



HAL
open science

Vesselness filters: A survey with benchmarks applied to liver imaging

Jonas Lamy, Odysée Merveille, Bertrand Kerautret, Nicolas Passat, Antoine Vacavant

► **To cite this version:**

Jonas Lamy, Odysée Merveille, Bertrand Kerautret, Nicolas Passat, Antoine Vacavant. Vesselness filters: A survey with benchmarks applied to liver imaging. International Conference on Pattern Recognition (ICPR), 2020, Milan, Italy. pp.3528-3535, 10.1109/ICPR48806.2021.9412362. hal-02544493v2

HAL Id: hal-02544493

<https://hal.science/hal-02544493v2>

Submitted on 3 Oct 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Vesselness Filters: A Survey with Benchmarks Applied to Liver Imaging

Jonas Lamy*, Odyssee Merveille†, Bertrand Kerautret*, Nicolas Passat§, and Antoine Vacavant‡

* Université Lyon 2, LIRIS (UMR 5205), Lyon, France

† Univ Lyon, INSA-Lyon, Université Claude Bernard Lyon 1, UJM-Saint Etienne, CNRS, Inserm, CREATIS UMR 5220, U1206, F-69XXX, LYON, France

§ Université de Reims Champagne Ardenne, CReSTIC, EA 3804, 51097 Reims, France

‡ Université Clermont Auvergne, CNRS, SIGMA Clermont, Institut Pascal, F-63000, Clermont-Ferrand, France

Abstract—The accurate knowledge of vascular network geometry is crucial for many clinical applications such as cardiovascular disease diagnosis and surgery planning. Vessel enhancement algorithms are often a key step to improve the robustness of vessel segmentation. A wide variety of enhancement filters exists in the literature, but they are often difficult to compare as the applications and datasets differ from a paper to another and the code is rarely available. In this article, we compare seven vessel enhancement filters covering the last twenty years literature in a unique common framework. We focus our study on the liver vascular network which is under-represented in the literature. The evaluation is made from three points of view: in the whole liver, in the vessel neighborhood and near the bifurcations. The study is performed on two publicly available datasets: the Ircad dataset (CT images) and the VascuSynth dataset adapted for MRI simulation. We discuss the strengths and weaknesses of each method in the hepatic context. In addition, the benchmark framework including a C++ implementation of each compared method is provided. An online demonstration ensures the reproducibility of the results without requiring any additional software.

I. INTRODUCTION

A. Context

Vascular diseases are among the principal causes of death worldwide. This has motivated the development of a wide range of medical imaging technologies, dedicated to the visualization of vascular structures and the inner flowing blood, e.g. Magnetic Resonance Imaging (MRI) and X-ray Computed Tomography (CT). The induced angiographic images provide information of precious help for clinical applications, such as diagnosis and surgery planning. However, these images are complex to analyze, due to the sparseness of blood vessels, their complex topological and geometric properties, the low signal-to-noise ratio, and the presence of noise and artifacts inherent to the flowing blood signal. In this challenging context, computer-aided tools dedicated to angiographic image processing and analysis [1] are indeed mandatory to help the clinicians in their daily practice.

In particular, vessel segmentation—that consists of accurately extracting the whole vascular volume or the vessel medial axes—is a crucial task and an abundant literature deals with this issue. From a methodological point of view, many approaches were investigated. Non-exhaustively, one can cite model-based, region-growing, level-sets and machine

learning approaches. With the rise of computational power, deep-learning is also increasingly investigated. However, the lack of large annotated vascular datasets constitutes a current limit for such approaches in the 3D case. An exhaustive state of the art is beyond the scope of this article that focuses on vessels enhancement filters, and the interested reader is referred to [2] for a survey of the methods before 2010 and to [3] for a more recent survey that includes in particular machine/deep-learning approaches.

Beyond their methodological variability, most (3D) angiographic image segmentation methods share a common preprocessing step, namely a filtering used to reduce noise and non-vascular signal whereas increasing the vascular signal. The efficiency of this preliminary step has a strong influence on the accuracy of the segmentation carried out afterward. As a consequence, the development of so-called vessel enhancement filters has also been an active research field over the last twenty years.

B. Vessel Filtering

The most popular approaches in vessel filtering rely on differential information, and in particular the second derivatives of the angiographic images. Indeed, due to the contrast and geometric properties of the vessels, it is generally assumed that the observation of the local curvatures via the Hessian matrix analysis can allow for the determination of the position and orientation of the putative vascular structures. Pioneering works based on this paradigm were proposed by Sato *et al.* [4], and Frangi *et al.* [5] who developed so-called vesselness functions by combining eigenvalue information in a scale-space framework.

