

DeepVesselNet: Vessel Segmentation, Centerline Prediction, and Bifurcation Detection in 3-D Angiographic Volumes

Giles Tetteh^a, Velizar Efremov^{a,d}, Nils D. Forkert^c, Matthias Schneider^{e,d}, Jan Kirschke^b, Bruno Weber^d, Claus Zimmer^b, Marie Piraud^a, Björn H. Menze^a

^a*Department of Computer Science, TU München, München, Germany.*

^b*Neuroradiology, Klinikum Rechts der Isar, TU München, München, Germany.*

^c*Department of Radiology, University of Calgary, Calgary, Canada.*

^d*Institute of Pharmacology and Toxicology, University of Zurich, Zurich, Switzerland.*

^e*Computer Vision Laboratory, ETH Zurich, Zurich, Switzerland*

Abstract

We present DeepVesselNet, an architecture tailored to the challenges to be addressed when extracting vessel networks and corresponding features in 3-D angiography using deep learning. We discuss the problems of low execution speed and high memory requirements associated with full 3-D convolutional networks, high class imbalance arising from low percentage (less than 3%) of vessel voxels, and unavailability of accurately annotated training data - and offer solutions that are the building blocks of DeepVesselNet.

First, we formulate 2-D orthogonal cross-hair filters which make use of 3-D context information. Second, we introduce a class balancing cross-entropy score with false positive rate correction to handle the high class imbalance and high false positive rate problems associated with existing loss functions. Finally, we generate synthetic dataset using a computational angiogenesis model, capable of generating vascular networks under physiological constraints on local network structure and topology, and use these data for transfer learning.

DeepVesselNet is optimized for segmenting vessels, predicting centerlines, and localizing bifurcations. We test the performance on a range of angiographic volumes including clinical Time-of-Flight MRA data of the human brain, as

Email address: giles.tetteh@tum.de (Giles Tetteh)

well as synchrotron radiation X-ray tomographic microscopy scans of the rat brain. Our experiments show that, by replacing 3-D filters with 2-D orthogonal cross-hair filters in our network, speed is improved by 23% while accuracy is maintained. Our class balancing metric is crucial for training the network and pre-training with synthetic data helps in early convergence of the training process.

Keywords: vessel segmentation, centerline prediction, bifurcation detection, deepvesselnet, cross-hair filters, class balancing, synthetic data.

1. Introduction

Angiography offers insights into blood flow and conditions of vascular network. Three dimensional volumetric angiography information can be obtained using magnetic resonance (MRA), ultrasound, or x-ray based technologies like computed tomography (CT). A common first step in analyzing these data is vessel segmentation. Still, moving from raw angiography images to vessel segmentation alone might not provide enough information for clinical use, and other vessel features like centerline, diameter, or bifurcations of the vessels are also needed to accurately extract information about the vascular network. In this work, we present a deep learning approach, called DeepVesselNet, to perform vessel segmentation, centerline prediction, and bifurcation detection tasks. DeepVesselNet deals with challenges that result from speed and memory requirements, unbalanced class labels, and the difficulty of obtaining well-annotated data for curvilinear volumetric structures.

Vessel Segmentation. Vessel enhancement and segmentation is a longstanding task in medical image analysis (see reviews by Kirbas and Quek, 2004, Lesage et al., 2009). The range of methods employed for vessel segmentation reflect the development of image processing during the past decades, some examples including region growing techniques (Martínez-Pérez et al., 1999), active contours (Nain et al., 2004), statistical and shape models (Chung and Noble, 1999, Liao et al., 2013, Moreno et al., 2013, Young et al., 2001), particle filtering

(Dalca et al., 2011, Florin et al., 2006, Wörz et al., 2009) and path tracing (Wang et al., 2013). All of these examples are interactive, starting from a set of seed label as root and propagating towards the branches. Other approaches aim at an unsupervised enhancement of vascular structures: Frangi et al. (1998) examined the multi scale second order local structure of an image (Hessian) with the purpose of developing a vessel enhancement filter. A measure of vessel-likeness is then obtained as a function of all eigenvalues of the Hessian. Law and Chung (2008) proposed a novel curvilinear structure detector, called Optimally Oriented Flux (OOF). OOF finds an optimal axis on which image gradients are projected to compute the image gradient flux. OOF has a lower computational load than the calculation of the Hessian matrix proposed in Frangi et al. (1998). Forkert et al. (2013, 2011) presented and evaluated a level-set segmentation approach with vesselness-dependent anisotropic energy weights, in 3-D time-of-flight (TOF) MRA. Phellan and Forkert (2017) presented a comparative analysis of the accuracy gains in vessel segmentation generated by the use of nine vessel enhancement algorithms on time-of-flight MRA that included multi scale vesselness algorithms, diffusion-based filters, and filters that enhance tubular shapes. A machine learning approach was followed by Schneider et al. (2015), combining joint 3-D vessel segmentation and centerline extraction using oblique Hough forest with steerable filters. In a similar fashion, Ciresan et al. (2012) used deep artificial neural network as a pixel classifier to automatically segment neuronal structures in stacks of electron microscopy images, a task somewhat similar to vessel segmentation. One example using deep learning architecture is, Phellan et al. (2017) who used a deep convolutional neural network to automatically segment the vessels of the brain in TOF MRA by extracting manually annotated bi-dimensional image patches in the axial, coronal, and sagittal directions as an input to the training process.

Centerline Prediction. Identifying the center of a vessel is relevant for calculating the vessel diameter, but also for extracting the 'skeleton' of a vessel when extracting the vascular network (see Fig. 1). The vessels' skeleton and center

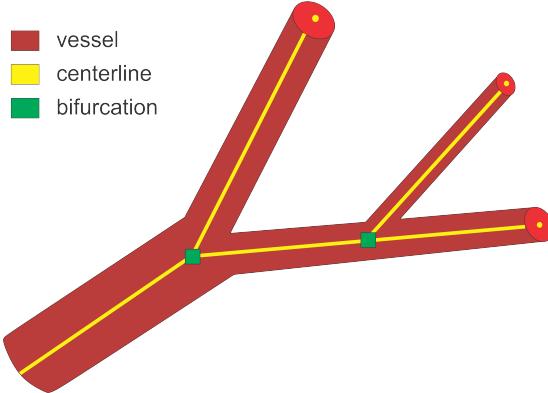


Figure 1: An overview of the three main tasks tackled in this paper. For bifurcations, we predict a neighbourhood cube around the indicated point.

can be found by post-processing a previously generated vessel segmentation or deal with centerline extraction in raw images with high vessel contrast. Shagufa et al. (2014) developed a method based on morphological operations, by performing erosion using 2×2 neighborhoods of a pixel to determine if a pixel is a centerline candidate. Maddah et al. (2003) applied the idea of active contour models as well as path planning and distance transforms for extracting centerline in vessels, and Chen and Cohen (2015) proposed a geodesic or minimal path technique. Santamaría-Pang et al. (2007) performed centerline extraction using a morphology-guided level set model by learning the structural patterns of a tubular-like object, and estimating the centerline of a tubular object as the path with minimal cost with respect to outward flux in gray level images. Zheng et al. (2012) adopted vesselness filters to predict the location of the centerline, while Macedo et al. (2010) used Houghs transforms. Schneider et al. (2015, 2012) designed Hough random forest with local image filters to predict the centerline, and trained on centerline data previously extracted using one of the level set approaches. The Application of deep learning to the extraction of vessel centerline has not been explored. One reason may be the lack of annotated data necessary to train deep architectures that is hard to obtain especially in 3-D datasets.

Bifurcation Detection. Vessel bifurcation refers to the point on a vessel centerline where the vessel splits into two or more smaller vessels (see Fig. 1). Bifurcations represent the nodes of the vascular network and knowing their locations is important both for network extraction and for studying its properties (Rempfler et al., 2015). They represent structures that can easily be used as landmarks in image registration, but also indicate the locations of modified blood flow velocity and pressure within the network itself (Chaichana et al., 2017). Bifurcations are hard to detect in volumetric data, – as they are rare point-like features varying in size and shape significantly. Similar to centerline extraction, the detection of bifurcations often happens by post-processing a previously generated vessels segmentation or by searching a previously extracted vessel graph. Zheng et al. (2015) proposed a two staged deep learning architecture for detecting carotid artery bifurcations as a specific landmark in volumetric CT data by first training a shallow network for predicting candidate regions followed by a sparse deep network for final prediction. Chaichana et al. (2017) proposed a three stage algorithm for detecting bifurcations in digital eye fundus images, a 2-D task, and their approach included image enhancement, clustering, and searching the graph for bifurcations. The direct predicting of the location of bifurcations in a full volumetric data set is a task which has – to the best of our knowledge – not been attempted yet. Same to centerline prediction, the lack of annotated training data is limiting the use of learning based approaches.

Convolutional Neural Networks for Angiography Analysis. The use of Convolutional Neural Networks (CNNs) for image segmentation has seen rapid progress during the past years. Originally aiming at predicting global labels or scores. Ciresan et al. (2012) employed CNNs for pixel-wise classification using as input the raw intensity values of a square window centered on the pixel of interest. Unfortunately, these patch-wise architectures suffer from replication of information in memory, which limits the amount of training data that can be used during training and affecting speed during testing. The problems were addressed in Long et al. (2014), Ronneberger et al. (2015) through inverting the down-

sampling and convolutional operations in CNN with upsampling and deconvolutional operators leading to a Fully Convolutional Neural Network (FCNN) which is trained end-to-end for semantic segmentation. Based on this idea different variants of state-of-the-art FCNN have been presented for medical image segmentation (Christ et al., 2017, Maninis et al., 2016, Milletari et al., 2016, Nogues et al., 2016, Roth et al., 2016, Sekuboyina et al., 2017, Tetteh et al., 2017). Most of these architectures were either employed on 2-D images or extended to 3-D volumes in a fashion which leads to loss of 3-D object information or to high memory needs. For example, Christ et al. (2017) applied the U-NET architecture proposed by Ronneberger et al. (2015) on 3-D data in a 2-D slice-wise fashion (so called 2.5D CNNs), which does not account for inter-slice information. At the same time, full 3-D architectures, such as the V-Net by Milletari et al. (2016), come with the cost of increasing the number of parameters to be learned during training by at least a factor of three over 2-D architectures of similar depth.

Challenges and Contributions. In this work, we address vessel segmentation, centerline prediction, and bifurcation detection as a learning problem, predicting each of the three labels directly from volumetric angiography data. We employ a 3-D CNN architecture, DeepVesselNet, that is optimized with respect to structure, as well as training procedures to deal with this type of image data. Specifically, DeepVesselNet addresses the following three key limitations for using CNN in the three tasks described above:

First, processing 3-D medical volumes poses a memory consumption and speed challenge. Using 3-D CNNs leads to drastic increase in number of parameters and computations compared to 2-D CNNs. At the same time, applying a 2-D CNN in a slice-wise fashion discards valuable 3-D context information that is crucial for tracking curvilinear structures in 3-D. Inspired by the ideas of Rigamonti et al. (2013) who proposed separable 2-D filters, Roth et al. (2014) who used three intersecting planes as 2-D input channels, and the triple-crossing idea of Liu et al. (2017), we demonstrate the use of cross-hair filters from three

intersecting 2-D filters, which helps to avoid the memory and speed problems of classical 3-D networks, while at the same time making use of 3-D information in volumetric data. Unlike the existing ideas where 2-D planes are extracted at a pre-processing stage and used as input channels, our cross-hair filters are implemented on a layer level which help retain the 3-D information throughout the network.

