A Benchmark Framework for Multiregion Analysis of Vesselness Filters

Jonas Lamy, Odyssee Merveille, Bertrand Kerautret, Nicolas Passat

A Benchmark Framework for Multi-Region Analysis of Vesselness Filters

Jonas Lamy, Odyssée Merveille, Bertrand Kerautret, Nicolas Passat

***Abstract*— This paper is an updated version of [1], following the correction of numerical errors. Vessel enhancement (aka vesselness) filters, are part of angiographic image processing for more than twenty years. Their popularity comes from their ability to enhance tubular structures while filtering out other structures, especially as a preliminary step of vessel segmentation. Choosing the right vesselness filter among the many available can be difficult, and their parametrization requires an accurate understanding of their underlying concepts and a genuine expertise. In particular, using default parameters is often not enough to reach satisfactory results on specific data. Currently, only few benchmarks are available to help the users choosing the best filter and its parameters for a given application. In this article, we present a generic framework to compare vesselness filters. We use this framework to compare seven gold standard filters. Our experiments are performed on three public datasets: the hepatic Ircad dataset (CT images), the Bullit dataset (brain MRA images) and the synthetic VascuSynth dataset. We analyse the results of these seven filters both quantitatively and qualitatively. In particular, we assess their performances in key areas: the organ of interest, the whole vascular network neighbourhood and the vessel neighbourhood split into several classes, based on their diameters. We also focus on the vessels bifurcations, which are often missed by vesselness filters. We provide the code of the benchmark, which includes up-to-date C++ implementations of the seven filters, as well as the experimental setup (parameter optimization, result analysis, etc.). An online demonstrator is also provided to help the community apply and visually compare these vesselness filters.**

***Index Terms*— Angiographic imaging, vessels, image enhancement/restoration, computed tomography, magnetic resonance imaging (MRI).**

I. Introduction

ANGIOGRAPHIC imaging provides information (position, size and shape of vessels) that can help clinicians to diagnose and perform follow-up tasks related to cardiovascular diseases. The ability to efficiently analyse the vascular structures in these data is then of high interest. In this context, the segmentation of vascular networks from angiographic images is the key step for further visualization and analysis. Yet crucial, this step remains an open problem, specifically in 3D images. Indeed, vessels are thin, elongated structures sparsely distributed in the image. They are also poorly contrasted, especially when reaching the image resolution, or due to physiological phenomena (e.g. flow turbulence). The problem is even more difficult in pathological cases, where the usual anatomical hypotheses are sometimes no longer satisfied.

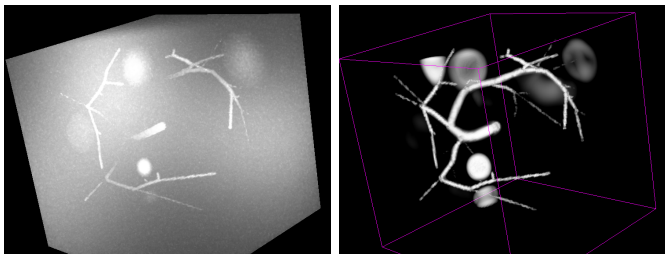


Fig. 1. Vessel enhancement (right) applied on a VascuSynth synthetic volume with MRI artifacts (left).

Popular techniques used to produce angiographic images, such as X-ray Computed Tomography Angiography (CTA) [2] and Magnetic Resonance Angiography (MRA) [3] are often undermined with poor vessels contrast and low signal-to-noise ratio (SNR), especially when they are used without contrast agent injection. MRA may be preferred over CTA as it is non-ionising; however the images present additional artifacts due to the magnetic field effects, such as signal inhomogeneity in tissues of the same nature. Finally, the acquisition procedure can also induce other artifacts, e.g. ghosting in MRA that appears during acquisition when the patient cannot hold his/her breath for a long time. The use of a contrast agent yields better images. However multiple physical and physiological parameters have to be considered for determining the adequate timing between the injection and the acquisition in order to get optimal contrast between the vessels of interest and the organ.

There is a vast literature dealing with angiographic image analysis. A complete review is beyond the scope of this article; see [4]–[6] for surveys on vessel segmentation and tracking. Many of the methods presented in these surveys—both "classical" and deep learning ones—rely on a preliminary step of vessel filtering / enhancement before carrying out their own processing. It is important to note that these vesselness

J. Lamy and B. Kerautret are with Univ Lyon, Univ Lyon 2, CNRS, INSA Lyon, UCBL, LIRIS, UMR5205, F-69676 Bron, France (e-mail:{jonas.lamy,bertrand.kerautret}@univ-lyon2.fr).

O. Merveille is with Univ Lyon, INSA-Lyon, Université Claude Bernard Lyon 1, UJM-Saint Etienne, CNRS, Inserm, CREATIS UMR 5220, U1294, F-69XXX, LYON, France (e-mail: odyssee.merveille@creatis.insa-lyon.fr).

N. Passat is with Université de Reims Champagne-Ardenne, CReSTIC, Reims, France (e-mail: nicolas.passat@univ-reims.fr).

This work was supported by the French *Agence Nationale de la Recherche*: grants ANR-18-CE45-0018, ANR-18-CE45-0014, ANR-20-CE45-0011.

filters do not require any annotation to be computed, contrary to supervised machine learning approaches. They are thus even more interesting when no annotated dataset is available, which is common in 3D vascular-related applications.

A. Vessel enhancement

Vesselness filters refer to operators designed to increase the vessel contrast by enhancing the blood signal and/or removing the signal of non-vessel structures. An example of vesselness filter is depicted in Fig. 1. The normalized filtering result can usually be interpreted as the probability of voxels to belong to a blood vessel, hence the terminology "vesselness".

One of the first vessel enhancement filters proposed in the literature was the Sato filter [7]. It uses the eigenvalues of the Hessian matrix of the angiographic image to analyse the local curvature of the voxel intensities and enhance tubular shapes. Frangi et al. [8] proposed a similar but improved formulation, which is often considered as the current gold standard. Other Hessian-based formulations were further developed by Meijering et al. [9], Jerman et al. [10] and Zhang et al. [11].

Alternatively, some methods compute the eigenvalues on different formulations of structure tensors, such as OOF (Optimally Oriented Flux) [12] and phase tensors [13].

Other methods use mathematical morphology paradigms to enhance vessels, based on geometrical properties. This is the case of RORPO (Ranking the Orientations of Robust Path Openings) [14], that relies on rank filtering of path-opening responses in the 3D principal directions.

Finally, some vesselness filters are based on diffusion frameworks, such as HDCS (Hybrid Diffusion with Continuous Switch) [15] that handles the trade-off between edge enhancement diffusion and coherence-enhancing diffusion depending on the local geometry. Other strategies that exploit the Frangi vesselness to choose smoothing directions [16] may also be mentioned. These filters have good performances, but require additional parametrization and generally induce higher computational costs.

A recent trend in the context of angiographic image preprocessing for further deep learning-based segmentation consists of designing features that provide necessary conditions of vesselness. Indeed, based on the assumption that many features may be combined as input of a neural network architecture [17], an operator that carries only a part of the information characterizing vessels can be relevant. A representative example of such operators is LIOT [18] that defines an ordering of intensity values in principal orientations, thus providing a qualitative assessment of gradient-like information. (It is worth mentioning that this approach presents several common properties with RORPO, namely the use of orientation-based responses and the ordering of these responses, but that it provides 2^d output images in dimension d, by contrast with RORPO that provides only one image.) These filtering approaches are generally not natively designed to characterize vessels, but more often to tackle issues related to deep learning image analysis, e.g. the robustness to inter-dataset contrast variation. Such operators, although useful for certain kinds of segmentation approaches, are not exactly vesselness operators and they fall out of the scope of our study.

B. Motivation and purpose

Vesselness filters are numerous, but none of them is able to perform optimally on all modalities and organs, especially as the definition of optimality depends on the targeted application. In this context, choosing the right filter for a specific task is difficult. Moreover, in the literature, most filters are tested either on toy examples or private datasets, which forbids the reproduction of the results and the comparison with new filters.

Our purpose is to tackle these issues. In the proposed framework, we compare seven vesselness filters, namely Sato, Frangi, OOF, Meijering, Jerman, Zhang and RORPO, described in Sec. II-D, based on a two-step parameter optimization process discussed in Sec. II-E.

These filters are evaluated on three public datasets, described in Sec. II-A: the "3D-IRCADb 01" CT dataset available on the Ircad website[1], the Bullitt/TubeTK MRA dataset available on the Kitware website[2] and a synthetic dataset[3] from the VascuSynth software [19].

The Ircad dataset presents vessels of widely different shapes and sizes. The Bullitt dataset exhibits long and tortuous small vessels. The modified VascuSynth dataset exhibits challenging tissues and vessels contrast. This synthetic dataset was added to include images with a controlled environment and complete ground truth (especially for small vessels and bifurcations), compared to real datasets with imperfect ground truth.

We study the behaviour of these filters in 6 regions of interest (ROIs) described in Sec. II-B. First, we assess the results in a global area to account for all possible sources of errors. Second, we focus on the vessel neighbourhood to ignore errors that may appear far from the vessels and may be irrelevant in some applications (e.g. if a post-processing is available to avoid border effect). We also refine this analysis by studying the filter performances separately depending on the size of the vessels (large, medium, small). Finally, we explore the bifurcation areas, where vesselness filters are usually assumed less efficient.

No additional pre-/post-processings have been added in this benchmark, as our goal is to compare the actual filters, with no bias induced by other operators.

The flow of our proposed framework is illustrated in Fig. 2 and Alg. 1.

C. Related works and novelty

Two vesselness filter benchmarks were proposed in the literature, by Luu et al. in 2015 [20], and by Phellan and Forkert in 2017 [21]. Our benchmark differs on several points.

First, reproducibility and reusability is a cornerstone of our proposal. We designed a modular and re-usable benchmark framework which allows the user to:

- apply and compare several vesselness filters;
- optimize the parameters of each filter with respect to a chosen metric;

[1] https://www.ircad.fr/fr/recherche/3d-ircadb-01-fr

[2] https://data.kitware.com/\#collection/591086ee8d777f16d01e0724/folder/58a372fa8d777f0721a64dfb

[3] https://vascusynth.cs.sfu.ca/Data.html

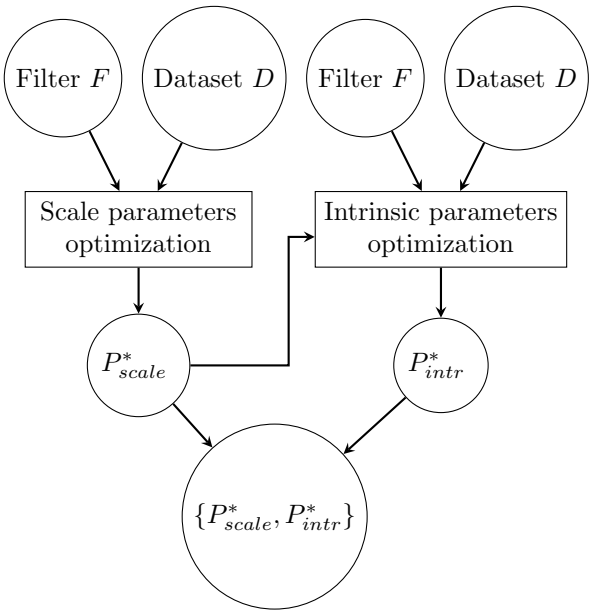


Fig. 2. Flowchart of the parameter optimization strategy. $\boldsymbol{P^\star}$ is the set of optimal parameters for filter $\boldsymbol{F}$ on dataset $\boldsymbol{D}$.