Many contributions then built upon this paradigm. Meijering *et al.* [6] proposed a neuriteness function to better detect long and thin tubular structures for neurite tracing in fluorescence microscopy images. To overcome the non-uniform response in different diameter vessels, Jerman *et al.* [7] proposed a volume ratio of tubular structures in the vesselness function. More recently, Zhang *et al.* [8] proposed to improve Jerman's method for CT data with an improved fuzzy affinity function.

Optimally Oriented Flux (OOF) was proposed by Law and Chung [9] to cope with the problem of fusion of closely located adjacent vessels of Hessian based framework. They

proposed an optimization framework based on image gradient flux that can be solved by an eigenvalue analysis. The resulting eigenvalues can then be involved in any classical vesselness function.

An alternative family of approaches makes use of morphological filters, based on openings by structuring elements. Merveille *et al.* [10] proposed to exploit this strategy by using orientation-based path opening responses, combined with a voting policy. Sazak *et al.* [11] proposed another related strategy by using openings with top-hat tensors and a combination of balls and cylinders structuring elements.

Finally, other approaches based on the phase tensors [12] or relying on wavelets [13] were proposed together with recent learning segmentation approaches [14], [15].

C. Purpose and Contributions

Performing comparisons of these enhancement methods is difficult if we only refer to the literature. Indeed, the filters are often compared by the authors against few other methods. The used datasets, parameters and metrics also vary from one paper to another.

Even if extensive reviews of vessel segmentation techniques exist [2], [3] they do not provide a common framework to evaluate performances on a specific image dataset. The reproduction of such algorithms is also complex and time consuming since implementations are not systematically provided by the authors.

Based on these considerations, our purpose is to propose a comprehensive comparison of seven classical vessel enhancement filters covering the literature of the last twenty years. More precisely, our contributions are the following: (i) we evaluate the filters with usual comparison criteria on whole images but also on focused areas (vessel neighborhood; bifurcations); (ii) we provide benchmark results on synthetic and real images with a focus on liver imaging that constitutes a target of increasing importance; (iii) we make available a reproducible implementation with the source codes and online demonstrations of all the selected methods to help the interested readers to reproduce our results and perform their own benchmark easily.

The compared vessel enhancement filters are Hessian-based methods, namely Sato [4], Frangi [5], Meijering [6], Jerman [7], Zhang [8] plus OOF [9] and RORPO [10]. In particular, we are interested in evaluating their robustness to CT and MRI artifacts and noise.

II. FILTERING METHODS

We briefly recall the various vessel enhancement methods investigated in the proposed evaluation framework. The purpose is not to exhaustively describe the different approaches, fully discussed in [4]–[10], but to emphasize the main similarities and differences between the considered filters.

As stated in Sec. I-B, most of the considered filtering methods rely on the second derivatives of the image intensities characterizing the curvature of the image structures. The

underlying hypothesis is that a vessel is a bright, elongated structure within a darker background.

Based on these facts, many vessel filtering approaches compute the Hessian matrix of the image. The Hessian matrix H of a function $f(x_1, x_2, x_3)$ is defined, at each point, as:

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

It is worth mentioning that f requires to be continuous and even twice differentiable, whereas the initial image I is digital, and then non-continuous / non-differentiable. To tackle this issue, it is generally chosen to define f from I by application of a Gaussian kernel of standard deviation σ . Convolving the image I with a Gaussian kernel also sets up a multiscale framework based on the linear scale-space paradigm [16] where the scale of the vessels of interest depends on σ .

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 $|\lambda_1| \leq |\lambda_2| \leq |\lambda_3|$. Locally, a vessel can be seen as a bright tubular structure on a dark background. In terms of eigen analysis, this is expressed by the following properties [17]:

$$\begin{aligned} |\lambda_1| &\approx 0 \\ \lambda_2 &\approx \lambda_3 \ll 0 \end{aligned}$$

A. Sato *et al.* [4]

The Sato vesselness is one of the first vessel enhancement approaches based on second derivative analysis. (N.B.: Sato *et al.* sort the eigenvalues λ_i^* such that $\lambda_1^* \geq \lambda_2^* \geq \lambda_3^*$.) In the Hessian based framework whenever $\lambda_2^*, \lambda_3^* < 0$, the eigenvector \mathbf{e}_1^* associated to λ_1^* corresponds to the direction of the putative vessel. Then, eigenvectors \mathbf{e}_2^* and \mathbf{e}_3^* form a basis of the vessel cross section where $|\lambda_2^*|$ and $|\lambda_3^*|$ represent the sizes of the cross section. The Sato vesselness introduces a ratio of the eigenvalues to get a high response in tubular structures and proposes an asymmetric formulation based on the sign of λ_1^* in order to improve the smoothness of the response and suppress noise. Two parameters α_1 and α_2 control the strength of this asymmetry:

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

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

B. Frangi *et al.* [5]

A second vesselness function—which is considered nowadays as the Hessian-based gold standard—relies on the hypothesis that using the three eigenvalues can be more discrim-

inative. Three measures are derived from these eigenvalues, namely:

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

which are dedicated to discriminate blobs (R_b); to distinguish plate and line structures (R_a); and to measure the norm of the Hessian to avoid enhancing low contrast structures (S). These three measures are involved in the following function:

$$F = \left(1 - \exp\left(-\frac{R_a^2}{2\alpha^2}\right)\right) \exp\left(-\frac{R_b^2}{2\beta^2}\right) \left(1 - \exp\left(-\frac{S^2}{2c^2}\right)\right)$$

if $\lambda_2, \lambda_3 \leq 0$ and $F = 0$ otherwise. This function is controlled by three parameters α, β, c .

C. Meijering et al. [6]

A third parameter-free vesselness function was developed to detect very elongated structures (e.g. neurites in fluoroscopy). This approach, initially proposed in 2D and then extended in 3D [12], relies on a modified Hessian matrix $H'(f)$ defined as:

$$\begin{bmatrix} h_{11} + \frac{\alpha}{2}(h_{22} + h_{33}) & (1 - \frac{\alpha}{2})h_{12} & (1 - \frac{\alpha}{2})h_{13} \\ (1 - \frac{\alpha}{2})h_{21} & h_{22} + \frac{\alpha}{2}(h_{11} + h_{33}) & (1 - \frac{\alpha}{2})h_{23} \\ (1 - \frac{\alpha}{2})h_{31} & (1 - \frac{\alpha}{2})h_{32} & h_{33} + \frac{\alpha}{2}(h_{11} + h_{33}) \end{bmatrix}$$

(generally with $\alpha = 1/3$). The three eigenvalues of $H'(f)$ are expressed with respect to those of $H(f)$ as:

$$\lambda'_i = \lambda_i + \alpha\lambda_j + \alpha\lambda_k$$

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

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

where $\lambda_{\max} = \max\{\lambda'_1, \lambda'_2, \lambda'_3\}$ is computed at each voxel, and λ_{\min} is the minimum of all the λ_{\max} of the image.

D. Jerman et al. [7]

Jerman *et al.* proposed a vesselness robust to bifurcations which builds upon the volume aspect ratio, a metric used to detect nearly spherical tensors. The function is defined by:

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

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

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

with $\tau \in [0, 1]$.

E. Zhang et al. [8]

In the specific context of hepatic vascular networks, Zhang *et al.* further proposed to improve Jerman vesselness by identifying the liver tissues and vessels mean intensity. A K-means classification was combined with a sigmoid filter dedicated to enhance the vessels whereas suppressing other tissues. In addition, they slightly modified the Jerman vesselness function F by adding a multiplicative term $1 - \exp\left(-\frac{3(\lambda_1^2 + \lambda_2^2 + \lambda_\rho^2)}{2\lambda_\rho}\right)$ in the ‘‘otherwise’’ case.

F. Optimally Oriented Flux (OOF) [9]

Alternatively to the Hessian-based analysis, Law and Chung proposed an optimization framework based on the projection of the image gradient on a sphere. The problem can be formulated as a generalized eigenvalue problem of a 3×3 matrix playing a similar role to the Hessian matrix in the five above vesselness functions. In particular, the sphere radius has an analogue role to σ with respect to scale-space analysis. In this benchmark, we consider the vesselness function accounting for the cross section:

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

G. Ranking the Orientation Responses of Path Operators (RORPO) [10]

This last method is the only that does not rely on differential information. By contrast to the previous (local, linear) filters, RORPO is semi-global and non-linear. It builds upon the notion of path opening defined in mathematical morphology [18]. Path opening consists of carrying out a morphological opening based on a set of structuring elements defining a family of paths defined by an adjacency relation. After computing path openings in the 7 main orientations of the 3D space, RORPO point-wise rank their responses. Then it extracts an operator discriminating blobs, planes and line-like structures according to the number of high and low responses at each voxel.

III. DATA

The proposed benchmark focuses on the hepatic vascular network. Image analysis for liver applications mostly involves CT acquisitions. However, the interest in MRI acquisitions is growing, in particular for vascular studies.