Second, the vessel, centerline and bifurcation prediction tasks is characterized by high class imbalances. Vessels account for less than 3% of the total voxels in a patient volume, centerlines represent a fraction of the segmented vessels, and visible bifurcations are often in the hundreds at best, even when dealing with volumes with 10^6 and more voxels. This bias towards the background class is a common problem in medical data (Grzymala-Busse et al., 2004, Christ et al., 2017, Haixiang et al., 2017). Unfortunately current class balancing loss functions for training CNNs turns out to be numerically unstable in extreme cases as ours. To this end, we offer a new class-balancing loss function that we demonstrate to work well with our vascular features of interest.

Third, manually annotating vessels, centerlines, and bifurcations requires many hours of work and expertise. To this end, we make use of simulation based frameworks (Schneider et al., 2012, Szczerba and Székely, 2005) that can be used for generating synthetic data with accurate labels for pre-training our networks, rendering the training of our supervised classification algorithm feasible.

2. Methodology

In Section 2.1, we formulate the cross-hair filters, which we design to replace the classical 3-D convolutional operator in our network. We then proceed with class balancing in Section 2.2, and discuss the numerical instability of class balancing loss function and the alternatives we propose. In Section 2.3, we provide a brief overview of the synthetic data generation and preparation process. Finally, Section 2.4 describes the network architectures in which we integrate

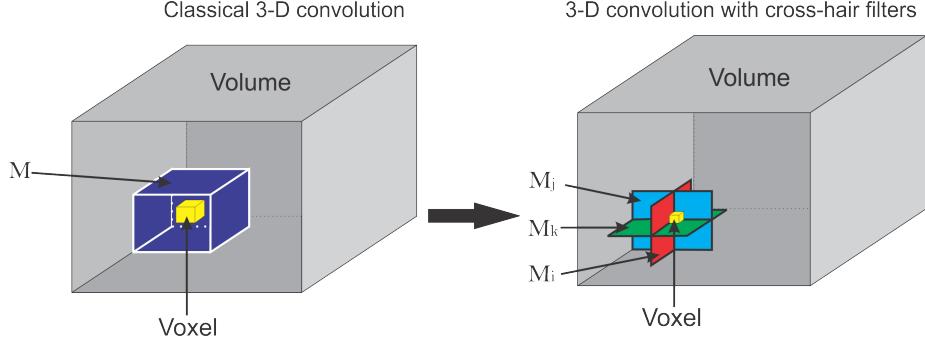


Figure 2: Graphical representation of cross-hair filters for 3-D convolutional operation. Left: A classical 3-D convolution with filter M . Right: Cross-hair 3-D convolutional with 2-D filter stack M_i, M_j, M_k .

the cross-hair filters to perform vessel segmentation, centerline prediction, and bifurcation detection tasks.

2.1. Cross-hair Filters Formulation

In this section, we discuss the formulation of the 3-D convolutional operator, which utilizes cross-hair filters to improve speed and memory usage while maintaining accuracy. Let I be a 3-D volume, M a 3-D convolutional kernel of shape (k_x, k_y, k_z) , and $*$ be a convolutional operator. We define $*$ as:

$$I * M = A = [a_{ijk}]; \quad a_{ijk} = \sum_{r=1}^{k_x} \sum_{s=1}^{k_y} \sum_{t=1}^{k_z} I_{(R,S,T)} M_{(r,s,t)}; \quad (1)$$

$$R = i + r - 1, \quad S = j + s - 1, \quad T = k + t - 1.$$

From equation (1), we see that a classical 3-D convolution involves $k_x k_y k_z$ multiplications and $k_x k_y k_z - 1$ additions for each voxel of the resulting image. For a $3 \times 3 \times 3$ kernel, we have 27 multiplications and 26 additions per voxel. Changing the kernel size to $5 \times 5 \times 5$ increases the complexity to 125 multiplications and 124 additions per voxel. This then scales up with the dimension of the input image. For example, a volume of size $128 \times 128 \times 128$ and a $5 \times 5 \times 5$ kernel results in about 262×10^6 multiplications and 260×10^6 additions. To

handle this increased computational complexity, we approximate the standard 3-D convolution operation by

$$\begin{aligned} a_{ijk} &= \alpha \sum_{s=1}^{k_y} \sum_{t=1}^{k_z} I_{(i,S,T)} M_{(s,t)}^i + \beta \sum_{r=1}^{k_x} \sum_{t=1}^{k_z} I_{(R,j,T)} M_{(r,t)}^j \\ &\quad + \gamma \sum_{r=1}^{k_x} \sum_{s=1}^{k_y} I_{(R,S,k)} M_{(r,s)}^k, \end{aligned} \quad (2)$$

where α, β, γ are weights given to the axial, sagittal, coronal planes respectively, and M^i, M^j, M^k are 2-D cross-hair filters. Using cross-hair filters results in $(k_y k_z + k_x k_z + k_x k_y)$ multiplications and $(k_y k_z + k_x k_z + k_x k_y - 1)$ additions. If we let k_{m1}, k_{m2}, k_{m3} be the sizes of the kernel M such that $k_{m1} \geq k_{m2} \geq k_{m3}$, we can show that

$$k_y k_z + k_x k_z + k_x k_y \leq 3(k_{m1} k_{m2}) \leq k_x k_y k_z, \quad (3)$$

where strict inequality holds for all $k_{m3} > 3$. Equation 3 shows a better scaling in speed and also in memory since the filters sizes in (1) and (2) are affected by the same inequality. With the approximation in (2), and using the same example as above (volume of size $128 \times 128 \times 128$ and a $5 \times 5 \times 5$ kernel), we now need less than 158×10^6 multiplications and 156×10^6 additions to compute the convolution leading to a reduction in computation by more than 100×10^6 multiplications and additions when compared to a classical 3-D convolution. Increasing the volume or kernel size, further increases the gap between the computational complexity of (1) and (2). Moreover, we will see later from our experiments that (2) still retains essential 3-D context information needed for the classification task.

Efficient Implementation. In equation (2), we presented our 2-D crosshair filters. However, applying (2) independently for each voxel (as defined in equation (2)) leads to a redundant use of memory. More precisely, voxels close to each other share some neighbourhood information and making multiple copies of it is not memory efficient. We now present an efficient implementation, which we develop and use in our experiments (Figure 3). Consider I as defined in

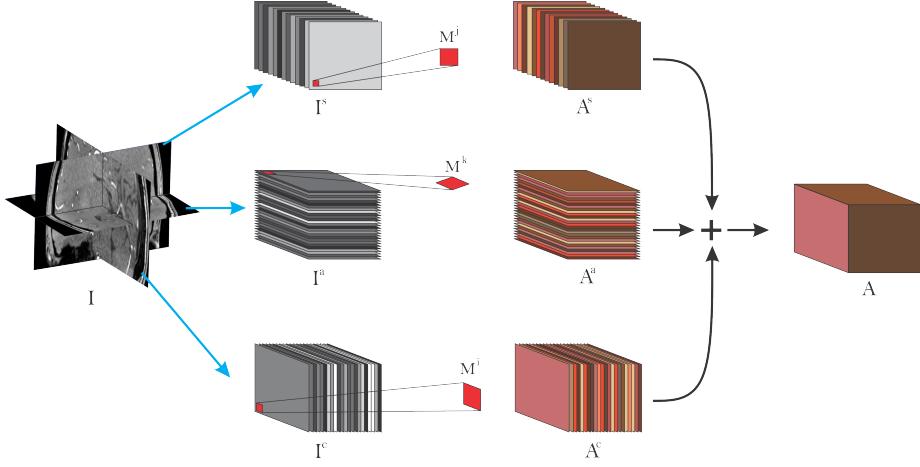


Figure 3: Pictorial view of efficient implementation of cross-hair filters. Grayscale stacks refer to input to the layer, red shaped squares refer to 2-D kernels used for each plane. Brown colored slices refer to extracted features after convolution operations.

Equation (1) and let us extract the sagittal, coronal, and axial planes as I^s, I^c , and I^a respectively. By application of equations (1), and (2), we have a final implementation as follows:

$$\begin{aligned}
 I \diamond M &= A = \alpha A^c + \beta A^s + \gamma A^a, \\
 A^c &= I^c \ast \ast M^i, \\
 A^s &= I^s \ast \ast M^j, \\
 A^a &= I^a \ast \ast M^k,
 \end{aligned} \tag{4}$$

where $\ast\ast$ refers to a 2-D convolution along the first and second axes of the left hand side matrix over all slices in the third axis and \diamond refers to our crosshair filter operation. This implementation is efficient in the sense that it makes use of one volume at a time instead of copies of the volume in memory where voxels share the same neighbourhood. In other words, we still have only one volume in memory but rather rotate the kernels to match the slices in the different orientations. This lowers the memory requirements during training and inference, allowing to train on more data with little memory.

2.5-D Networks vs. 3-D Networks with Cross-hair Filters. We end the presentation of cross-hair filters by discussing the difference between existing 2.5-D networks and our proposed cross-hair filters. Given a 3-D task (e.g. vessel segmentation in 3-D volume) a 2.5-D based network handles the task by considering one 2-D slice at a time. More precisely, the network takes a 2-D slice as input and classifies all pixels in this slice. This is repeated for each slice in the volume and the final results from the slices are fused again to form the 3-D result. Other 2.5-D methods include a pre-processing stage where several 2-D planes are extracted and used as input channels to the 2-D network (Liu et al., 2017, Roth et al., 2014). On the architecture level, 2.5-D networks are 2-D networks with a preprocessing method for extracting 2-D slices and a postprocessing method for fusing 2-D results into a 3-D volume. We note that the predictions of 2.5-D networks are solely based on 2-D context information. Example of 2.5-D networks is the implementation of U-Net in Christ et al. (2017) used for liver and lesion segmentation tasks in CT volumetric dataset and the network architecture of Sekuboyina et al. (2017) for annotation of lumbar vertebrae.

On the other hand, 3-D networks based on our proposed cross-hair filters take the whole 3-D volume as input and at each layer in the network, we apply the convolutional operator discussed in Section 2.1. Therefore, our filters make use of 3-D context information at each convolutional layer and do not require specific preprocessing or post processing. Our proposed method differs from classical 3-D networks in the sense that it uses less parameters and memory since it does not use full 3-D convolutions. However, it is worth noting that our filters scale exactly the same as 2.5-D (i.e. in only two directions) with respect to changes in filter and volume sizes. More precisely, given a square or cubic filter of size k , we have k^2 parameters in a 2.5-D network and $3k^2$ in our cross-hair filter based network. Increasing the filter size by a factor of r will scale up as $k + r$ quadratically in both situations (i.e. $(k + r)^2$ for 2.5-D and $3(k + r)^2$ in cross-hair filter case) as compared to full 3-D networks where the parameter size scales as a cube of $k + r$.