- evaluate the filter performances in specific ROIs;
- easily add extra vesselness filters via a Command Line Interface (CLI);
- test filters on any images from the online demonstration (without any installation)[4];
- reproduce and share results through structured and parametrized benchmark experiments;
- aggregate metric values, computed for each ROI, with possible export for further exploitation.

We chose well-known, public datasets representing different applications and image modalities (hepatic CT images and brain MRA images), as well as a synthetic dataset modified to create volumes with realistic, non-homogeneous contrast and intensities between tissues. We aim at creating a challenging environment in order to highlight the differences between the behaviours of the assessed filters. We also consider three filters not evaluated in the previous benchmarks, namely Meijering, Jerman and Zhang. Additionally, we optimize the filter parameters (scale and intrinsic parameters) on all volumes of each dataset, which was not considered in previous benchmarks. Finally, we investigate the bifurcation enhancement, which is an important, yet infrequent topic on whole datasets.

Our main contributions are then the following: a C++ benchmark framework, a reproducible quantitative comparison of seven optimized vesselness filters and an anatomical multi-scale analysis based on key ROIs from the whole organ down to the vessels bifurcations.

This article is an improved version of the conference paper [22] and includes novel elements. Regarding the data, the isotropic preprocessing of the Ircad images has been improved, leading to a higher resolution, whereas the VascuSynth images now contain realistic artifacts, leading to more challenging data (contrast and noise). Moreover, a new MRA dataset has been added to assess the filter performances on another application and image modality. Regarding the framework, we improved the parameter optimization scheme to be more realistic by mimicking the modus operandi usually considered in real applications. We also corrected a bias existing in our previous experiments by refining our analysis within the vessel neighbourhood with respect to the size of the vessels. We added a new metric, the PSNR (Peak Signal-to-Noise Ratio), in order to evaluate the filter performances independently of a thresholding process. Finally, we give a correct 3D formulation of Meijering filter with a proof of the parameter optimality (Appendix I).

[4]https://kerautret.github.io/LiverVesselnessIPOLDemo

Algorithm 1 Benchmark algorithm

Input:
Set of images $I = \{I_1, \ldots, I_N\}$
Set of ground truth $GT = \{GT_1, \ldots, GT_N\}$
Filter F
Region of interest ROI
Parameter set $P = \{P_{scale}, P_{intr}\}$
Metric M
Algorithm:
for i in $[1, N]$ **do**
$R_i \leftarrow ApplyFilter(F, I_i, P)$
$R_i^{masked}, GT_i^{masked} \leftarrow ApplyROI(R_i, GT_i, ROI)$
$m_i \leftarrow ComputeMetric(R_i^{masked}, GT_i^{masked}, M)$
end for
Output:
Mean metric value $\frac{1}{N}\sum_i m_i$

II. Material and Methods

A. Data

1) Ircad dataset: The Ircad dataset is composed of 20 3D chest CTA images (10 males, 10 females) in DICOM format. The image resolution range from 0.56 to 0.87 mm for the x, y axes and from 1.00 to 4.00 mm for the z axis. The size of the images is 512×512 voxels in the axial plane and ranges from 74 to 260 voxels for the z axis. The dataset is endowed with manual segmentations of organs, including the liver and both vena cava and portal vein as well as tumors.

Even thought some defects can be observed in this dataset, such as inconsistent naming conventions and erroneous voxels in ground truth masks (overestimated and/or missed vessels), the Ircad dataset is currently the best public dataset of CT scans available for liver vessel segmentation.

2) Bullitt dataset: The dataset available in the TubeTK framework website consists of 100 healthy subjects imaged by a 3T MRI scanner. It provides images for T1 ($1 \times 1 \times 1$ mm^3), T2 ($1 \times 1 \times 1$ mm^3), DTI ($2 \times 2 \times 2$ mm^3) and MRA ($0.5 \times 0.5 \times 0.8$ mm^3). In particular, 33 MRA images are provided with their manually annotated centerlines and radii. The volumic vessel annotations can be reconstructed using the TubeTK toolkit from these provided annotations. This automatic reconstruction leads to highly underestimated volumic vessel annotations. Sanchesa *et al.* [23] manually corrected these reconstructions, leading to better volumic annotations. We used these annotations in our benchmark.

Finally, as this dataset was acquired for research purpose, the images present a strong quality (very few artifacts, very

high contrast). In particular, contrary to most angiographic images acquired in clinical conditions, these images do not need filtering. As this dataset is the only public MRA dataset with annotation available, we chose to include it in the benchmark, but we altered the image quality to mimic clinical conditions. In particular, we added artifacts similar to surface coil flare and Rician noise with $\sigma = 4$.

3) VascuSynth dataset: The VascuSynth dataset is composed of 120 images of simulated vascular trees with varying bifurcation numbers. The image size is $101 \times 101 \times 101$ voxels with an isotropic 1.00 mm resolution. The dataset is composed of vascular trees with a constant null value background and a listing file providing vessel and bifurcation locations for each image. We modified this dataset in order to integrate MRA properties. More precisely, we added four types of artifacts: a gradual variation of vessels intensity, a global variation of intensities due to the lack of homogeneity of the magnetic field, Rician noise typical of MRI [24] and bright Gaussian artifacts to simulate non-vessel structures and make the segmentation task more complex.

A VascuSynth image with such artifacts is shown in Fig. 1. These modifications lead to three variations of the synthetic dataset with increasing noise level ($\sigma \in \{2, 4, 6\}$).

B. Regions of interest

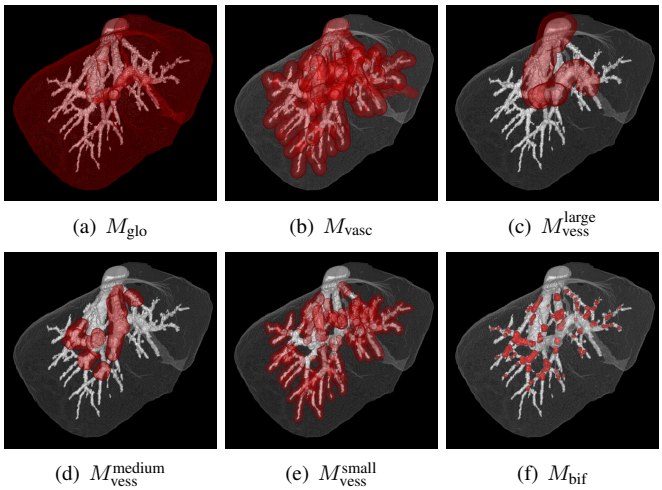
Analysing the results in several ROIs provides a useful insight on the filter performances depending on the targeted application and the available post-processing strategy. We observed in our previous work [22] that some filters may induce false positive responses in areas far from the vascular structures (e.g. the organ borders), whereas providing good results in the vascular areas. When a global automatic filtering / segmentation strategy is used, such filters may not be a good choice. However, if semi-automatic or manual post-processings are available, or if the goal is vessel tracking, then these filters may be relevant.

We also observed that some filters may have good performances on large vessels but behave poorly on small ones, which still leads to good average performances. In some applications, a good detection of small vessels is of very high interest. In this case, it is important to rely on metrics computed specifically on small vessels, instead of on the whole image, to choose an adequate filter.

Finally, the bifurcation areas are often poorly enhanced by filters. When considering tracking applications, such areas are of paramount interest to handle the topological modifications of centerlines.

Based on these observations, we study the behavior of each filter on 6 ROIs defined as binary masks, exemplified in Fig. 3:

- a global mask M_{glo}, that corresponds to the organ of interest volume (the liver for the Ircad dataset, the brain for the Bullitt dataset and the whole image for the VascuSynth dataset);
- a vascular neighbourhood mask M_{vasc}, corresponding to the union of areas inside the vessels and areas close to the vessels (see details on paragraph ROI construction);



(a) M_{glo} (b) M_{vasc} (c) $M_{\text{vess}}^{\text{large}}$

(d) $M_{\text{vess}}^{\text{medium}}$ (e) $M_{\text{vess}}^{\text{small}}$ (f) M_{bif}

Fig. 3. The six ROIs (in red), illustrated on an Ircad image (in white).

- scale-based vascular neighbourhood masks $M_{\text{vess}}^{\text{large}}$, $M_{\text{vess}}^{\text{medium}}$ and $M_{\text{vess}}^{\text{small}}$ which provide a partition of M_{vasc} into three areas with respect to the radius of the vessels;
- a mask M_{bif} which focuses on the vessel bifurcations.

In particular, we have $M_{\text{bif}} \subset M_{\text{vasc}} \subset M_{\text{glo}}$ and $M_{\text{vasc}} = M_{\text{vess}}^{\text{large}} \cup M_{\text{vess}}^{\text{medium}} \cup M_{\text{vess}}^{\text{small}}$.

ROI construction: The vascular neighbourhood masks are built with respect to the minimal, mean and maximal vessel diameters of each dataset: $[0, 3]$, $]3, 6]$ and $]6, \infty[$ mm, respectively for the Ircad dataset, and $[0, 1]$, $]1, 2]$ and $]2, \infty[$ mm, respectively for the VascuSynth dataset. The brain vessel diameters in the Bullitt dataset do not vary much; thus we used only two masks: $[0, 0.513]$ mm ($M_{\text{vess}}^{\text{small}}$) and $]0.513, \infty[$ mm ($M_{\text{vess}}^{\text{medium}}$).

The vascular (M_{vasc}) and vascular neighbourhood ($M_{\text{vess}}^{\text{small}}$, $M_{\text{vess}}^{\text{medium}}$ and $M_{\text{vess}}^{\text{large}}$) masks are obtained by dilating the vessel ground truth by a ball with a diameter which varies according to the size of the vessels. These values were experimentally set to 9, 7, 5 voxels for the Ircad dataset; 5, 3 voxels for the Bullit dataset; and 7, 5, 3 voxels for the VascuSynth dataset, respectively. When two vascular neighbourhood masks overlap, the overlapping region is assigned to the mask corresponding to the larger vessel neighbourhood. This guarantees that the three masks constitute a partition of M_{vasc} while preserving the coherence of the induced subdivisions.

The M_{bif} mask is built from the bifurcation points extracted from the annotated centerlines, which are dilated by a factor $k\rho$, where ρ is the vessel radius and $k = 3$ if $\rho \leq 1$ voxel and 2 otherwise. An intersection with the volumic vessel ground truth is finally performed to ensure that M_{bif} is included in the vessels.

C. Metrics

The benchmark implements several metrics computed from the confusion matrix between the thresholded filter output and the corresponding binary ground truth: true positives (tp), true negatives (tn), false positives (fp), false negatives (fn).