Liver datasets including liver images, liver masks and vascular ground-truths are rare as manual segmentation is a tedious and time consuming task requiring experts. In addition, regulations regarding patient’s data protection and privacy have become increasingly strict in Europe. To the best of our knowledge, only two such CT datasets are publicly available and none for MRI.

One of the two CT datasets was recently released for the Medical Segmentation Decathlon¹. However this dataset presents a low axial resolution (5 mm) leading to an insufficient precision with respect to vascular structures and ground-truths are defined as a mixture of tumors and vessels, making vascular-only filtering a difficult evaluation.

¹<http://medicaldecathlon.com>

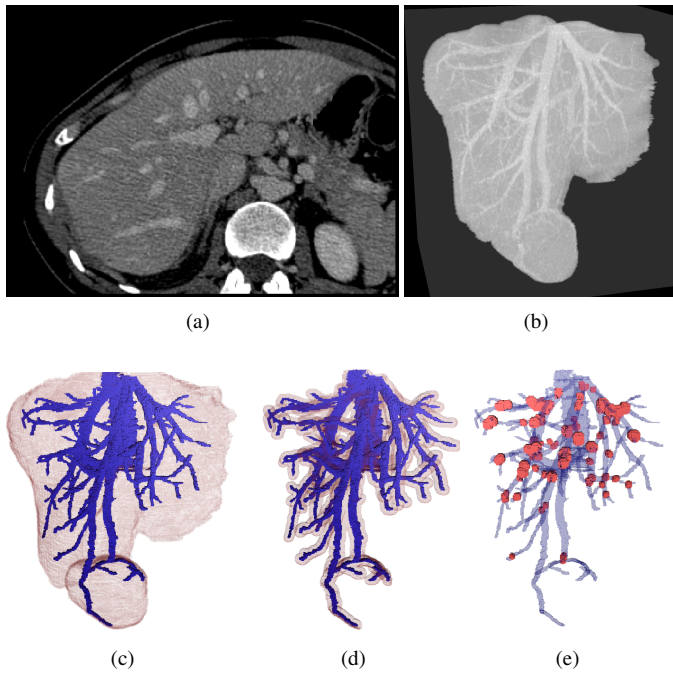


Fig. 1. CT image of the Ircad dataset. (a) Axial slice of the liver. (b) 3D maximum intensity projection (MIP) visualization inside the liver mask. (c–e) Vascular ground-truth (in blue), together with the mask of the liver (c, in red), the mask of the vessel neighborhood (d, in red) and the mask of the vessel bifurcations (e, in red).

Based on these considerations, we present hereafter the two datasets that we chose, namely the Ircad dataset for (real) CT images and the VascuSynth dataset for (simulated) MR Images.

A. Ircad Dataset

The Ircad dataset² contains images from 20 patients (50% men/women; 75% hepatic tumors). The following images are available for each patient: a CT image including the liver (Fig. 1(a,b)) (image size between $512^2 \times 74$ and $512^2 \times 260$ voxels; axial slice resolution between 0.56 mm and 0.87 mm, axial slice thickness between 1.00 mm and 4.00 mm); a mask for all organs in the abdomen, including the liver (Fig. 1(b,c)); and separate ground-truths of hepatic vessels and tumors.

Since the investigated filtering methods require isotropic data to provide relevant vesselness responses, the CT images were preprocessed using B-spline interpolation. This led to a resolution of $1 \times 1 \times 1 \text{ mm}^3$ for all the images, allowing for an homogeneous scale-space handling, and providing a relevant trade-off between axial upsampling and coronal/sagittal downsampling.

The vascular ground-truths were obtained by merging the hepatic vessels and portal vein ground-truths natively provided, and removing parts outside the liver mask (Fig. 1(c–e)).