Unlike the existing 2.5-D ideas where 2-D planes are extracted at a pre-

processing stage and used as input channels to a 2-D network architecture, our cross-hair filters are implemented on a layer level which help retain the 3-D information throughout the network making it a preferred option when detecting curvilinear objects in 3-D.

2.2. Class Balancing Loss Function with Stable Weights

Often in medical image analysis, the object of interest (e.g. vessel, tumor etc.) accounts for a minority of the total voxels of the image. Figure 4 shows that the objects of interest in the datasets used in this work account for less than 2.5% of the voxels (the different datasets are described in Section 3.1). Using a standard cross entropy loss function given by

$$\begin{aligned} C &= \frac{1}{N} \sum_{j=1}^N y_j \log P(y_j = 1|X; \mathbf{W}) + (1 - y_j) \log [1 - P(y_j = 1|X; \mathbf{W})], \\ C &= \frac{1}{N} \left(\sum_{y \in Y_+} \log P(y = 1|X; \mathbf{W}) + \sum_{y \in Y_-} \log P(y = 0|X; \mathbf{W}) \right), \end{aligned} \quad (5)$$

where y_j is the label for the j th example, X is the feature set, \mathbf{W} is the set of parameters of the network, Y_+ is the set of positive labels, and Y_- is the set of negative (background) labels, could cause the training process to be biased towards detecting background voxels at the expense of the object of interest. This normally results in predictions with high precision against low recall. To remedy this problem, Hwang and Liu (2015) proposed a biased sampling loss function for training multi scale convolutional neural networks for a contour detection task. This loss function introduced additional trade-off parameters and then samples twice more edge patches than non-edge ones for positive cost-sensitive fine-tuning, and vice versa, for negative cost-sensitive fine-tuning. Based on this, Xie and Tu (2015) proposed a class-balancing cross entropy loss function of the form

$$\mathcal{L}(\mathbf{W}) = -\beta \sum_{j \in Y_+} \log P(y_j = 1|X; \mathbf{W}) - (1 - \beta) \sum_{j \in Y_-} \log P(y_j = 0|X; \mathbf{W}), \quad (6)$$

where \mathbf{W} denotes the standard set of parameters of the network, which are trained with backpropagation and β and $1 - \beta$ are the class weighting multipliers,

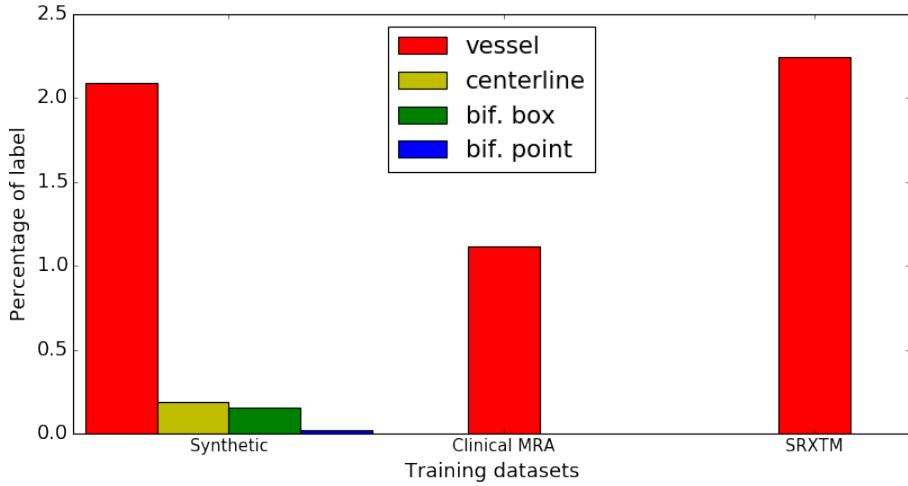


Figure 4: Distribution of labels in the datasets used for experiments in this work. The distribution shows that the objects of interest account for less than 2.5% of the total data points (voxels) in the training set. Description of datasets is given in Section 3.1

which are calculated as $\beta = \frac{|Y_-|}{|Y|}$, $1 - \beta = \frac{|Y_+|}{|Y|}$. $P(\cdot)$'s are the probabilities from the final layer of the network, and Y_+ and Y_- are the set of positive and negative class labels respectively. This idea, which is to give more weight to the cost associated with the class with the lowest count, has been used in other recent works (Christ et al., 2017, Maninis et al., 2016, Nogues et al., 2016, Roth et al., 2016). However, our experiments (in Section 3.3) show that the above loss function raises two main challenges:

Numerical Instability. The gradient computation is numerically unstable for very big training sets due to the high values taken by the loss. More precisely, there is a factor of $\frac{1}{N}$, that scales the final sum to the mean cost in the standard cross-entropy loss function in Equation (5). This factor ensures that the gradients are stable irrespective of the size of the training data N . However, in Equation (6), the weights β and $1 - \beta$ do not scale the cost to the mean. For high values of $|Y|$ (usually the case of voxel-wise tasks), the sums explode leading to numerical instability. For example, given a perfectly balanced data,

we have $\beta = 1 - \beta = 0.5$, irrespective of the value of $|Y|$. Thus, increasing the size of the dataset (batch size) has no effect on the weights (β). However, the number of elements in the sums increases, causing the computations to be unstable.

High False Positive Rate. We observe a high rate of false positives leading to high recall values. This is caused by the fact that in most cases the object of interest accounts for less than 5% of the total voxels (about 2.5 % in our case). Therefore, we have a situation where $1 - \beta < 0.05$, which implies that wrongly predicting 95 background voxels as foreground is less penalized in the loss than predicting 5 foreground voxels as background. This leads to high false positive rate and, hence, high recall values.

To address these challenges, we introduce different weighting ratios and an additional factor to take care of the high false positive rate; and define:

$$\begin{aligned}
\mathcal{L}(\mathbf{W}) &= \mathcal{L}_1(\mathbf{W}) + \mathcal{L}_2(\mathbf{W}) & (7) \\
\mathcal{L}_1(\mathbf{W}) &= -\frac{1}{|Y_+|} \sum_{j \in Y_+} \log P(y_j = 1 | X; \mathbf{W}) - \frac{1}{|Y_-|} \sum_{j \in Y_-} \log P(y_j = 0 | X; \mathbf{W}) \\
\mathcal{L}_2(\mathbf{W}) &= -\frac{\gamma_1}{|Y_+|} \sum_{j \in Y_{f+}} \log P(y_j = 0 | X; \mathbf{W}) - \frac{\gamma_2}{|Y_-|} \sum_{j \in Y_{f-}} \log P(y_j = 1 | X; \mathbf{W}) \\
\gamma_1 &= 0.5 + \frac{1}{|Y_{f+}|} \sum_{j \in Y_{f+}} |P(y_j = 0 | X; \mathbf{W}) - 0.5| \\
\gamma_2 &= 0.5 + \frac{1}{|Y_{f-}|} \sum_{j \in Y_{f-}} |P(y_j = 1 | X; \mathbf{W}) - 0.5|
\end{aligned}$$

\mathcal{L}_1 is a more numerically stable version of equation 6 since it computes the voxel-wise, cost which scales well with the size of the dataset or batch. But the ratio of β to $1 - \beta$ is maintained as desired. \mathcal{L}_2 (FP Rate Correction) is introduced to penalize the network for false predictions. However, we do not want to give false positive (Y_{f+}) and false negatives (Y_{f-}) the same weight as total predictions (Y_+, Y_-), since we will end up with a loss function without any class balancing because the weights will offset each other. Therefore, we introduce γ_1 and γ_2 , which depend on the mean absolute distance of the wrong predicted probabilities to 0.5 (the value can be changed to suit the task). This

allows us to penalize false predictions, which are very far from the central point (0.5). Experimental results from application of FP rate correction can be found in Section 3.3.

2.3. Synthetic Data Generation and Preparation

To generate synthetic data, we follow the method of Schneider et al. (2012) which considers the mutual interplay of arterial oxygen (O_2) supply and vascular endothelial growth factor (VEGF) secreted by ischemic cells to achieve physiologically plausible results. Each vessel segment is modeled as a rigid cylindrical tube with radius r and length l . It is represented by a single directed edge connecting two nodes. Semantically, this gives rise to four different types of nodes, namely root, leaf, bifurcation, and inter nodes. Each node is uniquely identified by the 3-D coordinate $\vec{P} = (x, y, z)^T$. Combining this with connectivity information, fully captures the geometry of the approximated vasculature. Radius of parent bifurcation branch r_p , and the radius of left (r_l) and right (r_r) daughter branches are related by a bifurcation law (also known as Murray's law) given by $r_p^\gamma = r_l^\gamma + r_r^\gamma$, where γ is the bifurcation exponent. Further constraints:

$$\cos(\phi_l) = \frac{r_p^4 + r_l^4 - r_r^4}{2r_p^2 r_l^2}, \quad \cos(\phi_r) = \frac{r_p^4 + r_r^4 - r_l^4}{2r_p^2 r_r^2}, \quad (8)$$

are placed on the bifurcation angles of the left (ϕ_l) and right (ϕ_r) respectively, which geometrically corresponds to the optimal position of the branching point \vec{P}_b with respect to a minimum volume principle (Schneider et al., 2012). The tree generation model and the bifurcation configuration is shown in Figure 5. In the arterial tree generation experiment, the parameters in Table 1 of Schneider et al. (2012) are used. We use the default (underlined) values for all model parameters and generate 136 volumes.

The output of the generation process is a tree with information on the 3-D position \vec{P} of the nodes, their type (root, bifurcation, inter, leaf), and connectivity information, which includes the edge E_{ij} between two nodes N_i and N_j , and its radius R_{ij} . We used this information to construct the actual 3-D volume by modeling each vessel segment as a cylinder in 3-D space. Vessel intensities are

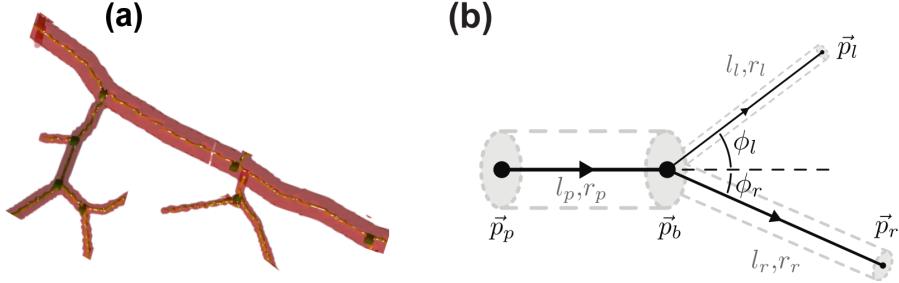


Figure 5: Sample of generated tree (a) with red, yellow, and green representing vessel, centerline, and bifurcation, respectively. (b) represents the constrained bifurcation configuration, as presented in Schneider et al. (2012), where l_p , l_r , and l_l are the length of the parent, right daughter, and left daughter segments, respectively. P_r and P_l are the right and left daughter nodes, respectively.

randomly chosen in the interval [128, 255] and non-vessel intensities are chosen from the interval [0 – 100]. Gaussian noise is then applied to the generated volume randomly changing the mean (i.e. in the range [−5, 5]) and the standard deviation (i.e. in the range [−15, 30]) for each volume.