We compute the Dice and the Matthew's correlation coefficients (MCC). The Dice (or F_1 score) is more common in the literature; however as highlighted by Chicco and Jurman

[25], the MCC accounts for true negatives and is a more stable measure when it comes to sparse structures such as vessels:

$$\text{Dice} = \frac{2\text{tp}}{\text{fp} + \text{fn} + 2\text{tp}}$$

$$\text{MCC} = \frac{\text{tp} \cdot \text{tn} - \text{fp} \cdot \text{fn}}{\sqrt{(\text{tp} + \text{fp})(\text{tp} + \text{fn})(\text{tn} + \text{fp})(\text{tn} + \text{fn})}}$$

We also compute the Peak Signal-to-Noise Ratio (PSNR), which does not depend on the thresholding of the filter output. The PNSR is usually computed to evaluate a denoising result, which is not the case here, as the ground truth is binary. In this benchmark, we use the PSNR as a similarity metric defined as follows:

$$\text{PSNR} = \log_{10}\left(\frac{(\max_x I(x))^2}{\text{MSE}(I_{\text{GT}}, I_{\text{Filter}})}\right)$$

with I the normalized input image, I_{Filter} the normalized filtered image, I_{GT} the binary ground truth image, and MSE the mean squared error.

D. Filters

Most vesselness filters rely on the following assumptions. First, vessels are bright structures surrounded by a darker environment. Second, a multiscale analysis is required to capture vessels of different sizes. Third, the induced scale space is coupled with a measure able to differentiate vessels from other structures.

1) Scale space: Different scale analyses have been proposed in the literature. In this work, three types of scale spaces have been involved.

The first scale space is the one defined by [26] which shows the relationship between Gaussian smoothing and details perception. The stronger the smoothing, the less present the small structures. In particular, for tubular structures, the standard deviation σ of the Gaussian kernel is assumed proportional to the diameter of the putative tubular object. This Gaussian scale space is most of the time coupled with Hessian-based filters. Indeed, the curvature information is a good descriptor of tubularity. Such information is contained in the Hessian matrix defined for a 3D point (x_1, x_2, x_3) as:

$$H(f) = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \frac{\partial^2 f}{\partial x_1 \partial x_3} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \frac{\partial^2 f}{\partial x_2 \partial x_3} \\ \frac{\partial^2 f}{\partial x_3 \partial x_1} & \frac{\partial^2 f}{\partial x_3 \partial x_2} & \frac{\partial^2 f}{\partial x_3^2} \end{bmatrix}$$

Both operations can be combined using convolutions of second order Gaussian kernels. This process also ensures that the digital image is continuous and twice differentiable.

This strategy tends to overestimate the tubular object size. This is a drawback in various cases, such as adjacent tubular structures, which tend to be merged. The optimally oriented flux (OOF) approach [12] proposes an alternative framework in which the scale space is defined by the radius r of a sphere. A matrix describing the local geometry is computed from the outward gradient vector field of the image over the boundary of this sphere S_r. The authors formulated an optimization problem aiming at finding an optimal projection direction $\widehat{\rho}$ by minimizing the inward oriented spherical flux defined as:

$$f(x; r, \widehat{\rho}) = \int_{\delta S_r} ((v(x + r\widehat{n}).\widehat{\rho})\widehat{\rho}).\widehat{n}\, dA = \widehat{\rho}^T Q_{r,x} \widehat{\rho}$$

with $v(.)$ the image gradient, dA an infinitesimal area on the surface δS_r and $\widehat{n}$ the outward unit normal of δS_r. The ith row and jth column of Q are defined as:

$$q_{r,x}^{i,j} = \int_{\delta S_r} ((v_i(x + r\widehat{n}).\widehat{\rho})\widehat{\rho}).n_j\, dA = \widehat{\rho}^T Q_{r,x} \widehat{\rho}$$

Law and Chung proposed an efficient implementation in the Fourier domain. The solution matrix Q of this minimization problem is analogue to the Hessian without its disadvantages. In this framework, the sphere radius r controls the scale at which the vessels are detected.

Another way to define a scale space is to use morphological granulometry. In mathematical morphology, the opening operator is a composition of two basics operators: an erosion followed by a dilation with respect to a shape called structuring element. This operator acts as a filter that will preserve only structures large enough to include the structuring element. In this context, path opening [27] defines a set of flexible structuring elements defined over a grid of adjacency that controls the element shape. This scale space differs from the two previous ones, as the scale here refers to the length of the paths, instead of the vessel cross-section.

In practice, scale spaces are expressed via logarithmic scale. Log scale spaces are computed as $\sigma_i = \exp(\log(\sigma_{min}) + Factor.i)$ with $i \in [0, NbScales]$. If the scale space is defined as a minimal and maximal bounds and number of scales $[\sigma_{\min}, \sigma_{\max}, NbScales]$, then $Factor$ can be deduced from $Factor = \log(\sigma_{\max} - \sigma_{\min})/(NbScales - 1)$.

2) Tubularity measures: One of the most common ways to define tubularity is to use the eigenvalues of the Hessian. Let $\mathbf{e_1}$, $\mathbf{e_2}$ and $\mathbf{e_3}$ be the three eigenvectors of $H(f)$, associated to the eigenvalues λ_1, λ_2 and λ_3, respectively, with $|\lambda_1| \leqslant |\lambda_2| \leqslant |\lambda_3|$. In terms of eigen analysis, tubularity can be expressed as follows [28]:

$$|\lambda_1| \approx 0$$
$$\lambda_2 \approx \lambda_3 \ll 0$$

Sato *et al.* [7] were among the first to use this formulation. (N.B.: the sorting order is slightly different for Sato's eigenvalues: $\lambda_i^\star$ such that $\lambda_1^\star \geqslant \lambda_2^\star \geqslant \lambda_3^\star$.) When $\lambda_2^\star, \lambda_3^\star < 0$, the eigenvector $\mathbf{e_1^\star}$ associated to $\lambda_1^\star$ points to the direction of the least variation of intensity, which is also the direction of the vessel. Then, eigenvectors $\mathbf{e_2^\star}$ and $\mathbf{e_3^\star}$ form a basis orthogonal to $\mathbf{e_1^\star}$ and correspond to the vessel cross-section. The cross-section semi-axes lengths are proportional to $|\lambda_2^\star|$ and $|\lambda_3^\star|$, respectively. The Sato vesselness makes use of an asymmetric ratio of eigenvalues to get a high response in tubular structures, based on the sign of $\lambda_1^\star$. It has the advantage to smooth the filter response and suppress noise. Two parameters α_1 and α_2 control the strength of this asymmetry:

$$F = \begin{cases} \lambda_c^\star \exp(-\frac{\lambda_1^{\star 2}}{2(\alpha_1 \lambda_c^\star)^2}) & \lambda_1^\star \leqslant 0, \lambda_c^\star \neq 0 \\ \lambda_c^\star \exp(-\frac{\lambda_1^{\star 2}}{2(\alpha_2 \lambda_c^\star)^2}) & \lambda_1^\star > 0, \lambda_c^\star \neq 0 \\ 0 & \lambda_c^\star = 0 \end{cases}$$

with $\lambda_c^\star = \min\{-\lambda_2^\star, -\lambda_3^\star\}$.

One year later, Frangi *et al.* [8] exploited the three eigenvalues to define a measure with more control over the geometry of the enhanced patterns. Three measures are derived from these eigenvalues:

$$\begin{aligned} R_b &= |\lambda_1|/\sqrt{|\lambda_2\lambda_3|} \\ R_a &= |\lambda_2|/|\lambda_3| \\ S &= \sqrt{\lambda_1^2+\lambda_2^2+\lambda_3^2} \end{aligned}$$

which discriminate blobs (R_b), plate and line structures (R_a), and enhance low contrast structures thanks to the norm of the Hessian (S). These three measures are unified in the following vesselness function:

$$F = \big(1-\exp\big(-\frac{R_a^2}{2\alpha^2}\big)\big)\exp\big(-\frac{R_b^2}{2\beta^2}\big)\big(1-\exp(-\frac{S^2}{2C^2})\big)$$

if $\lambda_2, \lambda_3 \leqslant 0$ and $F=0$ otherwise. This function is controlled by three parameters α, β, C, making the Frangi filter the one requiring the highest parameter tuning of our benchmark. This method is the gold standard of numerous segmentation applications.

For the purpose of detecting neurites in fluoroscopy, Meijering *et al.* [9] proposed a filter to enhance elongated structures of one or two voxel width in a low contrast setting. This method is parameter-free and was initially proposed in 2D; it was then tested in 3D in [13]. However, in 3D, a formulation has never been explicitly documented, and we clarify it hereafter. The filter relies on a modified Hessian matrix $H'(f)$:

$$\begin{bmatrix} \alpha h_{11}+h_{22}+h_{33} & h_{12} & h_{13} \\ h_{21} & \alpha h_{22}+h_{11}+h_{33} & h_{23} \\ h_{31} & h_{32} & \alpha h_{33}+h_{11}+h_{22} \end{bmatrix}$$

The α parameter is used to steer the filter so that its ridge is maximally flat in one direction, which is obtained when $\alpha = -2/3$ (see the proof in Appendix I). The three eigenvalues of $H'(f)$ are expressed with respect to those of $H(f)$ as:

$$\begin{aligned} \lambda_1' &= \alpha\lambda_1+\lambda_2+\lambda_3 \\ \lambda_2' &= \lambda_1+\alpha\lambda_2+\lambda_3 \\ \lambda_3' &= \lambda_1+\lambda_2+\alpha\lambda_3 \end{aligned}$$

for $i \neq j \neq k \neq i$. The vesselness is then defined by:

$$F = \begin{cases} \lambda_{\max}/\lambda_{\min} & \lambda_{\max} < 0 \\ 0 & \lambda_{\max} \geqslant 0 \end{cases}$$

where $\lambda_{\max} = \max\{\lambda_1', \lambda_2', \lambda_3'\}$ is computed at each voxel, and $\lambda_{\min}$ is the minimum of all the voxel-wise λ over the image.

Jerman *et al.* proposed a vesselness function that aims to be more robust to bifurcations and easier to parametrize, taking inspiration from the volume aspect ratio metric used to detect nearly spherical tensors. The function is defined by:

$$F = \begin{cases} 0 & \lambda_2 \leqslant 0 \text{ or } \lambda_\rho \leqslant 0 \\ 1 & \lambda_2 \geqslant \lambda_\rho/2 > 0 \\ \lambda_2^2(\lambda_\rho-\lambda_2)\big(\frac{3}{\lambda_2+\lambda_\rho}\big)^3 & \text{otherwise} \end{cases}$$

where λ_ρ is a regularized parametric version of λ_3, defined to reduce the sensitivity to weakly contrasted regions:

$$\lambda_\rho = \begin{cases} \lambda_3 & \lambda_3 > \tau\max_x \lambda_3(x) \\ \tau\max_x \lambda_3(x) & 0 < \lambda_3 \leqslant \tau\max_x \lambda_3(x) \\ 0 & \text{otherwise} \end{cases}$$

with $\tau \in [0,1]$. This produces a more homogeneous response, even with non-homogeneous vessels intensity.

Zhang *et al.* proposed to improve Jerman vesselness in the context of hepatic vessel segmentation, more specifically on masked liver. To circumvent a high response of the filter at the liver borders, they used a K-means classification to roughly estimate vessel intensity, combined with a sigmoid filter to suppress other tissues. In addition, they slightly modified the Jerman vesselness function F by adding a multiplicative term $1-\exp(-\frac{3(\lambda_1^2+\lambda_2^2+\lambda_\rho^2)}{2\lambda_\rho})$.