²<https://www.irca.fr/fr/recherche/3d-ircadb-01-fr>

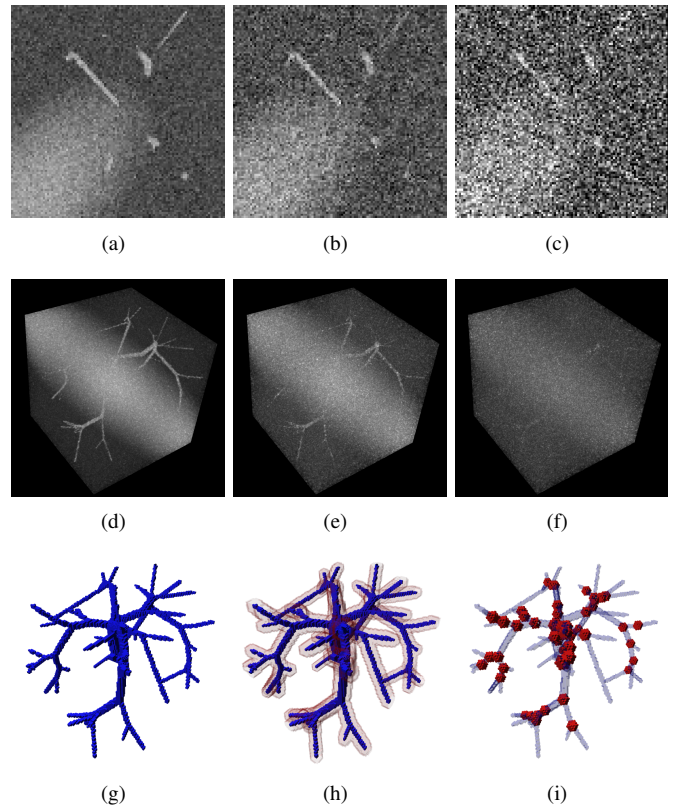


Fig. 2. Simulated images from the VascuSynth dataset. (a–c) Slices with different levels of noise ($\sigma = 5, 10, 20$) and bias effect. (d–f) 3D MIP visualizations for the same levels of noise and artifacts. (g–i) Vascular ground-truth (in blue), together with the mask of the vessel neighborhood (h, in red) and the mask of the vessel bifurcations (i, in red).

B. VascuSynth Dataset

In the absence of MRI hepatic vessel dataset, we turned to a synthetic dataset. We consider the 2013 VascuSynth [19] dataset³ that simulates vascular trees. The original dataset is composed of 120 3D isotropic images. The complexity of the vascular trees vary from 1 to 56 bifurcations. In this work, 30 images with a bifurcation complexity of 31, 41 and 51 were kept from the original dataset. In order to obtain realistic images, compliant with MRI acquisition properties, each of these 30 images was corrupted with three levels of Rician noise ($\sigma = 5, 10$ and 20) and bias effect (Fig. 2(a–f)). The final synthetic dataset thus contains 90 images.

The ground-truths simply correspond to the binary volumes used for simulating the images (Fig. 2(g–i)).

IV. EXPERIMENTAL SETTINGS

A. Regions of Interest

To compare the results from each method, a quantitative analysis is conducted by computing usual metrics (Sec. IV-B) in three different regions of interest: (1) the liver mask (resp. the whole image) for the Ircad dataset (resp. the VascuSynth

³<http://vascusynth.cs.sfu.ca/Data.html>

dataset) to obtain global results; (2) the vessel neighborhood for more detailed results around the objects of interest; and (3) the bifurcation areas, to focus on these complex regions where the vessel enhancement filters sometimes fail. Illustrations of these different regions of interest are provided in Figs. 1(c–e) and 2(h,i).

B. Evaluation Metrics

The goal of vessel enhancement filters is twofold: reducing noise and artifacts in the image whereas enhancing the vessels. It is usual to evaluate these filters by thresholding the result and comparing the best induced binary result to the vessel ground-truth. We follow this strategy and compute 100 thresholdings for each grey-level result. The confusion matrix is computed on each thresholded result to obtain the true positives (tp), true negatives (tn), false positives (fp) and false negatives (fn) with respect to the vessel ground-truth in each region of interest (Sec. IV-A).

Two similarity metrics are then computed based on these values, namely the Dice coefficient:

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

and the Matthews Correlation Coefficient:

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

The Dice coefficient focuses on the overlapping between the ground-truth and the binarized results, whereas the MCC takes into account the whole image (including also the true negatives). We chose the MCC instead of the usual Accuracy since the MCC is specifically designed for handling imbalanced classes (here, sparse vessels vs. large background). Dice score lies in $[0, 1]$ whereas MCC lies in $[-1, 1]$; the closer to 1 the better the results.

In addition to these two metrics, we also consider the ROC (Receiver Operating Characteristic) curve to have a more global overview on the grey-level results. A ROC curve depicts the true positive rate ($tpr = \frac{tp}{tp+fn}$) vs. the false positive rate ($fpr = \frac{fp}{fp+tn}$). In our experiments, a ROC curve is composed of 100 points, one for each binarized result.

C. Optimization Scheme

The different filtering methods depend on two kinds of parameters: the scale-space parameters (e.g. σ) and the intrinsic parameters of the methods (e.g. α , β and c in the Frangi vesselness). In the next section, we present filtering results with optimal parameters. In particular, these parameters are optimized with respect to the MCC.