2.4. Network Architecture and Implementations

We test the performance of DeepVesselNet discussed in sections 2.1, 2.2 and 2.3 through two main implementations:

DeepVesselNet-FCN (DVN-FCN). We construct a Fully Convolutional Network FCN with four convolutional layers and a sigmoid classification layer. In this implementation, we do not use any down-sampling layer and we carry out the convolutions in a way that the output image is of the same size as the input image by zero-padding. The removal of the down-sampling layer is motivated by the fact that the tasks (vessel segmentation, centerline prediction, and bifurcation detection) involve fine detailed objects and down-sampling has an effect of averaging over voxels which causes these fine details to be lost. With this network implementation, we have a very simple 5-layer fully-convolutional network, which takes a volume of arbitrary size and outputs a segmentation map

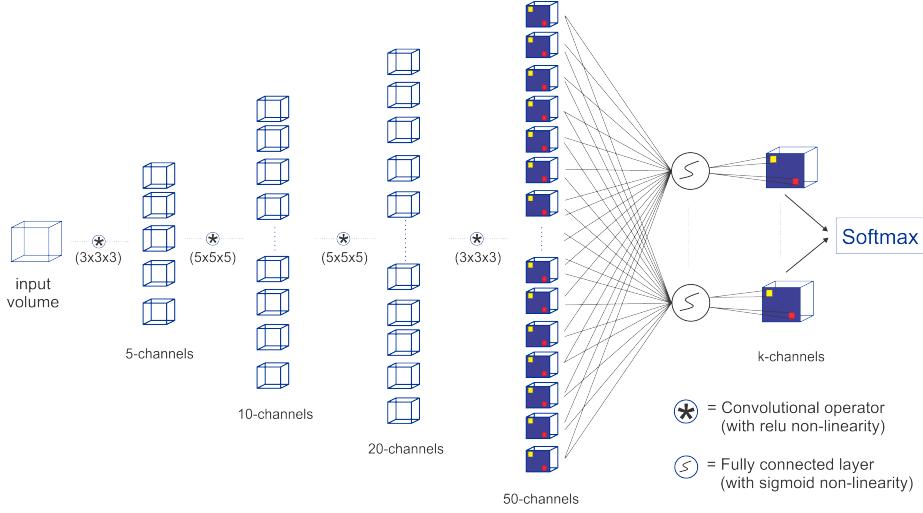


Figure 6: Our proposed DeepVesselNet-FCN architecture implementation with crosshair filters.

of the same size. For the network structure and a description of the parameters, see Figure 6.

DeepVesselNet-VNet (DVN-VNet). As a proof of principle, we take the V-Net architecture proposed by Milletari et al. (2016) (see Figure 7) and replace all 3-D convolutions with our proposed cross-hair filters discussed in section 2.1. The aim is to test the improvement in memory requirements as well as the speed up that our cross-hair implementation provides over the original architecture proposed in Milletari et al. (2016). We also use it to evaluate whether speed and memory consumption have a significant effect on prediction accuracy.

Network Configuration, Initialization, and Training. We use the above described architecture to implement three binary networks for vessel segmentation, centerline prediction, and bifurcation detection. Network parameters are randomly initialized, according to the method proposed in Bengio and Glorot (2010), by sampling from a uniform distribution in the interval $(-\frac{1}{\sqrt{k_x k_y k_z}}, \frac{1}{\sqrt{k_x k_y k_z}})$ where $(k_x \times k_y \times k_z)$ is the size of the given kernel in a particular layer. For each vol-

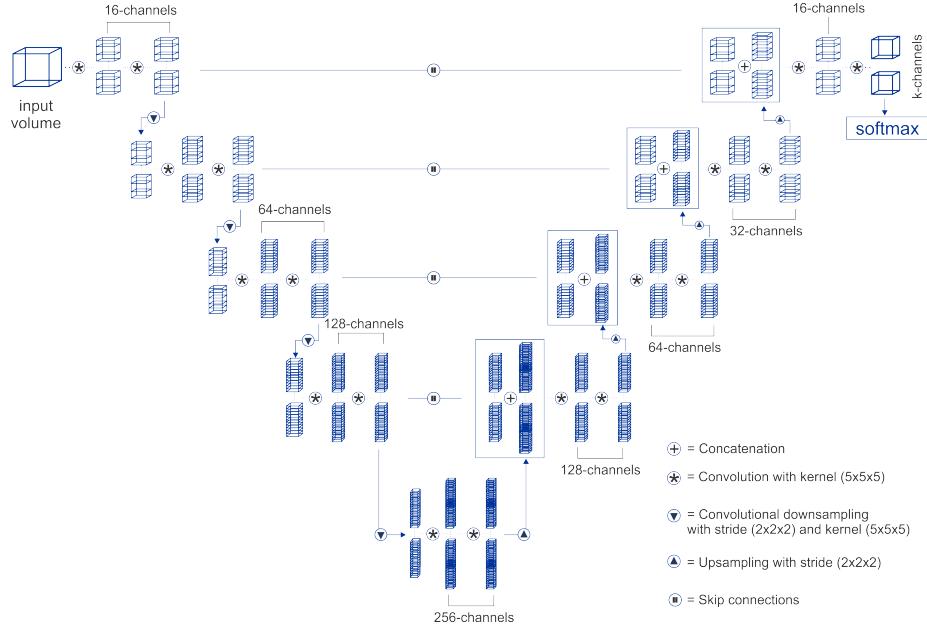


Figure 7: Our DeepVesselNet-VNet architecture implementation. We replace full 3-D filters in the V-Net architecture proposed by Milletari et al. (2016) with cross-hair filters and use it for comparison against the proposed DeepVesselNet-FCN

ume, we extract boxes of suitable sizes [e.g. $(64 \times 64 \times 64)$ or $(128 \times 128 \times 128)$] covering the whole volume and then feed them through the network for the fine-tuning of parameters. After this, we then train the network using a stochastic gradient descent without regularization. During pre-training, we use a learning rate of 0.01 and decay of 0.99, which is applied after every 200 iterations. For fine-tuning, we use a learning rate of 0.001 and a decay of 0.99 applied after every 200 iterations. We implement our algorithm using the THEANO (Theano Development Team, 2016) Python framework and train on a machine with 64GB of RAM and Nvidia TITAN X 12GB GPU.

3. Experiments and Results

3.1. Datasets

In this work, we use three different datasets to train and test the networks. In all three data sets, the test cases are kept apart from the training data and are used only for testing purposes.

Synthetic Dataset. Training convolutional networks from scratch typically requires significant amounts of training data. However, assembling a properly labeled dataset of 3-D curvilinear structures, such as vessels and vessel features, takes a lot of human effort and time, which turns out to be the bottleneck for most medical applications. To overcome this problem, we generate synthetic data based on the method proposed in Schneider et al. (2012). A brief description of this process has already been presented in Section 2.3. We initialize the processes with different random seeds and scale the resulting vessel sizes in voxels to match the sizes of vessels in clinical datasets. After this, we add different levels of Gaussian noise, as mentioned earlier in Section 2.3 to increase the randomness and to make it more realistic. We generate 136 volumes of size $325 \times 304 \times 600$ with corresponding labels for vessel segmentation, centerlines, and bifurcation detection. We then select twenty volumes out of the 136 as a test set for the pre-training phase and use the remaining volumes for pre-training in the various tasks at hand. An example of the synthetic dataset can be found in Figure 8(c).

Clinical MRA Time-of-Flight (TOF) Dataset. To fine-tune and test our network architectures on real data, we obtain 40 volumes of clinical TOF MRA, 20 of which are fully annotated and the remaining 20 partially annotated using the method proposed by Forkert et al. (2013). Each volume has a size of $580 \times 640 \times 136$ and spacial resolution of $0.3125mm \times 0.3125mm \times 0.6mm$ on the coronal, sagittal, and axial axes respectively. We select 15 out of the 20 fully annotated volumes for testing and use the remaining five as a validation set. We also correct the 20 partially annotated volumes by manually verifying some of the

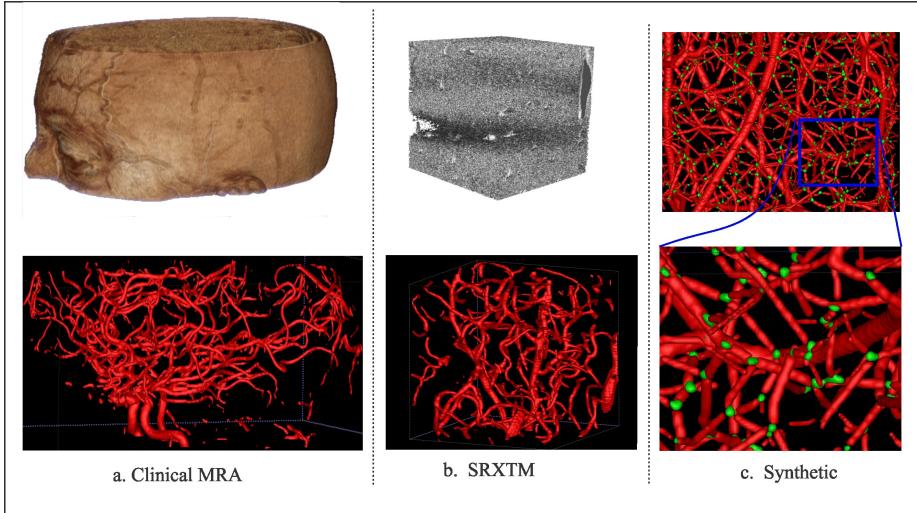


Figure 8: Sample of datasets used in our experiments with the corresponding ground truth segmentations

background and foreground voxels. This leads to three labels, which are true foreground (verified foreground), true background (verified background), and the third class, which represent the remaining voxels not verified. After this, we use the true foreground and background labels to fine-tune our network after pre-training with the synthetic dataset. This approach helps in avoiding any uncertainty with respect to using the partially annotated data for fine-tuning of the network. A sample of volume from the TOF MRA dataset can be found in Figure 8(a).