Any vesselness function can be embedded into the OOF framework. For this benchmark, we choose the geometric mean used by its authors in their experiments on real data:

$$F = \begin{cases} \sqrt{|\lambda_2\cdot\lambda_3|} & \lambda_2, \lambda_3 < 0 \\ 0 & \text{otherwise} \end{cases}$$

RORPO [14] is built upon the scale space defined by path openings [27]. To capture curvilinear structures, the structuring elements are defined as paths on an adjacency grid, which gives flexibility over the detected geometry. The openings are computed with structuring elements in seven orientations of the 3D space, capturing objects of different shapes such as blobs, line-like structures and planes. A final step consists of classifying the different shapes using the responses in the seven orientations. Indeed, for tubular objects, all structuring elements are oriented along the same direction.

E. Parameter settings

Vesselness filter parameters can be classified in two types: the scale parameters that relate to the size of the sought vessels; and the filter intrinsic parameters, which are responsible for the ability to capture the vessel shape. For instance, Frangi vesselness has 3 parameters linked to the scale space (corresponding to the minimal and maximal vessel sizes, and the number of scales between both) and 3 parameters linked to the enhancement.

When $k \gg 1$ parameters have to be tuned for a given filter (e.g. $k=6$ for Frangi), finding an optimal set of parametric values in the induced k-dimensional space is computationally intractable. A frequently adopted solution consists of using the default parameters suggested by the authors. However, this is generally not a relevant policy. Here, we chose to optimize the parameters in two steps:

1) the optimization of the scale parameters using fixed default intrinsic parameters provided by the authors;
2) the optimization of the intrinsic parameters using the scale parameters previously optimized.

In both steps, the optimal parameters are the ones that yield the best average MCC in the global mask (M_{glo}) of the whole dataset. This two-steps strategy, that first sets the scale parameters, can be compared to the natural choice when in

TABLE I

SCALE PARAMETERS FOR FILTERS USING DIAMETER-BASED SCALE SPACE: FRANGI, SATO, MEIJERING, JERMAN, ZHANG, OOF (GRID SEARCH INFORMATION). THE CONDITION ENSURES THAT THE SPACING BETWEEN EACH SCALE i IS HIGHER THAN THE VOXEL RESOLUTION.

Ircad and VascuSynth			
Parameter	Interval	Step	Conditions
$\sigma_{\min}$	$[0.4, 1.8]$	0.4	
$\sigma_{\max}$	$[1.4, 3.4]$	0.4	$\sigma_{min,i} - \sigma_{max,i} > \frac{1}{6}$ mm
Number of scales	$[\![3, 4]\!]$	1	
Bullitt			
Parameter	Interval	Step	Conditions
$\sigma_{\min}$	$[0.2, 1.6]$	0.4	
$\sigma_{\max}$	$[1.2, 3.2]$	0.4	$\sigma_{min,i} - \sigma_{max,i} > \frac{1}{6}$ mm
Number of scales	$[\![3, 4]\!]$	1	

TABLE II

SCALE PARAMETERS FOR RORPO FILTER WITHOUT DILATION (GRID SEARCH INFORMATION). THE CONDITION ENSURES THAT THE SPACING BETWEEN EACH SCALE i AVOIDS THE COMPARISON OF TOO SIMILAR PARAMETERS RANGES.

Ircad			
Parameter	Interval	Step	Conditions
Min. scale	$[30, 150]$	10	
Factor	$[1.1, 1.6]$	0.1	$20 < \text{scale}_i - \text{scale}_j < 200$
Nb. scales	$[\![2, 4]\!]$	1	
Bullitt			
Parameter	Interval	Step	Conditions
Min. scale	$[30, 90]$	10	
Factor	$[1.1, 1.5]$	0.1	$20 < \text{scale}_i - \text{scale}_j < 200$
Nb. scales	$[\![2, 4]\!]$	1	
VascuSynth			
Parameter	Interval	Step	Conditions
Min. scale	$[10, 90]$	10	
Factor	$[1.1, 1.5]$	0.1	$9 < \text{scale}_i - \text{scale}_j < 100$
Nb. scales	$[\![2, 4]\!]$	1	

TABLE III

INTRINSIC PARAMETERS (GRID SEARCH INFORMATION). MEIJERING AND RORPO DO NOT HAVE INTRINSIC PARAMETERS.

	Parameter	Interval	Step
Frangi	α	$[0.2, 1.0]$	0.2
—	β	$[0.2, 1.0]$	0.2
—	C	$[0, 60]$	30
Sato	α_1	$[0.2, 1.0]$	0.2
—	α_2	$[1, 2]$	0.2
OOF	σ	$[0.1, 1.0]$	0.1
Jerman	τ	$[0.1, 1.0]$	0.1
Zhang	τ	$[0.1, 1.0]$	0.1

practice, a user sets the scale parameters depending of the known biological structure size. A grid search strategy was used on these parameters; it is summarized in Tables I–II (scale parameters) and Table III (intrinsic parameters). The optimization strategy is as follows. Each filter is applied on the N images of a dataset with a parameter set p_j yielding the (normalized) resulting images $(R_{i,j})_{i\in[\![1,N]\!]}$. Then, thresholdings between 0 and 1 with a step of 0.005 are applied to each $R_{i,j}$ leading to 201 binary results $(B^t_{i,j})_{t\in[\![0,200]\!]}$. Each binary result is compared to the ground truth and the filter performance on an image I_i with parameter set p_j is given by its best MCC $s_{i,j} = \max_t\{\text{MCC}(B^t_{i,j})\}$. Finally, for each filter, the best mean parameter set p_j is the one with the best mean MCC, i.e. $\frac{1}{N}\sum_i s_{i,j}$.

III. RESULTS

In this section, we present and discuss the qualitative and quantitative results of the different filters. In addition to the seven filters presented in Sec. II-D, we also consider a simple thresholding approach as baseline. This baseline output is the normalized thresholded result, where the threshold parameter is optimized the same way as for the other filter parameters (see Sec. II-E).

In the following, we start by analysing the results of the filters in the global mask M_{glo} (Sec. III-A). Then we discuss the results in the vascular masks M_{vasc}, $M^{\text{large}}_{\text{vess}}$, $M^{\text{medium}}_{\text{vess}}$, $M^{\text{small}}_{\text{vess}}$ (Sec. III-B). We finally focus on the results in the bifurcation areas (Sec. III-C).

A. Global results

The ROC curves of the seven filters are presented in Fig. 4 and the quantitative results are summarized in Tables IV–VI. Qualitative results are presented in Figs. 5–7.

Ircad: Globally, the MCC and Dice of all filters, computed on the whole Ircad images, are low (less than 0.5). This was expected as we only perform a filtering of complex images. This justifies the need for a subsequent segmentation on such images.

Qualitatively, all filters except RORPO falsely enhance the organ border. Meijering seems to yield the worst results by highly enhancing the liver border, while enhancing the vessels but also noise in the liver tissue. By comparison, the baseline retrieve well the larger vessels, but the quality of the vessels decrease as they become smaller and disconnections become more frequent. The baseline has difficulties retrieving the small vessels and we observe noisy bits instead.

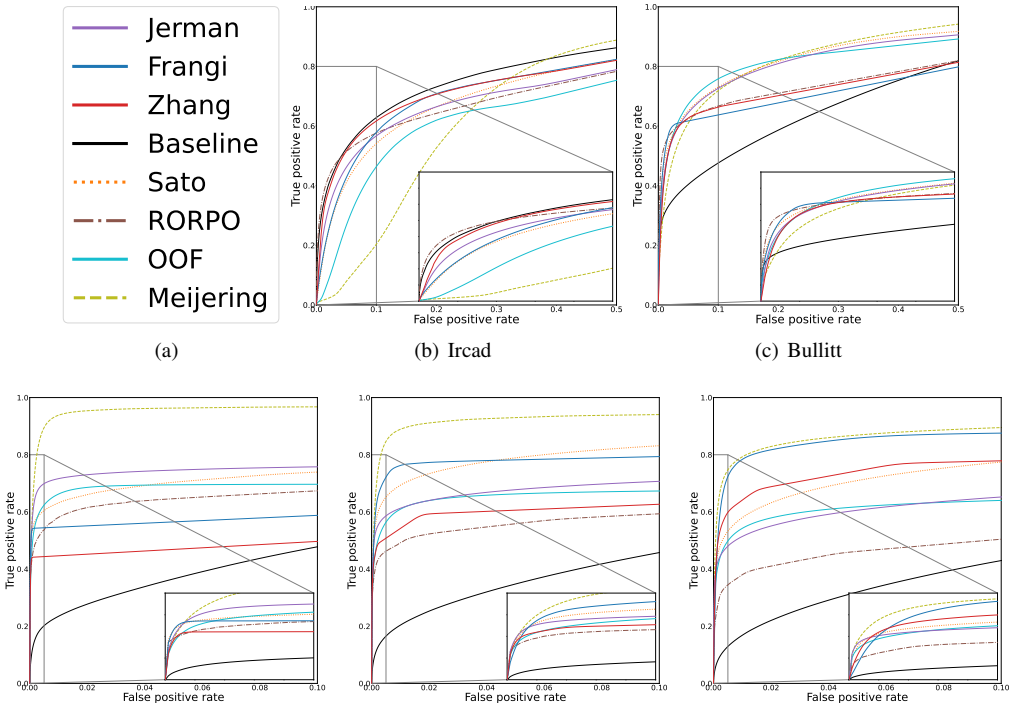
Quantitatively, RORPO yields the best results with a MCC of 0.475. The baseline obtains the second-highest MCC (0.452).

It is worth noting that, with regard to quantitative results, a simple thresholding performs better than most of the filters on these injected CTA images.However, this remark should be mitigated by the fact that the baseline is missing medium and small vessels in a larger proportion compared to the vesselness filters. In fact all filters are better at enhancing small and medium sized vessels.

Zhang provides the third best results (MCC $=$ 0.434), whereas OOF (MCC $=$ 0.277) and Meijering (MCC $=$ 0.232) have the lowest performances. The best two filters (RORPO and Zhang) show good results achieved by different means. RORPO yields a high precision (0.666) and an average sensitivity (0.379) whereas Zhang shows a high sensitivity (0.435) and an average precision (0.515).

One should keep in mind that these results are computed based on the best mean parameters on the whole dataset. The optimization process thus performs a trade-off between enhancing the vessels and decreasing the intensities of non-vessel structures (such as the liver border).

Bullitt: Qualitatively, RORPO seems to enhance the most vessels with a relatively low amount of noise. However, some poor contrasted vessels in the initial image showed irregular disconnections typical of an anti-extensive filter. Some



(a) (b) Ircad (c) Bullitt

(d) VascuSynth, $\sigma = 2$ (e) VascuSynth, $\sigma = 4$ (f) VascuSynth, $\sigma = 6$

Fig. 4. Mean ROC curves of the seven vesselness filters applied on (b) the Ircad dataset, (c) Bullitt dataset and (d–f) the VascuSynth dataset for three level of noise. ROC curves are zoomed, as vessels voxels represent $\mathbf{6.4\%}$ (b), $\mathbf{1.7\%}$ (c), and $\mathbf{0.2\%}$ (d–f) of the total number of voxels.

TABLE IV
QUANTITATIVE RESULTS (MEAN ± STD) IN THE GLOBAL MASK M_{GLO} ON THE IRCAD DATASET.