The first step of the optimization deals with the scale-space parameters. It is carried out independently for each dataset, each method and each mask (except the bifurcation mask). We fix the intrinsic parameters of each Hessian-based method to default values (generally provided by the authors, empirically chosen otherwise). Then, a grid search is performed on the scale-space bounds σ_{\min} and σ_{\max} with a number of scales fixed to 4 within the interval $[\sigma_{\min}, \sigma_{\max}]$, with $\sigma_{\min} \in$

$[0.6, 1.4]$ and $\sigma_{\max} \in [1.6, 3.0]$, and a step of 0.2, under the constraint that $\sigma_{\max} - \sigma_{\min} > 1$. For RORPO (the only non-Hessian-based method), three parameters are optimized: the size s_{\min} of the minimal path, with $s_{\min} \in [10, 60]$ and a step of 10; the (multiplicative) growing factor g of path size with $g \in [1.2, 1.6]$ and a step of 0.2; and the number k of different sizes of path, with $k \in \{3, 4\}$; under the constraint that the only considered paths have a size in $[40, 100]$. In this first step, the optimization is carried out groupwise. In other words, for given filter / dataset / mask, the same scale-space parameters are determined for all the images of the dataset by maximizing the mean MCC score over all these images.

The second step of the optimization deals with the intrinsic parameters of the considered filter. At this stage, the scale-space parameters are known and fixed. For each image of the considered dataset / mask, the parameters of the filter are then optimized, once again by considering MCC as objective function. This step is required only for Sato, Frangi, Jerman, Zhang and OOF. Indeed, Meijering and RORPO only rely on scale-space parameters. It is important to notice that such way of setting the intrinsic parameters corresponds to a “best case” scenario for parameter-dependent filters, since in real applications, a “mean case” scenario is generally considered, i.e. the filter is not parameterized per image, but parameters are learnt from a sample of images, and then applied on the whole dataset.

V. EXPERIMENTS AND RESULTS

A. Implementation and Reproducibility

In addition to the benchmark results presented hereafter, we also provide a common framework with a C++ implementation of all the tested filters. The source code is freely available on the following GITHUB repository: <https://github.com/JonasLamy/LiverVesselness>. Few dependencies are needed (mainly ITK and BOOST) and different scripts permit to reproduce our results. Providing these implementations allows any interested reader to design his/her own way of comparing the different filters, and to complete the proposed study.

In order to ensure the reproducibility of the results without software dependency, we also propose an online demonstration, available via the following link: http://ipol-geometry.loria.fr/~kerautre/ipol_demo/LiverVesselnessIPOLDemo. It allows one to test the seven methods on VasuSynth and Ircad images or any uploaded image. The user may experiment each method and observe their parameter sensitivity by directly visualizing the results online or download them for further analysis.

B. Results

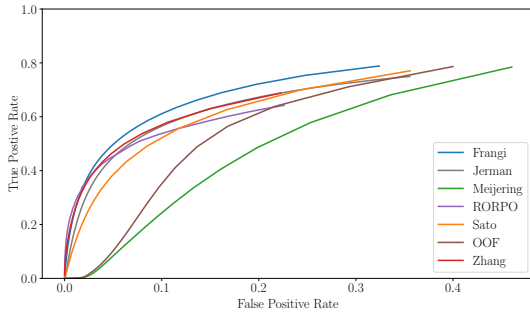
The MCC and Dice values are provided for each of the seven methods and each of the three masks, for the Ircad dataset (Tab. I) and the VasuSynth dataset (Tab. II). Regarding VasuSynth, we provide the results for $\sigma = 10$, which corresponds to a relatively low signal-to-noise ratio. This is justified by the fact that MR images acquired for the liver are

TABLE I
IRCAD DATASET: MCC AND DICE VALUES (MEAN \pm STD DEV) FOR THE 7 FILTERS AND THE 3 MASKS (BEST RESULTS HIGHLIGHTED).