Synchrotron Radiation X-ray Tomographic Microscopy (SRXTM). A 3-D volume of size $2048 \times 2048 \times 2740$ and spacial resolution $0.7mm \times 0.7mm \times 0.7mm$ is obtained from synchrotron radiation X-ray tomographic microscopy of a rat brain. From this large volume, we extract a dataset of 20 non-overlapping volumes of size $256 \times 256 \times 256$, which were segmented using the method proposed by Schneider et al. (2015) and use them to fine-tune the network. To create a test set, we manually annotate 52 slices in 4 other volumes (208 slices in total).

Detailed description of the SRXTM data can be found in Reichold et al. (2009), and a sample volume is presented in Figure 8(b).

3.2. Data Preprocessing with Intensity Projection

In our experiments, we use different datasets that come from different sources and acquisition modalities. Therefore, we need to normalize their intensity ranges and contrast. We test the following preprocessing strategies to achieve homogeneity between the datasets.

First, the original intensities are normalized to the range [0, 1] using $f(x) = \frac{x - \min(X)}{\max(X) - \min(X)}$ where x is the pixel intensity and X denotes the range of all intensities in the volume. The second strategy involves clipping the intensity values by $g(x) = \{c, x > c; \quad x, x \leq c\}$ and then normalizing the intensities by $f(x)$. Our experiments show that a value of $c = 190$ is optimal for the intensity range of the datasets. The final strategy builds on the second preprocessing strategy by clipping by $g(x)$, normalizing by $f(x)$, and then projecting the resulting intensities by the function $q(x) = x^p$. In our experiments, we test quadratic and cubic projections (i.e. $p = 2$ and $p = 3$, respectively).

The visual effect of the strategies discussed above can be found in Figure A.1 in the Appendix. We also consider the change in the distribution of the foreground (vessel) voxels and background voxels in the different preprocessing schemes. From Figure A.2 in the Appendix, it is evident that quadratic and cubic intensity projections have the effect of stretching the range of the foreground intensities and compressing the background intensities. This can be seen from the histogram of original and clipped intensities (**a**, and **b**) compared to the polynomial intensity projections (**c** and **d**). This change in intensity (in **c** and **d**) causes the distribution to look similar to that of the synthetic data (in **e**). Results from experiments with these preprocessing strategies are given in Section 3.3.

3.3. Evaluating the DeepVesselNet Components

Table 1: Result from Experiment 1: comparing the effect of preprocessing strategies on the vessel segmentation task of TOF MRA dataset after pre-training on synthetic data without fine-tuning. It can be seen that quadratic projection gives a more balanced precision and recall.

Preprocessing method	Precision	Recall	Dice
Original Intensities	0.0112	1.0	0.0221
Clipped Intensities	0.3117	0.9379	0.4679
Quadratic Projection	0.8276	0.8025	0.8149
Cubic Projection	0.8995	0.7033	0.7894

Experiment 1: Testing Preprocessing Strategies. In a first experiment we test the different preprocessing strategies proposed in Section 3.2. We carry out these experiments considering the synthetic dataset as ideal. Therefore, we train DeepVesselNet-FCN solely on our synthetic data for vessel segmentation and then use this trained network to segment vessels in the clinical MRA dataset after applying each of the four preprocessing strategies discussed in Section 3.2. This enables us to measure how similar our clinical MRA dataset is to the synthetic dataset after the preprocessing. Table 1 shows the outcome of these experiments. From the results, we observe that the intensity projections (quadratic and cubic) improve the results (a lot in terms of Dice score). Quadratic intensity projection performs slightly better than the cubic intensity projection. Therefore, we use the quadratic intensity projection in all following experiments.

Experiment 2: Fast Cross-hair Filters. To investigate the usefulness of the 3-D context information extracted by the cross-hair filters of DeepVesselNet, we construct a similar network as in Figure 6 (DeepVesselNet-FCN) but replace the cross-hair filters by 2-D filters. We train this 2.5-D network on 2-D slices (as described in Section 2.1) and then fuse the final slice-wise results to obtain the final volumetric predictions. We perform this experiment using the synthetic dataset for centerline prediction with the training and test splits discussed in

Section 3.1 and measure the performance of the two networks in terms of execution time per volume (Ex. time) and in terms of accuracy by the Dice score. Results are reported in Table 2. We observe a benefit of 2.5-D network in terms of run time (6s vs. 13s) when compared to the 3-D network but at a cost of accuracy in terms of Dice Score (0.70 vs. 0.80), which can be explained by lack of 3-D context information.

Table 2: Results from Experiment 2: comparison of 2-D and cross-hair filter versions of DeepVesselNet-FCN on centerline prediction using synthetically generated data.

Method	Precision	Recall	Dice	Ex. time
3-D (Cross-hair filters)	0.7763	0.8235	0.7992	13 s
2.5-D (2-D on slices)	0.6580	0.7472	0.7017	6 s

Experiment 3: The FP Rate Correction Loss Function ($\mathcal{L}_1 + \mathcal{L}_2$). To test the effect of FP rate correction loss function discussed in section 2.2, we train DeepVesselNet-FCN architecture on a sub-sample of four clinical MRA volumes from scratch, with and without FP rate correction described in Equation (7). We train for 5000 iterations and record the ratio of precision to recall every 5 iterations using a threshold of 0.5 on the probability maps. A plot of the precision - recall ratio during training without FP rate correction (\mathcal{L}_1 Only) and with FP rate correction ($\mathcal{L}_1 + \mathcal{L}_2$) is presented in Figure 9. The results of this experiments suggest that training with both factors in the loss function, as proposed in Section 2.2, keeps a better balance between precision and recall (i.e. a ratio closer to 1.0) than without the second factor. A balanced precision-recall ratio implies that the training process is not bias towards the background or the foreground. This helps prevent over-segmentation, which is caused by the introduction of the class balancing.

Experiment 4: Pre-training on Synthetic Data. We assess the usefulness of transfer learning with synthetic data by comparing the training convergence

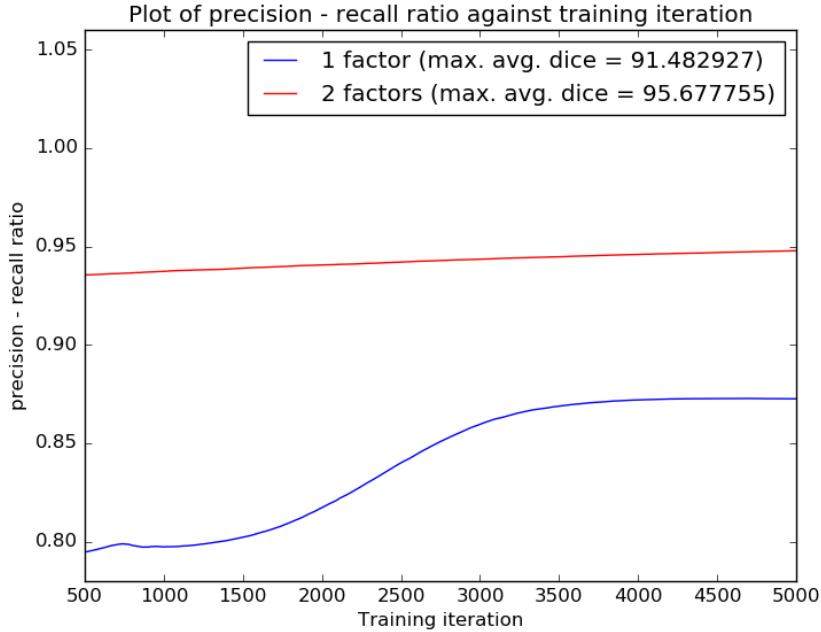


Figure 9: Precision - recall ratio during training, with FP rate correction and without FP rate correction in the loss function, on four selected clinical MRA volumes. **max. avg Dice** refers to the final training Dice value on these four volumes. A balanced precision-recall ratio (i.e. close to 1) implies that the training process is not bias towards the background or the foreground.

speed, and various other scores that we obtain when we pre-train DeepVesselNet-FCN on synthetic data and fine-tune on the clinical MRA dataset, compared to training DeepVesselNet-FCN from scratch on the clinical MRA. For this experiment, we only consider the vessel segmentation task, as no annotated clinical data is available for centerline and bifurcation tasks. Results of this experiment are reported in Table 3. We achieve a Dice score of 0.8639 for training from scratch without pre-training on synthetic data and 0.8668 when pre-training on synthetic data. This shows that training from scratch or pre-training on synthetic data does not make a big difference regarding the accuracy

of the results. However, training from scratch requires about 600 iterations more than pre-training on synthetic data for the network to converge (i.e. 50% more longer).

Table 3: Results from Experiment 4: pre-training DeepVesselNet-FCN on synthetic data and fine-tuning with the training set from the clinical MRA and training DeepVesselNet-FCN from scratch on clinical MRA. Iterations refers to training iterations required for the network to converge. Although the result in Dice score are not very different, it is clear that the pre-training on synthetic data leads to an earlier convergence of the network.

Method	Precision	Recall	Dice	Iterations
With pre-training	0.8644	0.8693	0.8668	1200
Without pre-training	0.8587	0.8692	0.8639	1800

3.4. Evaluating DeepVesselNet Performance

In this subsection, we retain the best training strategy from above and assess the performance of our proposed network architecture with other available methods on the three main tasks of vessel segmentation, centerline prediction, and bifurcation detection.

Experiment 1: Vessel Segmentation. We pre-train DeepVesselNet-(FCN and VNet) architectures (from Figures 6 and 7) on synthetic volumes for vessel segmentation and evaluate its performance on TOF MRA volumes. We then fine-tune the networks with additional clinical TOF MRA data, repeating the evaluation. Table 4 reports results of these tests, together with performances of competing methods. We obtain a Dice score of 0.81 for DeepVesselNet-FCN and 0.80 for DeepVesselNet-VNet before, and 0.86 (DeepVesselNet-FCN) as well as 0.84 (DeepVesselNet-VNet) after fine tuning. Generating vessel segmentations takes less than 13s per volume using DeepVesselNet-FCN and 20s for DeepVesselNet-VNet. Table 4 also reports results from the methods of Schneider et al. (2015) (V-Net) and Forkert et al. (2013) both of which are outperformed by DeepVesselNet-FCN both in terms of speed (execution time) and Dice score.

Comparing DeepVesselNet-VNet and original V-Net on the MRA data, we find a small advantage for the latter in terms of Dice score which cannot be considered as significant (0.8425 and 0.8497 respectively with sample standard error of 0.0066 and T-test significance probability of 0.1950). However, DeepVesselNet-VNet has the advantage of being six seconds faster (about 23% improvement) during prediction, a time difference that will scale up with volume size and filter sizes, and our results show that cross-hair filters can be used in DeepVesselNet at a little to no cost in terms of vessel segmentation accuracy.

Experiment 2: Centerline Prediction. For centerline prediction, we train DeepVesselNet on the synthetic dataset and test it on synthetic as well as clinical datasets. The network uses the probabilistic segmentation masks from experiment 1 as an input (together with the same training and test splits described in experiment 1) and uses the vessel predictions to restrict predicted centerlines to be within vessel regions. Qualitative results are presented in Figures 10 and 11 together with quantitative scores in Table 5. We obtain a Dice score of 0.79 for DeepVesselNet-FCN, outperforming all other methods by a margin of more than 5%. DeepVesselNet-FCN is also better than the competing methods in terms of execution time. Again, DeepVesselNet-VNet performs slightly worse than V-Net in terms of the Dice score (0.67 vs. 0.75). However, DeepVesselNet-VNet has an advantage in speed of execution (17s vs. 23s).

