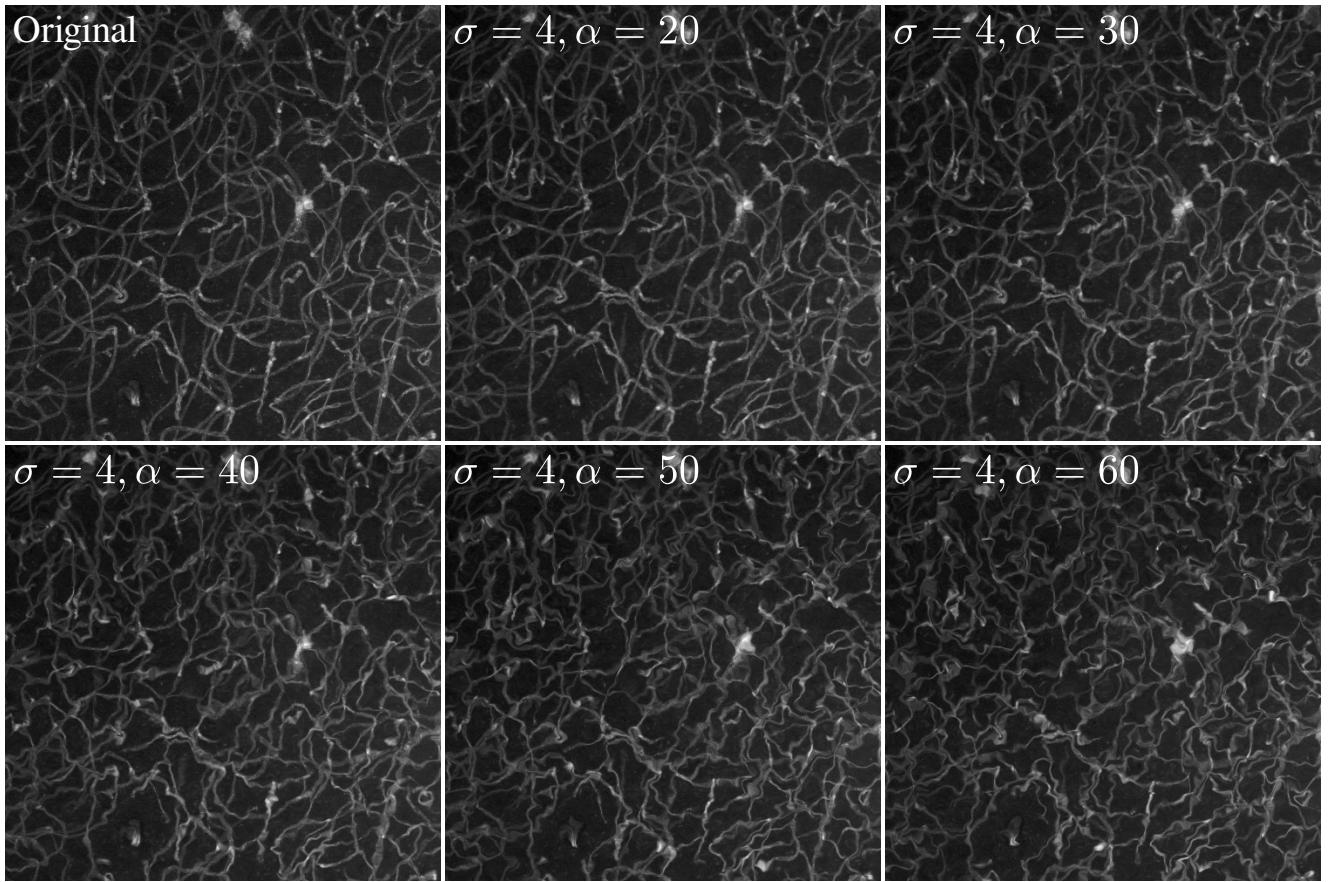


Figure 13: Examples of elastic transformations applied to an image depicting the mouse brain vasculature. The larger the value of α , the more tortuous each vessel segment appears to be.

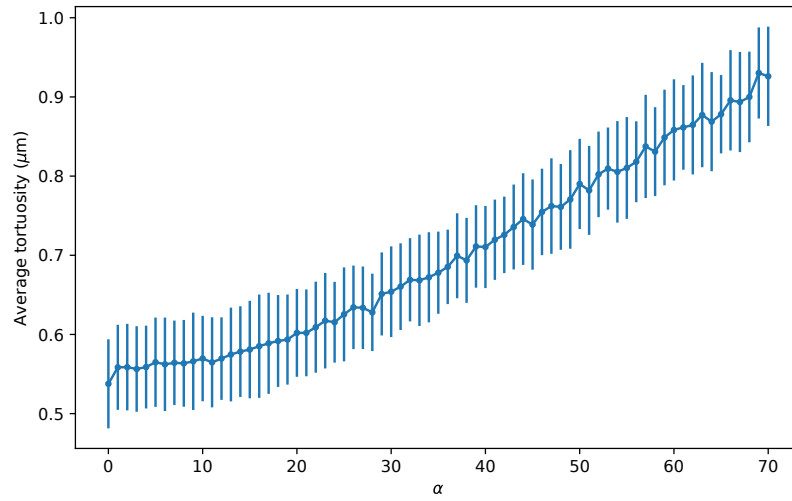


Figure 14: Average tortuosity values of the set of low tortuosity vessels after the application of elastic transformations with different values of α . Each point corresponds to the average tortuosity of 50 images randomly sampled from the 1600 images of the dataset containing low tortuosity vessels. The error bars depict the standard deviation of the tortuosity calculated over the 50 images. The same set of images was used for all values of α .