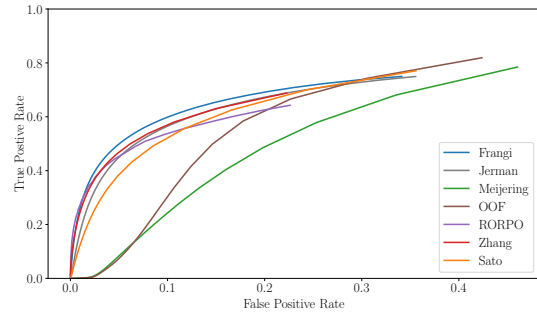
	Liver mask		Vessel neighbourhood		Bifurcations	
	MCC	Dice	MCC	Dice	MCC	Dice
Sato	0.275 \pm 0.066	0.286 \pm 0.073	0.534 \pm 0.061	0.722 \pm 0.072	0.288 \pm 0.057	0.289 \pm 0.068
Frangi	0.356 \pm 0.079	0.367 \pm 0.083	0.560 \pm 0.061	0.719 \pm 0.083	0.453 \pm 0.041	0.525 \pm 0.043
Meijering	0.138 \pm 0.038	0.150 \pm 0.048	0.466 \pm 0.069	0.651 \pm 0.079	0.264 \pm 0.044	0.302 \pm 0.076
Jerman	0.318 \pm 0.081	0.330 \pm 0.082	0.550 \pm 0.057	0.717 \pm 0.069	0.323 \pm 0.046	0.341 \pm 0.073
Zhang	0.346 \pm 0.106	0.344 \pm 0.108	0.503 \pm 0.134	0.685 \pm 0.170	0.308 \pm 0.079	0.348 \pm 0.110
OOF	0.190 \pm 0.041	0.200 \pm 0.047	0.556 \pm 0.058	0.732 \pm 0.059	0.320 \pm 0.046	0.378 \pm 0.059
RORPO	0.384 \pm 0.077	0.375 \pm 0.092	0.456 \pm 0.060	0.631 \pm 0.103	0.272 \pm 0.043	0.290 \pm 0.075

TABLE II
VASCUSYNTH ($\sigma = 10$): MCC AND DICE VALUES (MEAN \pm STD DEV) FOR THE 7 FILTERS AND THE 3 MASKS (BEST RESULTS HIGHLIGHTED).

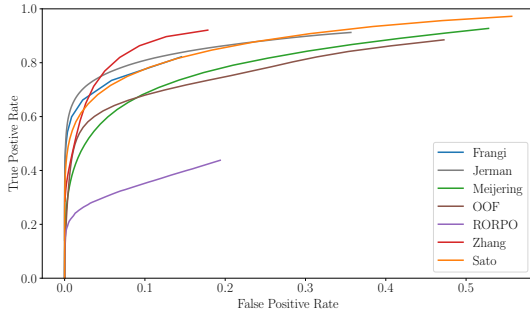
	Whole image		Vessel neighbourhood		Bifurcations	
	MCC	Dice	MCC	Dice	MCC	Dice
Sato	0.544 \pm 0.043	0.497 \pm 0.048	0.553 \pm 0.044	0.727 \pm 0.049	0.637 \pm 0.037	0.654 \pm 0.041
Frangi	0.602 \pm 0.042	0.577 \pm 0.049	0.558 \pm 0.049	0.735 \pm 0.056	0.650 \pm 0.036	0.665 \pm 0.042
Meijering	0.356 \pm 0.040	0.318 \pm 0.056	0.426 \pm 0.031	0.610 \pm 0.054	0.477 \pm 0.029	0.500 \pm 0.040
Jerman	0.612 \pm 0.040	0.589 \pm 0.052	0.545 \pm 0.044	0.725 \pm 0.052	0.629 \pm 0.033	0.656 \pm 0.039
Zhang	0.478 \pm 0.041	0.420 \pm 0.051	0.614 \pm 0.037	0.797 \pm 0.038	0.641 \pm 0.039	0.679 \pm 0.038
OOF	0.343 \pm 0.035	0.342 \pm 0.036	0.549 \pm 0.041	0.737 \pm 0.047	0.630 \pm 0.038	0.654 \pm 0.040
RORPO	0.311 \pm 0.032	0.219 \pm 0.032	0.402 \pm 0.043	0.589 \pm 0.035	0.355 \pm 0.035	0.322 \pm 0.040



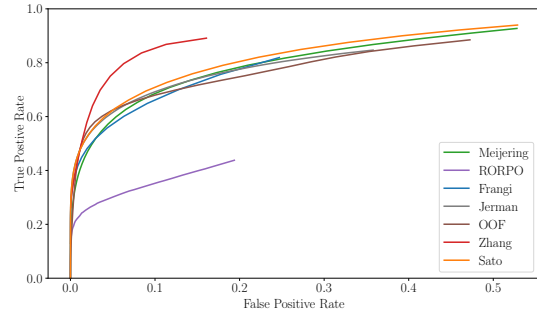
(a) Ircad – Liver mask



(b) Ircad – Vessel neighborhood mask



(c) VasculSynth – Whole image



(d) VasculSynth – Vessel neighborhood

Fig. 3. ROC curves of the seven filters on (a,b) the Ircad dataset and (c,d) the VasculSynth dataset ($\sigma = 10$): (a,c) liver mask/whole image, (b,d) vessel neighborhood mask.