	MCC	Dice	PSNR
Baseline	0.452 ± 0.129	$\mathbf{0.468} \pm 0.126$	9.352 ± 1.247
Frangi	0.355 ± 0.075	0.392 ± 0.074	19.899 ± 1.624
Jerman	0.382 ± 0.060	0.415 ± 0.059	18.926 ± 1.186
Meijering	0.232 ± 0.036	0.241 ± 0.050	19.079 ± 1.392
OOF	0.277 ± 0.049	0.316 ± 0.055	19.728 ± 1.575
RORPO	$\mathbf{0.475} \pm 0.073$	0.477 ± 0.076	$\mathbf{20.349} \pm 1.687$
Sato	0.340 ± 0.056	0.380 ± 0.057	19.915 ± 1.633
Zhang	0.434 ± 0.085	0.462 ± 0.079	20.274 ± 1.648

TABLE V
QUANTITATIVE RESULTS (MEAN ± STD) IN THE GLOBAL MASK M_{GLO} ON THE BULLIT DATASET.

	MCC	Dice	PSNR
Baseline	0.396 ± 0.049	0.340 ± 0.061	20.275 ± 0.732
Frangi	0.474 ± 0.027	0.481 ± 0.026	21.768 ± 0.510
Jerman	0.432 ± 0.030	0.438 ± 0.029	19.723 ± 1.051
Meijering	0.349 ± 0.040	0.354 ± 0.043	21.905 ± 0.463
OOF	0.417 ± 0.029	0.424 ± 0.030	21.875 ± 0.491
RORPO	$\mathbf{0.543} \pm 0.021$	$\mathbf{0.540} \pm 0.023$	$\mathbf{21.909} \pm 0.497$
Sato	0.475 ± 0.026	0.473 ± 0.028	21.799 ± 0.466
Zhang	0.423 ± 0.037	0.431 ± 0.037	21.261 ± 0.847

methods enhance noise in the brain tissue more than others, such as Jerman, Sato, Meijering, OOF. However, Jerman and Sato highly enhance the vessels, leading to an overall good contrast. The diameter of vess els is overestimated by Jerman, Zhang, Meijering and, to a lesser extent, OOF; this leads to the fusion of close vessels, *aka* kissing vessels (see Fig. 7). The vessel enhancement pattern of Zhang is irregular (some vessels are highly contrasted, some are poorly contrasted) as Zhang is based on a K-means that introduces a different level of enhancement depending on the class associated to neighbor vessel pixels. The reader should note that we do not observe organ border artifacts in this dataset, as we eroded the brain mask in order to avoid unlabeled veins of the ground truth that would have biased our metrics.

Quantitatively, RORPO outperforms the other filters (MCC $= 0.543$). Sato (MCC $= 0.475$) and Frangi (MCC $= 0.474$) come respectively in second and third position. While close in MCC, Frangi recall is better than Sato (0.469 and 0.399) but with a lower precision rate (0.498 vs. 0.585). Jerman (MCC $= 0.432$), Zhang (MCC $= 0.423$) and OOF (MCC $= 0.417$) results are better than the baseline (MCC $= 0.396$) , while Meijering performs poorly (MCC $= 0.349$) because of a very high false positive rate.

VascuSynth: Qualitatively, Meijering, Sato and Jerman seem to yield the best results. However, Meijering tends to highly enhance the noise close to the vessels, yielding an irregular contour aspect. Jerman yields a good vessel enhancement while also enhancing a significant amount of noise. Frangi, Zhang and Sato seem to be the best methods for filtering out the Rician noise. In this dataset, RORPO is more sensitive to noise as the higher the noise, the higher the enhancement of blob-like artifacts. This was expected as we set the RORPO dilation parameter (which handles image noise) to 0. This choice was motivated by the fact that this parameter cannot be optimized separately from the scale parameters. In real applications with a high noise level, the dilation setting should be investigated.

As the VascuSynth dataset does not present organ borders or similar artifacts, the potential of false enhancement artifacts is reduced compared to real datasets, except for OOF which

TABLE VI

QUANTITATIVE RESULTS (MEAN ± STD) IN THE GLOBAL MASK M_{GLO} ON THE VASCUSYNTH DATASET.

	MCC			Dice			PSNR		
	$\sigma=2$	$\sigma=4$	$\sigma=6$	$\sigma=2$	$\sigma=4$	$\sigma=6$	$\sigma=2$	$\sigma=4$	$\sigma=6$
Baseline	0.184 ± 0.136	0.143 ± 0.116	0.106 ± 0.089	0.162 ± 0.134	0.122 ± 0.114	0.089 ± 0.087	9.411 ± 0.231	9.397 ± 0.230	9.374 ± 0.229
Frangi	$\mathbf{0.634} \pm 0.051$	0.577 ± 0.070	0.500 ± 0.081	$\mathbf{0.621} \pm 0.049$	0.572 ± 0.074	0.485 ± 0.091	26.274 ± 2.813	$\mathbf{26.496} \pm 2.872$	$\mathbf{26.692} \pm 2.856$
Jerman	0.611 ± 0.064	0.565 ± 0.049	0.501 ± 0.048	0.603 ± 0.065	0.549 ± 0.046	0.464 ± 0.048	$\mathbf{26.774} \pm 1.296$	21.758 ± 0.399	21.831 ± 0.489
OOF	0.627 ± 0.061	0.496 ± 0.065	0.449 ± 0.069	0.530 ± 0.060	0.476 ± 0.063	0.419 ± 0.067	26.324 ± 1.802	24.594 ± 1.329	22.983 ± 1.072
Meijering	0.538 ± 0.061	$\mathbf{0.603} \pm 0.059$	$\mathbf{0.565} \pm 0.060$	0.619 ± 0.064	$\mathbf{0.599} \pm 0.061$	$\mathbf{0.564} \pm 0.059$	26.586 ± 2.331	25.902 ± 1.889	24.821 ± 1.395
RORPO	0.587 ± 0.155	0.517 ± 0.119	0.366 ± 0.123	0.554 ± 0.157	0.476 ± 0.117	0.325 ± 0.113	23.236 ± 2.472	20.672 ± 1.689	18.372 ± 1.571
Sato	0.618 ± 0.046	0.559 ± 0.058	0.488 ± 0.052	0.596 ± 0.044	0.548 ± 0.058	0.464 ± 0.050	26.602 ± 2.539	26.241 ± 1.803	24.801 ± 1.285
Zhang	0.553 ± 0.052	0.523 ± 0.051	0.481 ± 0.065	0.531 ± 0.051	0.498 ± 0.049	0.474 ± 0.067	26.221 ± 2.805	26.360 ± 2.826	26.543 ± 2.845

TABLE VII

OPTIMAL PARAMETERS FOR THE SEVEN FILTERS AND THE THREE DATASETS.

	Parameter	Ircad	VascuSynth			Bullitt
	Parameter		$\sigma=2$	$\sigma=4$	$\sigma=6$	
Frangi	α	1	0.2	0.2	0.2	1.0
—	β	1	0.6	0.4	0.4	0.4
—	C	30	30	30	30	30
Jerman	τ	0.1	0.2	0.2	0.3	0.2
OOF	σ	1	0.1	0.1	0.1	0.1
RORPO	Scale min	70	50	50	40	20
—	Factor	1.3	1.1	1.2	1.4	1.3
—	nbScales	4	3	2	2	4
Sato	α_1	0.3	0.3	0.3	0.5	0.7
Sato	α_2	1.6	2	1	2	1.0
Zhang	τ	0.3	0.4	0.8	0.9	0.1

enhances the image border. This is explained by the vesselness chosen for the OOF framework, which does not differentiate plane-like from tube-like structures. Another vesselness such as Frangi could be used in the OOF framework to avoid this problem. Interestingly, when the noise is not well filtered out, the vesselness filters tend to induce false tubular patterns.

Quantitatively, Frangi yields the best results for $\sigma = 2$ with a MCC of 0.634, and Meijering for the other two levels with a MCC of 0.603 for $\sigma = 4$ and 0.565 for $\sigma = 6$.

These good performances of Meijering are explained by the fact that the geometry of the VascuSynth vessels exactly follows the hypothesis of the Meijering model (straight, constant-radius vessels). Sato yields the third best results for the noise level $\sigma = 2$ (MCC $= 0.618$) while Jerman performs better for higher levels of noise (MCC $= 0.565$ for $\sigma = 4$ and MCC $= 0.501$ for $\sigma = 6$). Nonetheless, both filters provide similar performances as their sensitivities are similar, independently of the noise level. The differences lie in their precision: for high level of noise (e.g. $\sigma = 4$), Jerman shows a higher precision (0.717 vs. 0.661) while Sato has a higher precision for low level of noise (0.810 vs. 0.689). Overall, Frangi is the method that best removes the noise for high noise level with a PSNR of 26.496 and 26.692 for $\sigma = 4$ and $\sigma = 6$, respectively. The baseline results are very poor because of the presence of high intensity artifacts and a high intensity non-homogeneous background in the VascuSynth data. This motivates the use of vesselness filtering for applications with similar hypothesis. Globally, RORPO performs the worst on the VascuSynth dataset. It is interesting to note that Zhang seems to be the most robust to filter noise, as its performances remain stable when the noise level increases.

B. Vessels neighbourhood

Quantitative results of the seven filters are presented in Tables VIII–X for the Ircad, Bullitt and VascuSynth datasets, respectively.

Ircad: All hessian based and equivalent filters performances drastically increase when they are computed in M_{vasc} since they falsely enhance artifacts far from the vessels, which is not taken into account in M_{vasc}. In this mask, Frangi and Zhang perform the best with both a MCC of 0.535. When looking at the filter performances according to the size of the enhanced vessels ($M_{\text{vess}}^{\text{large}}$, $M_{\text{vess}}^{\text{medium}}$, $M_{\text{vess}}^{\text{small}}$), we observe that Frangi is able to enhance large (MCC $= 0.580$) and medium vessels (MCC $= 0.619$) correctly, but its performances drop for small vessels (MCC $= 0.460$). Conversely, OOF and Jerman correctly enhance small vessels (MCC $= 0.514$ and MCC $= 0.502$) while yielding lower performances on large vessels (MCC $= 0.513$ and MCC $= 0.480$).

Bullitt: Locally, the results do not change in this dataset when evaluating them in M_{vasc}, compared with M_{glo}, as this dataset does not contain many of non-vessel structures that could be falsely detected. In this context, RORPO still performs the best, followed by Sato and Frangi.

VascuSynth: Globally, the vessel enhancement is best performed by Frangi, except for low level of noise where Meijering performs the best. A finer analysis of the results based on the size of the vessels reveals that Meijering and Jerman perform better than Frangi on medium and small vessels for low level of noise. In other cases Frangi perform much better on M_{vasc}.

C. Bifurcations

Filtering results in bifurcation areas are exemplified in Fig. 6 for the VascuSynth (a–i) and the Ircad (j–r) datasets while bifurcations and kissing vessels are considered in Fig. 7 for the Bullitt dataset.

Qualitatively, we observe that Frangi and Sato filters show a signal loss in the centre of bifurcations for the Ircad and Bulitt datasets, and on the side of the bifurcations for the VascuSynth dataset. This signal loss is not observed for the other filters. The shifted location of the signal loss for the VascuSynth dataset can be explained by the specific bifurcation geometry modelling. Indeed, new branches start from the side of a large main vessel, with a significant lower diameter; thus, the bifurcation area is shifted from the intersection of the main and secondary branches to the base of the secondary vessel. Regarding kissing vessels, some filters exhibiting an

TABLE VIII
QUANTITATIVE RESULTS (MEAN ± STD) BY VESSELS SIZE CATEGORIES ON THE IRCAD DATASET.