Experiment 3: Bifurcation Detection. For a quantitative evaluation of DeepVesselNet in bifurcation detection, we use synthetically generated data, and adopt a two-input-channels strategy. We use the vessel segmentations from experiments 1 as one input channel and the centerline predictions from Experiment 2 as a second input channel relying on the same training and test splits as in the previous experiments. In our predictions we aim at localizing a cubic region of size $5 \times 5 \times 5$ around the bifurcation points, which are contained within the vessel segmentation. We evaluate the results based on a hit-or-miss criterion: a bifurcation point in the ground truth is counted as hit if a region of a cube of size $5 \times 5 \times 5$ centered on this point overlaps with the prediction, and

Table 4: Results for Experiment 1: vessel segmentation results on test datasets. Results on TOF MRA for all methods are evaluated within the brain region using brain masks. Average execution time (Ex. time) is the average time for segmentation of one full volume in each dataset during test and DVN refers to DeepVesselNet, (Pre) refers to the result we obtained on the test set after pre-training, and (Fine) is the result after fine-tuning.

Dataset	Method	Precision	Recall	Dice	Ex. time
Synthetic	DVN-FCN	0.9984	0.9987	0.9986	13 s
	DVN-VNet	0.9954	0.9959	0.9956	17 s
	V-Net	0.9948	0.9950	0.9949	23 s
	Schneider et al.	0.9947	0.9956	0.9952	n/a
TOF MRA	DVN-FCN (Fine)	0.8644	0.8693	0.8668	13 s
	DVN-FCN (Pre)	0.8276	0.8025	0.8148	13 s
	DVN-VNet (Fine)	0.8500	0.8351	0.8425	20 s
	DVN-VNet (Pre)	0.8332	0.7712	0.8010	20 s
	V-Net (Fine)	0.8434	0.8562	0.8497	26 s
	V-Net (Pre)	0.8241	0.7582	0.7898	26 s
	Schneider et al.	0.8481	0.8215	0.8346	100 mins
	Forkert et al.	0.8499	0.7300	0.7857	n/a
SRXTM	DVN-FCN	0.9672	0.9582	0.9627	4 s
	DVN-VNet	0.9583	0.9618	0.9601	7 s
	V-Net	0.9525	0.9584	0.9555	11 s
	Schneider et al.	0.9515	0.9151	0.9330	23 mins

counted as a miss otherwise; a hit is considered as true positive (TP) and a miss is considered as false negative (FN); a positive label in the prediction is counted as false positive (FP) if a cube of size $5 \times 5 \times 5$ centered on this point contains no bifurcation point in the ground truth. Qualitative results on synthetic and clinical MRA TOF are shown in Figures 12 and 13, respectively. Results for Schneider et al. (2015) are obtained by first extracting the vessel network and then all nodes with two or more splits are treated as bifurcations. In Figure

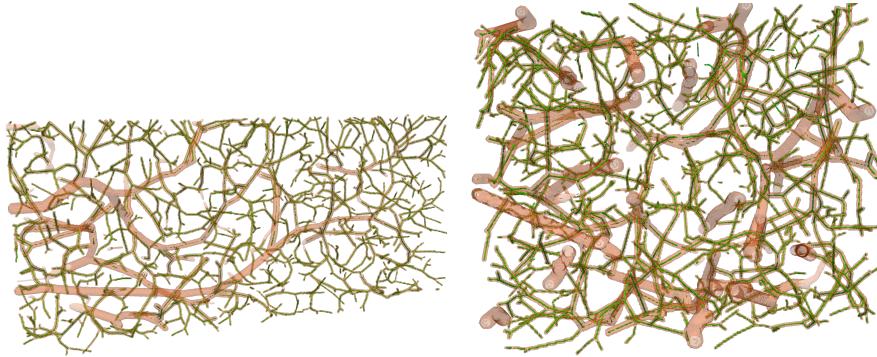


Figure 10: Centerline prediction on synthetic test data using DeepVesselNet (centerline in green)

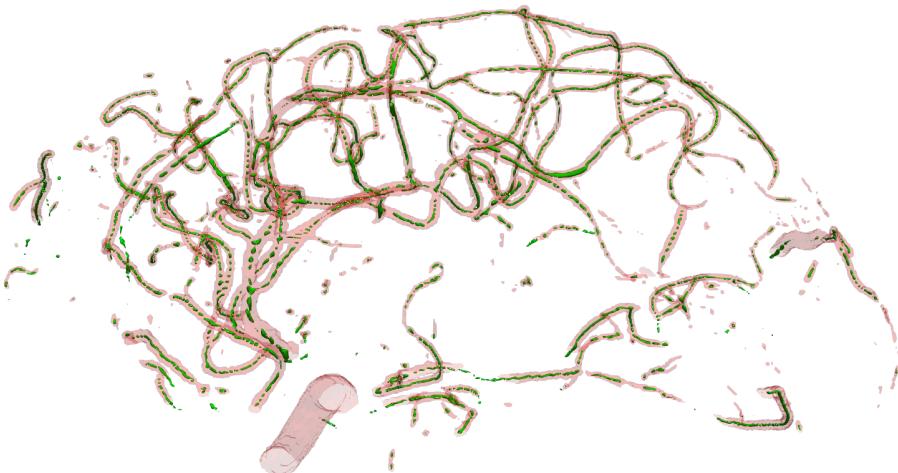


Figure 11: Centerline prediction on clinical MRA test data using DeepVesselNet (centerline in green)

14, we present the precision-recall curve obtained by varying the threshold for converting probability maps, from the networks, into binary predictions. Results from Table 6 and Figure 14 show that DeepVesselNet-FCN performs better than the other architectures in 5 out of 6 metrics. In our experiments, it became

Table 5: Results for Experiment 2: centerline prediction tasks. Average execution time (Ex. Time) is the average time for predicting one full volume during test and DVN refers to DeepVesselNet.

Method	Precision	Recall	Dice	Ex. Time
DVN-FCN	0.7763	0.8235	0.7992	13 s
DVN-VNet	0.6515	0.6887	0.6696	17 s
V-Net	0.7641	0.7330	0.7482	23 s
Schneider et al.	0.4807	0.8603	0.6168	n/a

evident that V-Net tends to over-fit, possibly suffering from its high number of parameters to be determined from the rather few bifurcations present in our training data. This may explain why results for V-Net are worse than all other methods, also suggesting that in cases where little training data is available, the DeepVesselNet-FCN architecture may be the preferable.

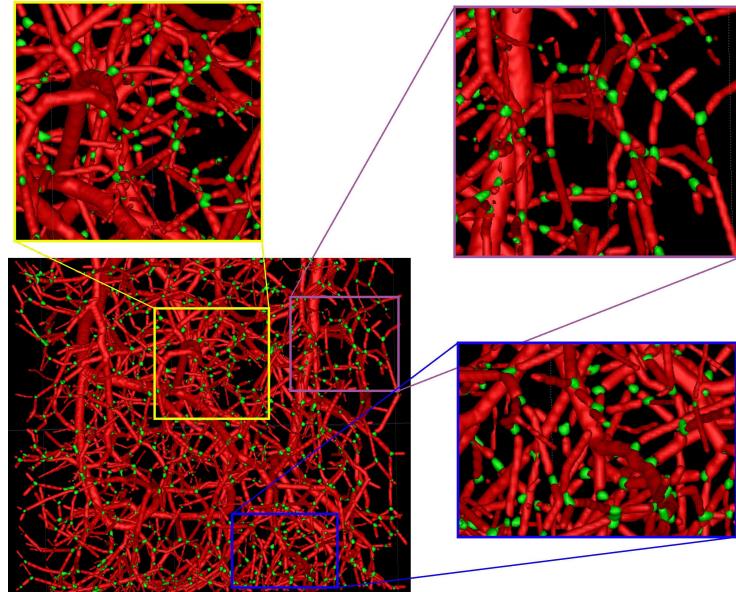


Figure 12: Bifurcation detection on synthetic test data using DeepVesselNet-FCN (bifurcations in green)

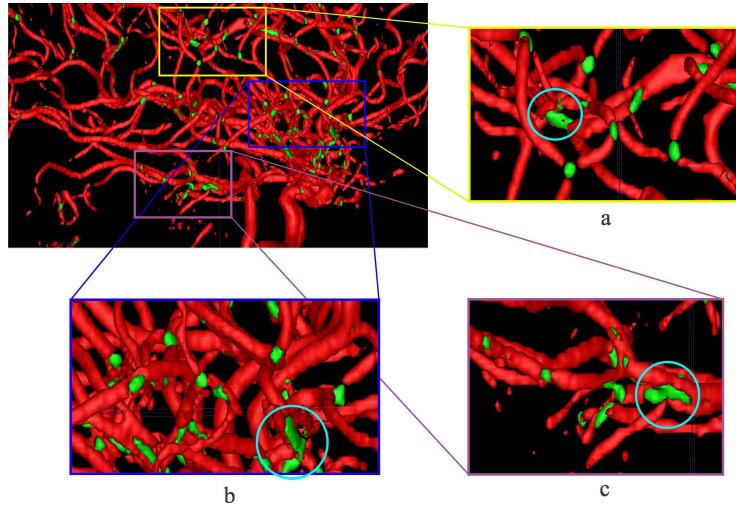


Figure 13: Bifurcation detection on clinical MRA test data using DeepVesselNet-FCN (bifurcations in green). At regions where a lot of vessels intersect, the network predicts it as a big bifurcation, this can be seen in the circled regions in zoomed images (a, b, and c).

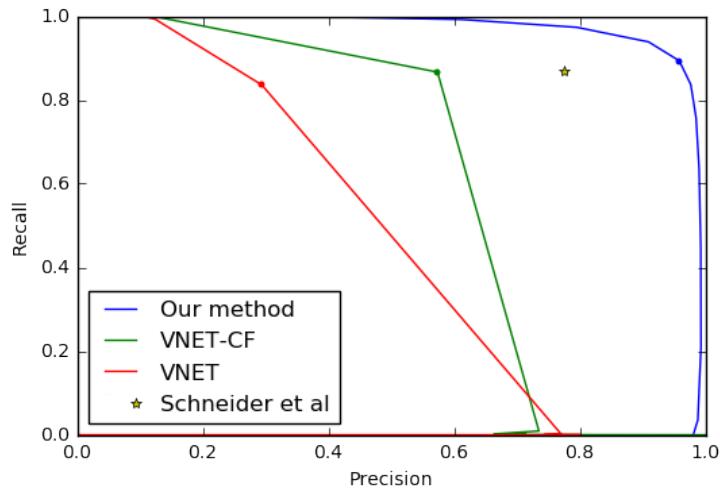


Figure 14: Precision-recall curve of the results from the bifurcation detection task obtained by varying the threshold for converting probabilities maps into binary predictions.