	Vessel neighbourhood			Large vessels		
	MCC	Dice	PSNR	MCC	Dice	PSNR
Baseline	0.491 ± 0.118	0.527 ± 0.110	13.110 ± 1.795	0.552 ± 0.130	0.597 ± 0.136	20.938 ± 2.637
Frangi	0.535 ± 0.073	**0.581** ± 0.065	19.989 ± 1.653	**0.580** ± 0.072	**0.627** ± 0.087	22.189 ± 1.867
Jerman	0.501 ± 0.054	0.521 ± 0.060	**21.464** ± 1.757	0.480 ± 0.065	0.496 ± 0.083	**24.119** ± 1.972
Meijering	0.451 ± 0.061	0.522 ± 0.049	20.091 ± 1.646	0.545 ± 0.055	0.669 ± 0.044	22.407 ± 1.850
OOF	0.498 ± 0.063	0.556 ± 0.051	19.912 ± 1.642	0.513 ± 0.060	0.574 ± 0.067	22.056 ± 1.850
RORPO	0.491 ± 0.066	0.501 ± 0.075	20.463 ± 1.765	0.491 ± 0.069	0.504 ± 0.080	22.580 ± 1.948
Sato	0.508 ± 0.054	0.542 ± 0.057	19.996 ± 1.679	0.512 ± 0.067	0.548 ± 0.086	22.130 ± 1.861
Zhang	**0.535** ± 0.064	0.551 ± 0.074	20.940 ± 1.857	0.541 ± 0.078	0.561 ± 0.101	23.199 ± 2.032
	Medium vessels			Small vessels		
	MCC	Dice	PSNR	MCC	Dice	PSNR
Baseline	0.509 ± 0.121	0.557 ± 0.117	21.250 ± 2.989	0.391 ± 0.103	0.424 ± 0.097	18.687 ± 2.209
Frangi	0.619 ± 0.115	0.660 ± 0.113	27.387 ± 2.554	0.460 ± 0.123	0.506 ± 0.118	26.624 ± 2.232
Jerman	0.604 ± 0.095	0.622 ± 0.110	**30.111** ± 3.155	0.502 ± 0.093	0.525 ± 0.104	**27.991** ± 2.120
Meijering	0.542 ± 0.085	0.602 ± 0.082	27.432 ± 2.474	0.419 ± 0.088	0.462 ± 0.077	26.723 ± 2.187
OOF	**0.642** ± 0.097	0.681 ± 0.097	27.334 ± 2.467	**0.514** ± 0.103	**0.559** ± 0.096	26.692 ± 2.251
RORPO	0.547 ± 0.102	0.573 ± 0.115	28.138 ± 2.637	0.417 ± 0.093	0.435 ± 0.104	27.157 ± 2.354
Sato	0.602 ± 0.096	**0.629** ± 0.105	27.437 ± 2.548	0.488 ± 0.092	0.522 ± 0.091	26.777 ± 2.277
Zhang	0.602 ± 0.110	0.619 ± 0.126	28.808 ± 3.119	0.481 ± 0.110	0.497 ± 0.124	27.471 ± 2.311

TABLE IX
QUANTITATIVE RESULTS (MEAN ± STD) BY VESSELS SIZE CATEGORIES ON THE BULLITT DATASET.

	Vessels neighbourhood		
	MCC	Dice	PSNR
Baseline	0.371 ± 0.038	0.341 ± 0.062	**22.291** ± 0.513
Frangi	0.415 ± 0.028	0.506 ± 0.026	21.641 ± 0.517
Jerman	0.377 ± 0.037	0.466 ± 0.029	21.990 ± 0.687
Meijering	0.288 ± 0.041	0.412 ± 0.045	22.076 ± 0.509
OOF	0.353 ± 0.026	0.456 ± 0.032	21.771 ± 0.506
RORPO	**0.506** ± 0.022	**0.556** ± 0.025	21.784 ± 0.506
Sato	0.435 ± 0.027	0.491 ± 0.029	21.698 ± 0.478
Zhang	0.348 ± 0.029	0.460 ± 0.026	21.430 ± 0.553
	Medium vessels		
	MCC	Dice	PSNR
Baseline	0.542 ± 0.129	0.555 ± 0.163	33.836 ± 2.891
Frangi	0.605 ± 0.034	0.684 ± 0.034	33.262 ± 2.827
Jerman	0.580 ± 0.048	0.660 ± 0.055	**34.066** ± 3.226
Meijering	0.402 ± 0.054	0.521 ± 0.067	34.034 ± 2.820
OOF	0.620 ± 0.049	**0.690** ± 0.044	33.621 ± 2.971
RORPO	**0.647** ± 0.047	0.712 ± 0.047	33.370 ± 2.882
Sato	0.594 ± 0.064	0.663 ± 0.088	33.072 ± 2.928
Zhang	0.577 ± 0.074	0.655 ± 0.101	33.789 ± 2.790
	Small vessels		
	MCC	Dice	PSNR
Baseline	0.358 ± 0.038	0.315 ± 0.057	**22.771** ± 0.591
Frangi	0.419 ± 0.024	0.491 ± 0.026	22.153 ± 0.609
Jerman	0.376 ± 0.031	0.445 ± 0.030	22.352 ± 0.863
Meijering	0.287 ± 0.040	0.382 ± 0.044	22.549 ± 0.593
OOF	0.353 ± 0.025	0.436 ± 0.038	22.275 ± 0.586
RORPO	**0.510** ± 0.023	**0.544** ± 0.027	22.307 ± 0.590
Sato	0.436 ± 0.024	0.473 ± 0.031	22.236 ± 0.572
Zhang	0.349 ± 0.022	0.440 ± 0.023	21.861 ± 0.680

extensive behaviour, such as Jerman, Meijering or Zhang tend to merge erroneously some parallel, close vascular structures. Other filters, such as Frangi, Sato, OOF or RORPO seem more robust to such difficulties (in particular, for RORPO, this is guaranteed by its anti-extensivity).

However, quantitatively, we do not observe a significant drop in performance in the bifurcation area (M_{bif}). Indeed, by optimizing the filter parameters, the threshold used to compute the metrics compensate this observed signal loss.

IV. DISCUSSION

Based on these experiments, we can formulate several guidelines regarding the choice of a filter depending on the task of interest.

We observed that Meijering overall performances are low on the real datasets, as it highly enhances noise. This behavior is consistent with the fact that Meijering is a vesselness designed to enhance vessels in very low-contrasted images for vessel tracking applications. However, Meijering presented good results on the synthetic VascuSynth dataset. Meijering is a good choice of vesselness filter in the following cases:

- the image of interest presents a very low constrast and/or the vessels geometry is overall straight without two much diameter variations;
- the user plans to apply a segmentation based on a tracking strategy and/or apply a post-processing to remove the false positives.

OOF, coupled with the geometric mean vesselness, presents overall low results on our datasets which can be explained by two facts: (1) the choice of the vesselness which discrimination power of plate-like and tubular structures is low; and (2) OOF is based on the hypothesis of circular vessel cross-sections, which is in practice not often the case, especially on the Ircad dataset. One should keep in mind, that OOF is a scale-space framework and different vesselnesses may then be used with it, depending on the application of interest. OOF is a good choice of framework in the following cases:

- vessels with a circular cross-section;
- vascular networks presenting many kissing vessel patterns.

Sato and Frangi both present good performances on our three datasets. These filters propose a good trade-off between sensitivity and specificity but present a loss of signal in bifurcations. Frangi tends to overestimate the vessel volume; however it yields the best results on large vessels. Frangi is a good choice of filter in the following cases:

- images with large to medium vessels;

TABLE X
QUANTITATIVE RESULTS (MEAN ± STD) BY VESSELS SIZE CATEGORIES ON THE VASCUSYNTH DATASET ($\sigma = 2$).

	Vessels neighbourhood			Large vessels		
	MCC	Dice	PSNR	MCC	Dice	PSNR
Baseline	0.392 ± 0.145	0.340 ± 0.172	25.198 ± 3.109	0.515 ± 0.293	0.491 ± 0.307	30.738 ± 1.278
Frangi	0.700 ± 0.037	0.688 ± 0.044	26.275 ± 2.814	0.757 ± 0.022	0.747 ± 0.025	32.939 ± 0.934
Jerman	0.710 ± 0.055	0.702 ± 0.067	$\mathbf{29.880} \pm 3.007$	0.735 ± 0.027	0.722 ± 0.031	36.312 ± 0.973
Meijering	$\mathbf{0.725} \pm 0.034$	$\mathbf{0.753} \pm 0.035$	27.305 ± 2.969	$\mathbf{0.818} \pm 0.028$	$\mathbf{0.834} \pm 0.028$	34.217 ± 0.978
OOF	0.648 ± 0.043	0.624 ± 0.053	28.275 ± 2.968	0.714 ± 0.025	0.696 ± 0.029	35.068 ± 0.980
RORPO	0.639 ± 0.082	0.619 ± 0.096	29.406 ± 3.407	0.713 ± 0.165	0.694 ± 0.197	$\mathbf{37.049} \pm 1.574$
Sato	0.661 ± 0.034	0.642 ± 0.042	26.938 ± 2.867	0.731 ± 0.024	0.719 ± 0.028	33.699 ± 0.945
Zhang	0.624 ± 0.042	0.594 ± 0.053	26.225 ± 2.809	0.713 ± 0.041	0.696 ± 0.050	32.893 ± 0.936
	Medium vessels			Small vessels		
	MCC	Dice	PSNR	MCC	Dice	PSNR
Baseline	0.415 ± 0.176	0.355 ± 0.202	26.699 ± 1.974	0.284 ± 0.125	0.218 ± 0.129	27.954 ± 3.858
Frangi	0.715 ± 0.035	0.698 ± 0.041	28.986 ± 2.026	0.683 ± 0.056	0.657 ± 0.069	30.328 ± 3.627
Jerman	0.708 ± 0.047	0.691 ± 0.057	$\mathbf{32.360} \pm 2.236$	$\mathbf{0.730} \pm 0.073$	$\mathbf{0.719} \pm 0.090$	$\mathbf{34.315} \pm 4.028$
OOF	$\mathbf{0.768} \pm 0.031$	$\mathbf{0.785} \pm 0.031$	30.159 ± 2.141	0.660 ± 0.072	0.652 ± 0.086	31.054 ± 3.723
Meijering	0.672 ± 0.040	0.646 ± 0.049	31.016 ± 2.152	0.615 ± 0.071	0.572 ± 0.091	32.202 ± 3.802
RORPO	0.670 ± 0.087	0.647 ± 0.101	32.281 ± 2.506	0.566 ± 0.108	0.523 ± 0.118	32.660 ± 3.946
Sato	0.675 ± 0.035	0.649 ± 0.042	29.659 ± 2.078	0.647 ± 0.047	0.615 ± 0.059	30.939 ± 3.674
Zhang	0.648 ± 0.044	0.616 ± 0.054	28.936 ± 2.024	0.580 ± 0.054	0.528 ± 0.071	30.278 ± 3.618

- the user needs an easy way to control the level of noise to filter out, at the price of loosing small vessels.

Sato is a good choice of filter in the following cases:

- images with a low-level of noise;
- vessels with circular cross sections;
- vacular networks presenting many kissing vessel patterns.

Jerman presents good performances with a high sensitivity and sharp vessel contours. However it tends to be more sensitive to noise than classic Hessian filters (i.e. Frangi and Sato), and share a same trend to overestimate the vessel volumes. Jerman is a good choice of filter in the following cases:

- images with small vessels;
- images with a low-level of noise.