Table 6: Result from Experiment 3: bifurcation detection experiments. Precision and recall are measured on the basis of the $5 \times 5 \times 5$ blocks around the bifurcation points. Mean error and its corresponding standard deviation are measured in voxels away from the bifurcation points (not blocks) and DVN refers to DeepVesselNet.

Method	Precision	Recall	Det.	Mean	Err	Ex. Time
	%			Err	Std	
DVN-FCN	0.7880	0.9297	86.87	0.2090	0.6671	13 s
DVN-VNet	0.4680	0.5670	84.21	1.6533	0.9645	17 s
V-Net	0.2550	0.6871	70.29	1.2434	1.3857	23 s
Schneider et al.	0.7718	0.8508	84.30	0.1529	0.7074	n/a

4. Summary and Conclusions

We present DeepVesselNet, an architecture tailored to the challenges of extracting vessel networks and network features using deep learning. Our experiments have shown that the cross-hair filters, which is one of the components of DeepVesselNet, performs comparably well as 3-D filters and, at the same time, improves significantly both speed and memory usage, easing an upscaling to larger data sets. Another component of DeepVesselNet, the introduction of new weights and the FP rate correction in the class balancing loss function helps in maintaining a good balance between precision and recall during training. This turns out to be crucial for preventing over and under-segmentation problems, which are common problems in vessel segmentation. Finally, we successfully demonstrated that transfer learning of DeepVesselNet through pre-training on synthetically generated data improves segmentation and detection results, especially in situations where obtaining manually annotated data is a challenge.

As future work, we will generalize DeepVesselNet to multiclass segmentation, handling vessel segmentation, centerline prediction, and bifurcation detection simultaneously, rather than in three subsequent binary tasks. We also expect that network architectures tailored to our three hierarchically nested classes will improve the performance of the DeepVesselNet, for example, in a single,

but hierarchical approach starting from a base network for vessel segmentation, additional layers for centerline prediction, and a final set of layers for bifurcation detection. The current implementation (cross-hair layers, the networks, cost function), and future extensions of DeepVesselNet, will be made publicly available on Github.

References

- Y. Bengio and X. Glorot. Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the 13th International Conference on Artificial Intelligence and Statistics (AISTATS)*, volume 9, 2010.
- Thanapong Chaichana, Zhonghua Sun, Mark Barrett-Baxendale, and Atulya Nagar. Automatic location of blood vessel bifurcations in digital eye fundus images. In *Proceedings of Sixth International Conference on Soft Computing for Problem Solving: SocProS 2016, Volume 2*, pages 332–342, Singapore, 2017. Springer Singapore.
- Da Chen and Laurent D. Cohen. Piecewise geodesics for vessel centerline extraction and boundary delineation with application to retina segmentation. In *Scale Space and Variational Methods in Computer Vision: 5th International Conference, SSVM 2015, Lège-Cap Ferret, France, May 31 - June 4, 2015, Proceedings*, pages 270–281. Springer International Publishing, 2015.
- P. F. Christ, F. Ettlinger, F. Grün, M. E. A. Elshaera, J. Lipkova, S. Schlecht, F. Ahmaddy, S. Tatavarty, M. Bickel, P. Bilic, M. Rempfler, F. Hofmann, M. D Anastasi, S.-A. Ahmadi, G. Kaassis, J. Holch, W. Sommer, R. Braren, V. Heinemann, and B. Menze. Automatic liver and tumor segmentation of ct and mri volumes using cascaded fully convolutional neural networks. *ArXiv e-prints*, February 2017.
- Albert C. S. Chung and J. Alison Noble. Statistical 3d vessel segmentation using a rician distribution. In Chris Taylor and Alain Colchester, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI: Second International Conference, Cambridge, UK, Proceedings*, pages 82–89, Berlin, Heidelberg, 1999. Springer.
- D. C. Ciresan, A. Giusti, L. M. Gambardella, and J. Schmidhuber. Deep neural networks segment neuronal membranes. *electron microscopy images. In NIPS*, page 28522860, 2012.

- Adrian Dalca, Giovanna Danagoulian, Ron Kikinis, Ehud Schmidt, and Polina Golland. Segmentation of nerve bundles and ganglia in spine mri using particle filters. In *Medical Image Computing and Computer-Assisted Intervention — MICCAI: 14th International Conference, Toronto, Canada, Proceedings*, pages 537–545, Berlin, Heidelberg, 2011. Springer.
- Charles Florin, Nikos Paragios, and Jim Williams. Globally optimal active contours, sequential monte carlo and on-line learning for vessel segmentation. In *Computer Vision – ECCV 2006: 9th European Conference on Computer Vision, Graz, Austria, Proceedings*, pages 476–489, Berlin, Heidelberg, 2006. Springer.
- N. D. Forkert, A. Schmidt-Richberg, J. Fiehler, T. Illies, D. Möller, H. Handels, and D. Säring. Fuzzy-based vascular structure enhancement in time-of-flight mra images for improved segmentation. *Methods of Information Medicine*, 50:74–83, 2011.
- N. D. Forkert, A. Schmidt-Richberg, J. Fiehler, T. Illies, D. Möller, D. Säring, H. Handels, and J. Ehrhardt. 3d cerebrovascular segmentation combining fuzzy vessel enhancement and level-sets with anisotropic energy weights. *Magnetic Resonance Imaging*, 31:262–271, 2013.
- Alejandro F. Frangi, Wiro J. Niessen, Koen L. Vincken, and Max A. Viergever. Multiscale vessel enhancement filtering. In William M. Wells, Alan Colchester, and Scott Delp, editors, *Medical Image Computing and Computer-Assisted Intervention — MICCAI*, pages 130–137, Berlin, Heidelberg, 1998. Springer.
- Jerzy W. Grzymala-Busse, Linda K. Goodwin, Witold J. Grzymala-Busse, and Xinqun Zheng. An approach to imbalanced data sets based on changing rule strength. In Sankar K. Pal, Lech Polkowski, and Andrzej Skowron, editors, *Rough-Neural Computing: Techniques for Computing with Words*, pages 543–553, Berlin, Heidelberg, 2004. Springer.
- Guo Haixiang, Li Yijing, Jennifer Shang, Gu Mingyun, Huang Yuanyue, and

- Gong Bing. Learning from class-imbalanced data: Review of methods and applications. *Expert Systems with Applications*, 73:220 – 239, 2017.
- Jyh-Jing Hwang and Tyng-Luh Liu. Pixel-wise deep learning for contour detection. In *ICLR*, 2015.
- C. Kirbas and F. Quek. A review of vessel extraction techniques and algorithms. *ACM Comput. Surv.*, 36:81–121, 2004.
- Max W. K. Law and Albert C. S. Chung. Three dimensional curvilinear structure detection using optimally oriented flux. In David Forsyth, Philip Torr, and Andrew Zisserman, editors, *Computer Vision – ECCV 2008*, pages 368–382, Berlin, Heidelberg, 2008. Springer.
- D. Lesage, E.D. Angelini, I. Bloch, and G. Funka-Lea. A review of 3d vessel lumen segmentation techniques: models, features and extraction schemes. *Med. Image Anal.*, 13(6):819845, 2009.
- Wei Liao, Karl Rohr, and Stefan Wörz. Globally optimal curvature-regularized fast marching for vessel segmentation. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI*, pages 550–557, Berlin, Heidelberg, 2013. Springer.
- Siqi Liu, Donghao Zhang, Yang Song, Hanchuan Peng, and Weidong Cai. Triple-crossing 2.5d convolutional neural network for detecting neuronal arbours in 3d microscopic images. In *Machine Learning in Medical Imaging*, pages 185–193. Springer International Publishing, 2017.
- J. Long, E. Shelhamer, and T. Darrell. Fully convolutional networks for semantic segmentation. *CoRR*, *abs/1411.4038*, 2014.
- Maysa M. G. Macedo, Choukri Mekkaoui, and Marcel P. Jackowski. Vessel centerline tracking in cta and mra images using hough transform. In *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications: 15th Iberoamerican Congress on Pattern Recognition, CIARP 2010*,

Sao Paulo, Brazil, November 8-11, 2010. Proceedings, pages 295–302, Berlin, Heidelberg, 2010. Springer.

M. Maddah, A. Afzali-khusha, and H. Soltanian. Snake modeling and distance transform approach to vascular center line extraction and quantification. *Computerized Med. Imag. and Graphics*, 27 (6):503–512, 2003.

Kevis-Kokitsi Maninis, Jordi Pont-Tuset, Pablo Arbeláez, and Luc Van Gool. Deep retinal image understanding. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016: 19th International Conference, Athens, Greece, October 17-21, 2016, Proceedings, Part II*, pages 140–148. Springer International Publishing, 2016.

M. Elena Martínez-Pérez, Alun D. Hughes, Alice V. Stanton, Simon A. Thom, Anil A. Bharath, and Kim H. Parker. Retinal blood vessel segmentation by means of scale-space analysis and region growing. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI*, pages 90–97, Berlin, Heidelberg, 1999. Springer.

F. Milletari, N. Navab, and S. Ahmadi. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In *Fourth International Conference on 3D Vision (3DV)*, pages 565–571, 2016.

Rodrigo Moreno, Chunliang Wang, and Örjan Smedby. Vessel wall segmentation using implicit models and total curvature penalizers. In *Image Analysis: 18th Scandinavian Conference, Proceedings*, pages 299–308, Berlin, Heidelberg, 2013. Springer.

Delphine Nain, Anthony Yezzi, and Greg Turk. Vessel segmentation using a shape driven flow. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI*, pages 51–59, Berlin, Heidelberg, 2004. Springer.

Isabella Nogues, Le Lu, Xiaosong Wang, Holger Roth, Gedas Bertasius, Nathan Lay, Jianbo Shi, Yohannes Tsehay, and Ronald M. Summers. Automatic lymph node cluster segmentation using holistically-nested neural networks

and structured optimization in ct images. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016: 19th International Conference, Athens, Greece, October 17-21, 2016, Proceedings, Part II*, pages 388–397. Springer International Publishing, 2016.

Renzo Phellan and Nils D. Forkert. Comparison of vessel enhancement algorithms applied to time-of-flight mra images for cerebrovascular segmentation. *Medical Physics*, 44(11):5901–5915, 2017.

Renzo Phellan, Alan Peixinho, Alexandre Falcão, and Nils D. Forkert. Vascular segmentation in tof mra images of the brain using a deep convolutional neural network. In *Intravascular Imaging and Computer Assisted Stenting, and Large-Scale Annotation of Biomedical Data and Expert Label Synthesis*, pages 39–46. Springer International Publishing, 2017.

Johannes Reichold, Marco Stampanoni, Anna Lena Keller, Alfred Buck, Patrick Jenny, and Bruno Weber. Vascular graph model to simulate the cerebral blood flow in realistic vascular networks. *Journal of Cerebral Blood Flow & Metabolism*, 29(8):1429–1443, 2009.

M. Rempfler, M. Schneider, G. D. Ielacqua, X. Xiao, S. R. Stock, J. Klohs, G. Székely, B. Andres, and B. H. Menze. Reconstructing cerebrovascular networks under local physiological constraints by integer programming. *Medical Image Analysis*, pages 86–94, 2015.

Roberto Rigamonti, Amos Sironi, Vincent Lepetit, and Pascal Fua. Learning separable filters. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2013.

O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 9351:234–241, 2015.

Holger R. Roth, Le Lu, Ari Seff, Kevin M. Cherry, Joanne Hoffman, Shijun Wang, Jiamin Liu, Evrim Turkbey, and Ronald M. Summers. A new 2.5d rep-

resentation for lymph node detection using random sets of deep convolutional neural network observations. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2014*, pages 520–527. Springer International Publishing, 2014.

Holger R. Roth, Le Lu, Amal Farag, Andrew Sohn, and Ronald M. Summers. Spatial aggregation of holistically-nested networks for automated pancreas segmentation. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016: 19th International Conference, Athens, Greece, October 17-21, 2016, Proceedings, Part II*, pages 451–459. Springer International Publishing, 2016.

A. Santamaría-Pang, C. M. Colbert, P. Saggau, and I. A. Kakadiaris. Automatic centerline extraction of irregular tubular structures using probability volumes from multiphoton imaging. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2007: 10th International Conference, Brisbane, Australia, October 29 - November 2, 2007, Proceedings, Part II*, pages 486–494, Berlin, Heidelberg, 2007. Springer.

M. Schneider, J. Reichold, B. Weber, G. Székely, and S. Hirsch. Tissue metabolism driven arterial tree generation. *Medical Image Analysis*, pages 1397–1414, 2012.

Matthias Schneider, Sven Hirsch, Bruno Weber, Gábor Székely, and Bjoern H. Menze. Joint 3-d vessel segmentation and centerline extraction using oblique hough forests with steerable filters. *Medical Image Analysis*, 19(1):220–249, 2015.

A. Sekuboyina, A. Valentinitsch, J.S. Kirschke, and B. H. Menze. A localisation-segmentation approach for multi-label annotation of lumbar vertebrae using deep nets. *arXiv*, 1703.04347, 2017.

B. Shagufta, S. A. Khan, A. Hassan, and A. Rashid. Blood vessel segmentation and centerline extraction based on multilayered thresholding in ct images.

Proceedings of the 2nd International Conference on Intelligent Systems and Image Processing, pages 428–432, 2014.

Dominik Szczerba and Gábor Székely. Simulating vascular systems in arbitrary anatomies. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI: 8th International Conference, Palm Springs, CA, USA, October 26-29, 2005, Proceedings*, pages 641–648, Berlin, Heidelberg, 2005. Springer.

Giles Tetteh, Markus Rempfler, Claus Zimmer, and Bjoern H. Menze. Deep-fext: Deep feature extraction for vessel segmentation and centerline prediction. In *Machine Learning in Medical Imaging*, pages 344–352. Springer International Publishing, 2017.

Theano Development Team. Theano: A Python framework for fast computation of mathematical expressions. *arXiv e-prints*, abs/1605.02688, May 2016.

Shijun Wang, Brandon Peplinski, Le Lu, Weidong Zhang, Jianfei Liu, Zhuoshi Wei, and Ronald M. Summers. Sequential monte carlo tracking for marginal artery segmentation on ct angiography by multiple cue fusion. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI: 16th International Conference, Nagoya, Japan, Proceedings*, pages 518–525, Berlin, Heidelberg, 2013. Springer.

Stefan Wörz, William J. Godinez, and Karl Rohr. Probabilistic tracking and model-based segmentation of 3d tubular structures. In *Bildverarbeitung für die Medizin 2009: Algorithmen — Systeme — Anwendungen Proceedings des Workshops*, pages 41–45, Berlin, Heidelberg, 2009. Springer.

Saining Xie and Zhuowen Tu. Holistically-nested edge detection. In *Proceedings of the IEEE international conference on computer vision*, pages 1395–1403, 2015.

S. Young, V. Pekar, and J. Weese. Vessel segmentation for visualization of mra with blood pool contrast agent. In *Medical Image Computing and Computer-*

Assisted Intervention – MICCAI, pages 491–498, Berlin, Heidelberg, 2001. Springer.

Yefeng Zheng, Jianhua Shen, Huseyin Tek, and Gareth Funka-Lea. Model-driven centerline extraction for severely occluded major coronary arteries. In *Machine Learning in Medical Imaging: Third International Workshop, MLMI 2012, Held in Conjunction with MICCAI: Nice, France, Revised Selected Papers*, pages 10–18, Berlin, Heidelberg, 2012. Springer.

Yefeng Zheng, David Liu, Bogdan Georgescu, Hien Nguyen, and Dorin Comaniciu. 3d deep learning for efficient and robust landmark detection in volumetric data. In Nassir Navab, Joachim Hornegger, William M. Wells, and Alejandro Frangi, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI*, pages 565–572. Springer International Publishing, 2015.

Appendix A. Effect of Preprocessing Strategies

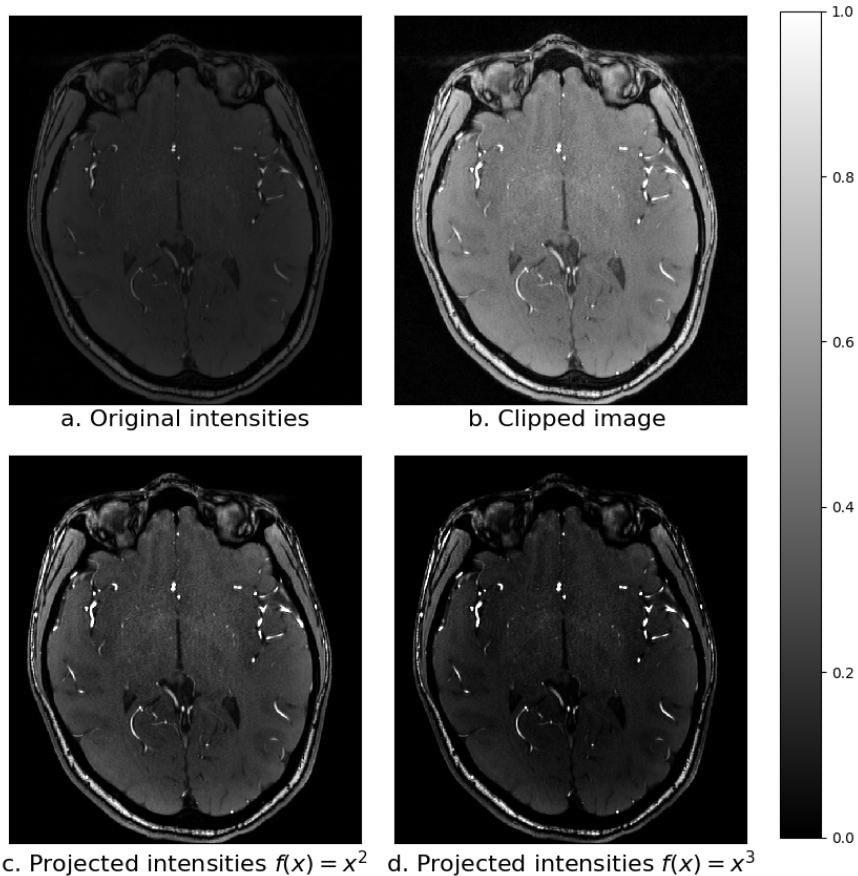


Figure A.1: Sample of slide from TOF MRA dataset after application of the discussed preprocessing schemes. Intensities in c. and d. have been further normalized between [0,1].

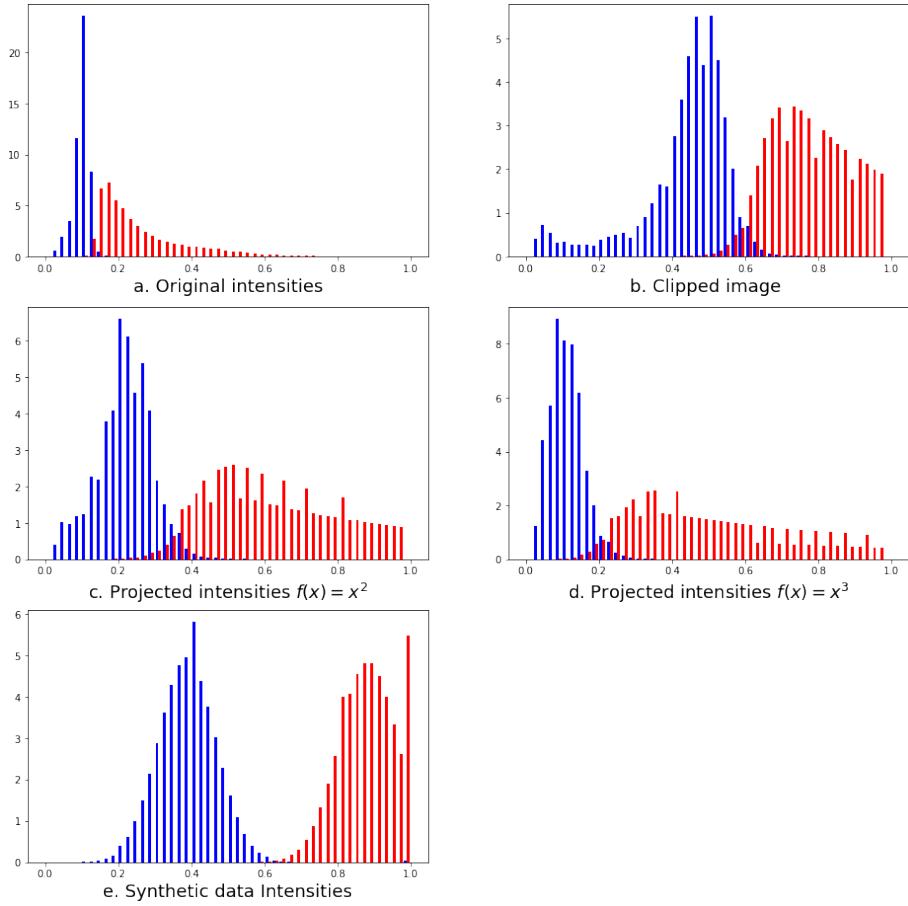


Figure A.2: Distribution of foreground and background labels before and after the pre-processing schemes. Subplots **a**, **b**, **c**, and **d**. refer to original intensities, clipped intensities, quadratic projection, and cubic projection, respectively. Subplot **e** shows the distribution of the labels in the synthetic dataset. Blue colored bars represent background intensities and red colored bars represent foreground intensities. All intensities have been normalized to the range [0,1]. From the histograms, we can see that intensity projection with quadratic and cubic polynomial stretches the intensity range for the foreground and compresses that of the background.