Zhang is a modified version of Jerman, designed to be less sensitive to noise and border artefacts, but dedicated to a specific organ (namely, the liver). Thus, to yield good results, it requires prior knowledge on the distribution of the image intensities only available for some applications of interest (e.g. masked liver CT images). Zhang is a good choice of filter in the following cases:

- applications where a mask of the organ is available;
- a constant vessel contrast in the image.

RORPO achieves first place in the two real datasets, but yields poor results on the VascuSynth dataset. It accurately differentiates organ borders from tubular structures and never overestimates the vessel volume. However RORPO favors specificity over sensitivity and thus yields a higher number of false negatives than other filters. RORPO is a good choice of filter in the following cases:

- applications requiring precise vessel contours;
- applications with a low tolerance to false positives (e.g. no post-processing available).

In this benchmark, we compared all the filters using their respective optimal parameters for a fair comparison; see Table VII. However, one may not have access to these optimal parameters, and should tune them manually. In this case, the reader should take into account how easy it is to tune these parameters and favor filters yielding stable results to small parameter changes.

As an intrinsic parameter-free filter, Meijering is the easiest filter to tune. RORPO and OOF are also easy to tune as RORPO has only one intrinsic binary parameter (the dilation parameter) and OOF depends on only one smoothing parameter σ. Jerman and Zhang also have one intrinsic parameter to tune, τ, which controls both the vessel contour sharpness and the enhancement contrast. In most case, optimal τ had a low value resulting in a high homogeneous response.

Sato and Frangi are the more complex filters to tune. They have 2 and 3 intrinsic parameters respectively which are difficult to tune as they have not an intuitive meaning. These parameters directly influence the geometry of detected structures and the results are thus very sensitive to parameter changes. We found that optimal parameters for both filters were tuned to relax the tubularity constraint and enhance vessels with more varying shapes. For instance, Frangi performances may vary up to 11% between the default and the optimized parameters.

The average computation times of each filter on three volumes of each dataset are summarized in Table XI. All Hessian-based methods are similarly fast to compute (around 1 minute per volume), except for Zhang which adds a K-means step and is thus a bit longer to compute.

OOF and RORPO require much more computational power as their complexity is not linear with the size of the image volume. However, RORPO implementation is multi-threaded to reduce the computation time. OOF can also be implemented in GPU, as demonstrated by Law [29].

Finally, it is important to note that the experimental results highly depend on the dataset and its associated ground truth. Public annotated datasets are scarce and thus very precious. However, it is important to be aware of their limitations. In particular, the ground truth of the Ircad dataset lacks some large vessels, which may influence the computed metrics. Moreover, vessel diameters are usually overestimated, which

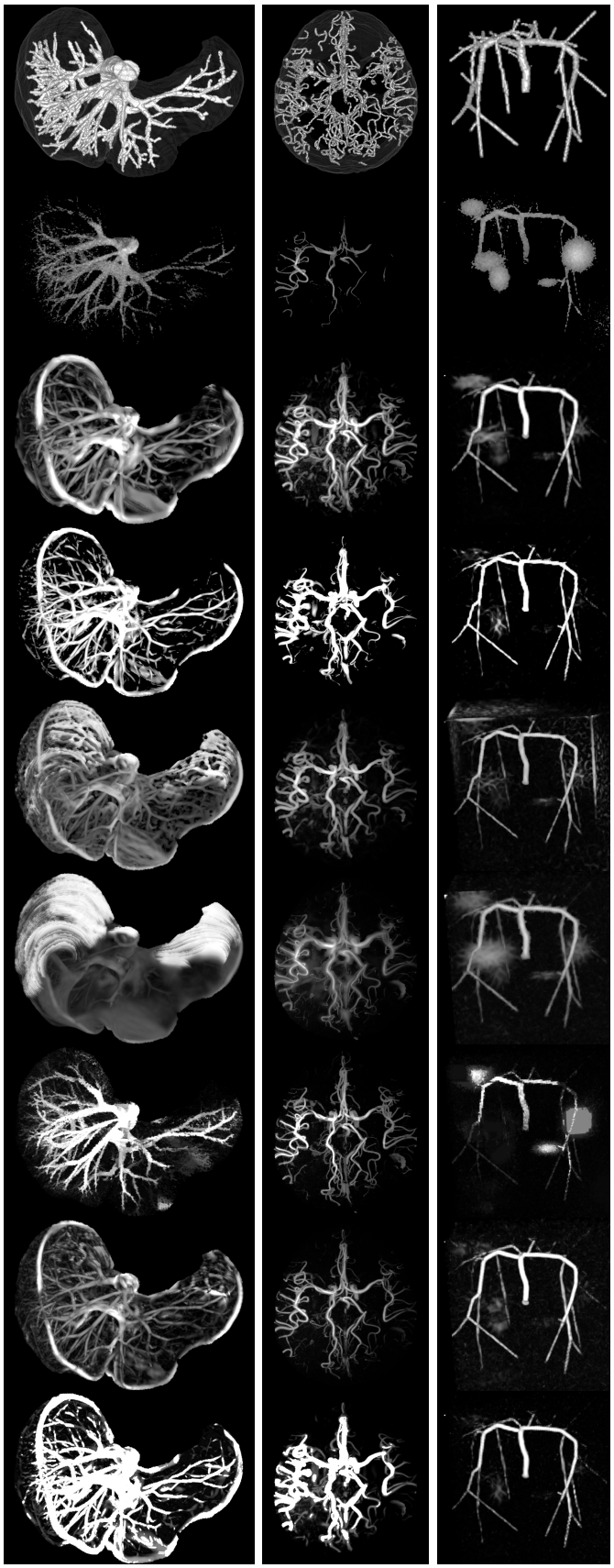


Fig. 5. Illustration of the filtering results on the Ircad (first column), Bullitt (second column) and VascuSynth (third column) datasets in $\boldsymbol{M}_{\text{glo}}$. First row is the ground truth, and the following rows are respectively the baseline, Frangi, Jerman, OOF, Meijering, RORPO, Sato and Zhang results.

TABLE XI
COMPUTATIONAL COST THE THE SEVEN FILTERS (CPU TIME IN SECONDS) MEAN OVER 3 VOLUMES OF EACH DATASET.

	Ircad	Bullit	VascuSynth
Frangi	72	47	6
Sato	67	44	5
Meijering	43	36	6
OOF	231	274	17
RORPO	1776	1227	160
Jerman	39	34	6
Zhang	106	80	10

adds a systematic bias (positive or negative depending on the filter) in the quantitative results. The VascuSynth dataset being synthetic, the vessel geometry is simpler than for real vessels, which impacts the filter results. Moreover, the background context of these images is much simpler than the one of real images that contain other organs, structures, etc.

V. CONCLUSION

In this article, we presented a benchmark framework to apply, compare and evaluate vesselness filters. This analysis can be done globally on whole images or on specific user-defined ROIs depending on the user application and needs. The benchmark includes: an automatic optimization strategy to help the user tune the filter parameters; and the up-to-date implementation of 7 state-of-the-art filters in a common framework. All the benchmark quantitative results, based on several metrics, are gathered in easily accessible formatted files. We made several recommendations on which filters should be used depending on the type of image and application of interest. The code of the benchmark and all associated files required to reproduce its results are available on *GitHub*: `https://github.com/JonasLamy/LiverVesselness` and an online demonstration allows to individually test each method on any other images: `https://kerautret.github.io/LiverVesselnessIPOLDemo`.

Finally, this benchmark, contrary to other proposed in the literature, is designed to be modular and extensible. The user may indeed easily add new filters, metrics or ROIs. We hope that it will help the community compare and analyse current and future vesselness filters.

Future works include adding other families of vesselness filters such as phase-based vesselness, and comparing the filters on different datasets. In the case of the hepatic vascular network, it would be relevant to compare the filters on real MRA datasets.

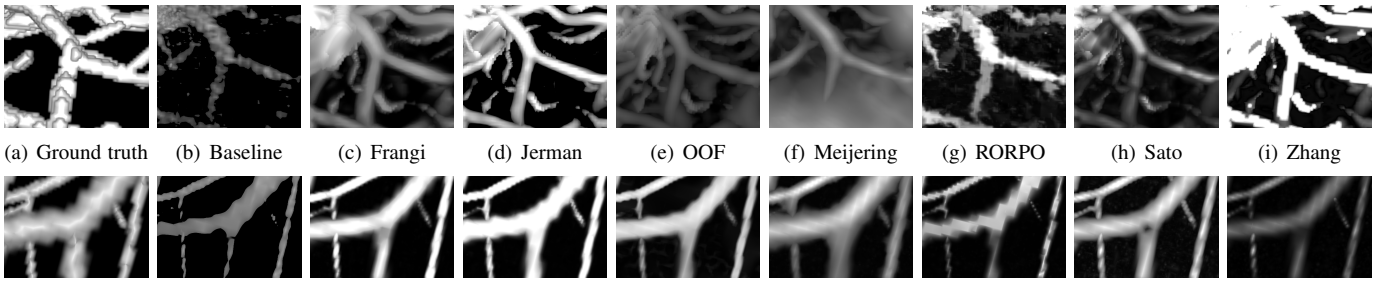
APPENDIX I
PROOF OF MEIJERING'S MAXIMAL FLATNESS (3D CASE)

We extend to 3D the 2D study performed in the Appendix of [30]. Up to translation and rotation, we can assume without loss of generality that the putative tubular structure is oriented in the x-axis of the 3D space $\mathbb{R}^3$ and that the point of interest is located at $\mathbf{x} = \mathbf{0}$. Following [30], we aim to have

$$\lim_{\mathbf{x}\to\mathbf{0}} (\mathbf{e}_x.\nabla)^2 h'(\mathbf{x}) = 0 \tag{1}$$

with

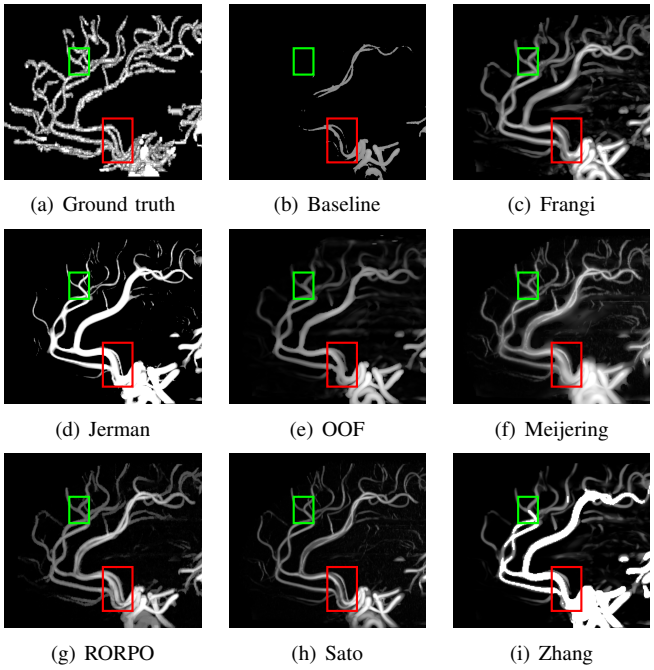
$$h'(\mathbf{x}) = (\alpha(\mathbf{e}_x.\nabla)^2 + (\mathbf{e}_y.\nabla)^2 + (\mathbf{e}_z.\nabla)^2)G(\mathbf{x})$$
$$G(\mathbf{x}) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2+z^2}{2\sigma^2}}$$



(a) Ground truth (b) Baseline (c) Frangi (d) Jerman (e) OOF (f) Meijering (g) RORPO (h) Sato (i) Zhang

(j) Ground truth (k) Baseline (l) Frangi (m) Jerman (n) OOF (o) Meijering (p) RORPO (q) Sato (r) Zhang

Fig. 6. Maximum Intensity Projection (MIP) view of the best filtering result on one bifurcation of the VascuSynth dataset with $\boldsymbol{\sigma} = \mathbf{2}$ (first row) and the Ircad dataset (second row). Baseline is the original image masked by the optimal thresholding.



(a) Ground truth (b) Baseline (c) Frangi

(d) Jerman (e) OOF (f) Meijering

(g) RORPO (h) Sato (i) Zhang

Fig. 7. Maximum Intensity Projection (MIP) view of central vessels in the Bullit dataset. Baseline is the original image masked by the optimal thresholding on M_{glo}. These vessels are located on the same plane, illustrating both kissing vessels (red) and bifurcations (green).

where $\{\mathbf{e}_x, \mathbf{e}_y, \mathbf{e}_z\}$ is the canonical basis of $\mathbb{R}^3$. We have

$$\begin{aligned} h'(\mathbf{x}) &= \alpha\frac{\partial^2 G(\mathbf{x})}{\partial x^2} + \frac{\partial^2 G(\mathbf{x})}{\partial y^2} + \frac{\partial^2 G(\mathbf{x})}{\partial z^2} \\ &= \frac{\alpha x^2 + y^2 + z^2 - (\alpha+2)\sigma^2}{\sigma^4} G(\mathbf{x}) = p(\mathbf{x}).G(\mathbf{x}) \end{aligned}$$

and

$$\begin{aligned} (\mathbf{e}_x.\nabla)^2 h'(\mathbf{x}) &= \frac{\partial^2 h'(\mathbf{x})}{\partial x^2} = \frac{\partial^2 p(\mathbf{x}).G(\mathbf{x})}{\partial x^2} \\ &= \frac{\partial^2 p(\mathbf{x})}{\partial x^2}.G(\mathbf{x}) + 2\frac{\partial p(\mathbf{x})}{\partial x}.\frac{\partial G(\mathbf{x})}{\partial x} + p(\mathbf{x}).\frac{\partial^2 G(\mathbf{x})}{\partial x^2} \end{aligned}$$

Since we have

$$\lim_{\mathbf{x}\to\mathbf{0}} G(\mathbf{x}) = \frac{1}{2\pi\sigma^2} \; ; \lim_{\mathbf{x}\to\mathbf{0}} \frac{\partial G(\mathbf{x})}{\partial x} = 0 \; ; \lim_{\mathbf{x}\to\mathbf{0}} \frac{\partial^2 G(\mathbf{x})}{\partial x^2} = -\frac{1}{2\pi\sigma^4}$$
$$\lim_{\mathbf{x}\to\mathbf{0}} p(\mathbf{x}) = -\frac{\alpha+2}{\sigma^2} \; ; \lim_{\mathbf{x}\to\mathbf{0}} \frac{\partial p(\mathbf{x})}{\partial x} = 0 \; ; \lim_{\mathbf{x}\to\mathbf{0}} \frac{\partial^2 p(\mathbf{x})}{\partial x^2} = \frac{2\alpha}{\sigma^4}$$

it comes

$$\lim_{\mathbf{x}\to\mathbf{0}} (\mathbf{e}_x.\nabla)^2 h'(\mathbf{x}) = \frac{3\alpha+2}{2\pi\sigma^6}$$

and Eq. (1) is solved iff $\alpha = -\frac{2}{3}$.

REFERENCES

[1] J. Lamy, *et al.*, "A benchmark framework for multiregion analysis of vesselness filters," *IEEE Transactions on Medical Imaging*, vol. 41, no. 12, pp. 3649–3662, 2022.

[2] G. D. Rubin, *et al.*, "CT angiography after 20 years: A transformation in cardiovascular disease characterization continues to advance," *Radiology*, vol. 271, pp. 633–652, 2014.

[3] D. G. Nishimura, A. Macovski, and J. M. Pauly, "Magnetic Resonance Angiography," *IEEE T Med Imaging*, vol. 5, pp. 140–151, 1986.

[4] D. Lesage, *et al.*, "A review of 3D vessel lumen segmentation techniques: Models, features and extraction schemes," *Med Image Anal*, vol. 13, pp. 819–845, 2009.

[5] S. Moccia, *et al.*, "Blood vessel segmentation algorithms — Review of methods, datasets and evaluation metrics," *Comput Meth Prog Bio*, vol. 158, pp. 71–91, 2018.

[6] D. Jia and X. Zhuang, "Learning-based algorithms for vessel tracking: A review," *Comput Medical Imaging Graph*, vol. 89, p. 101840, 2021.

[7] Y. Sato, *et al.*, "3D multi-scale line filter for segmentation and visualization of curvilinear structures in medical images," in *CVRMed-MRCAS*, 1997, pp. 213–222.

[8] A. F. Frangi, *et al.*, "Multiscale vessel enhancement filtering," in *MICCAI*, 1998, pp. 130–137.

[9] E. Meijering, *et al.*, "Neurite tracing in fluorescence microscopy images using ridge filtering and graph searching: Principles and validation," in *ISBI*, 2004, pp. 1219–1222.

[10] T. Jerman, *et al.*, "Enhancement of vascular structures in 3D and 2D angiographic images," *IEEE T Med Imaging*, vol. 35, pp. 2107–2118, 2016.

[11] R. Zhang, *et al.*, "An improved fuzzy connectedness method for automatic three-dimensional liver vessel segmentation in CT images," *J Healthc Eng*, vol. 2018, 2018.

[12] M. W. K. Law and A. C. S. Chung, "Three dimensional curvilinear structure detection using optimally oriented flux," in *ECCV*, 2008, pp. 368–382.

[13] B. Obara, *et al.*, "Contrast-independent curvilinear structure detection in biomedical images," *IEEE T Image Process*, vol. 21, pp. 2572–2581, 2012.

[14] O. Merveille, *et al.*, "Curvilinear structure analysis by ranking the orientation responses of path operators," *IEEE T Pattern Anal*, vol. 40, pp. 304–317, 2018.

[15] A. Mendrik, *et al.*, "Noise reduction in computed tomography scans using 3-D anisotropic hybrid diffusion with continuous switch," *IEEE T Med Imaging*, vol. 28, pp. 1585–1594, 2009.

[16] C. Cañero and P. Radeva, "Vesselness enhancement diffusion," *Pattern Recogn Lett*, vol. 24, pp. 3141–3151, 2003.

[17] S. Survarachakan, *et al.*, "Effects of enhancement on deep learning based hepatic vessel segmentation," *Electronics*, vol. 10, no. 10, 2021.

[18] T. Shi, *et al.*, "Local intensity order transformation for robust curvilinear object segmentation," *IEEE Transactions on Image Processing*, vol. 31, pp. 2557–2569, 2022.

[19] G. Hamarneh and P. Jassi, "VascuSynth: Simulating vascular trees for generating volumetric image data with ground-truth segmentation and tree analysis," *Comput Med Imag Grap*, vol. 34, pp. 605–616, 2010.

[20] H. M. Luu, *et al.*, "Quantitative evaluation of noise reduction and vesselness filters for liver vessel segmentation on abdominal CTA images," *Phys Med Biol*, vol. 60, pp. 3905–3926, 2015.

[21] R. Phellan and N. D. Forkert, "Comparison of vessel enhancement algorithms applied to time-of-flight MRA images for cerebrovascular segmentation," *Med Phys*, vol. 44, pp. 5901–5915, 2017.
[22] J. Lamy, *et al.*, "Vesselness filters: A survey with benchmarks applied to liver imaging," in *ICPR*, 2020, pp. 3528–3535.
[23] P. Sanchesa, *et al.*, "Cerebrovascular network segmentation of MRA images with deep learning," in *ISBI*, 2019, pp. 768–771.
[24] H. Gudbjartsson and S. Patz, "The Rician distribution of noisy MRI data," *Magn Reson Med*, vol. 34, pp. 910–914, 1995.
[25] D. Chicco and G. Jurman, "The advantages of the matthews correlation coefficient (mcc) over f1 score and accuracy in binary classification evaluation," *BMC genomics*, vol. 21, no. 1, pp. 1–13, 2020.
[26] T. Lindeberg, *Scale-Space Theory in Computer Vision*. Springer, 1994.
[27] H. J. A. M. Heijmans, M. Buckley, and H. Talbot, "Path openings and closings," *J Math Imaging Vis*, vol. 22, pp. 107–119, 2005.
[28] C. Lorenz, *et al.*, "Multi-scale line segmentation with automatic estimation of width, contrast and tangential direction in 2D and 3D medical images," in *CVRMed-MRCAS*, 1997, pp. 233–242.
[29] M. W. K. Law and A. C. S. Chung, "Efficient implementation for spherical flux computation and its application to vascular segmentation," *IEEE T Image Process*, vol. 18, pp. 596–612, 2009.
[30] E. Meijering, *et al.*, "Design and validation of a tool for neurite tracing and analysis in fluorescence microscopy images," *Cytometry A*, vol. 58, pp. 167–176, 2004.

Jonas Lamy graduated from the ENSICAEN engineering school in Caen, France, in 2019. He is a PhD candidate at the LIRIS Laboratory, Lyon, France since 2019. His work focuses on medical image processing and analysis, especially in the context of angiographic imaging. He is also interested in reproducible research and open-source software development.

Odyssée Merveille received her PhD from Paris Est University, Paris, France in 2016 (PhD award from the French *Association Française de Reconnaissance et Interprétation des Formes*, AFRIF). After two post-doctoral fellowships at Strasbourg University, Strasbourg, France, she has been an associate professor at INSA Lyon (CREATIS), Lyon, France, since 2019. She has a long-standing interest in the analysis and modelling of vascular networks to help clinicians diagnose and treat cardiovascular diseases. Her research interests include deep learning and convex optimization approaches, with a particular focus on open science and designing robust and reproducible tools for the community. She is the coauthor of 15 articles in top-ranking international journals and conferences.

Bertrand Kerautret is a Full Professor at the LIRIS Laboratory at Lyon (University Lyon 2) since 2018. He received a PhD degree in Computer Science from Université of Bordeaux in 2004 and Habilitation from Lorraine University in 2017. In 2005 he was Assistant Professor at the LORIA laboratory. His research deals with digital geometry, image processing and its applications. He worked on noise and scale detection on digital surface with applications to medical or wood imagery. He is also interested in reproducible research and was at the initiative of the first workshop on Reproducible Research in Pattern Recognition in 2016. He has coauthored about 50 papers in international journals or conferences. He actively contributes to the DGtal library and he is the manager of the DGtalTools project.

Nicolas Passat received the M.Sc and Ph.D. degrees from the Université Strasbourg 1, France, in 2002 and 2005, and habilitation degree from the Université de Strasbourg, France, in 2011. He was a postdoctoral fellow with the Université Marne-la-Vallée, France, between 2005 and 2006 and an assistant professor with the Université de Strasbourg, France, between 2006 and 2012. He is now a first-class professor with the Université de Reims Champagne-Ardenne, France. He is the coauthor of more than 150 articles and book chapters in the area of mathematical morphology, discrete geometry and topology, image analysis, medical imaging and remote sensing.