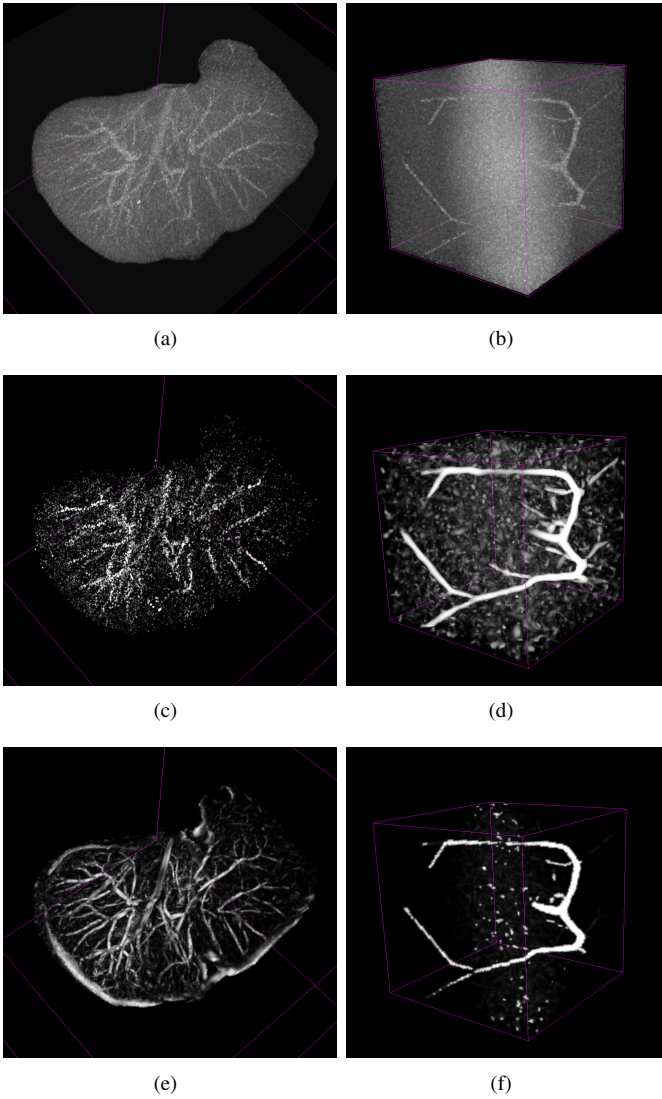


Fig. 4. Illustration of the best methods on the Ircad dataset (a,c,e) and the VascuSynth dataset (b,d,f). (a,b) Initial images (MIP). (c) RORPO and (e) Frangi results on (a) (MIP). (d) Jerman and (f) Zhang results on (b) (MIP).

generally of low quality (in particular compared to CT images or MRIs of other kinds of structures, e.g. the brain).

The ROC curves of the seven methods and two of the three masks (whole image/liver mask; vessel neighborhood mask) are provided for the Ircad and the VascuSynth dataset with $\sigma = 10$ (Fig. 3).

Some examples of filtering results are illustrated in Fig. 4 in order to exemplify the behaviour of the best filters.

VI. DISCUSSION AND CONCLUSION

A. Dice, MCC and ROC Curve Analysis

From these results, one can first observe that the scores of the filters are mostly in the same range. However the behaviour of the filters sometimes differs between the datasets and the studied regions of interest.

Overall, the scores on the Ircad dataset are rather low, confirming vessel enhancement as only a first step for segmentation. All the methods present a high number of false positives (Fig. 3(a)) which often derive from liver border enhancement. Because $\#positives \ll \#negatives$, a point gain of 1% of the FPR results in a much higher number of falsely detected voxels than the number of correctly detected voxels for the same gain in TPR. In this context, the anti-extensivity property of RORPO favors a high precision over a high sensitivity, making it the top filter in MCC and Dice despite a lower ROC curve. Frangi and Zhang yield comparable scores whereas favoring a high sensitivity over a high precision.

The experiments on the vessel neighbourhood confirm this observation. While focusing on an area around the vascular network, most of the false positives detected by the filters are excluded which yields to overall better results. In this case, Frangi and OOF present the best MCC and Dice results closely followed by Sato and Jerman, as they present higher true positive rates. A significant increase of the OOF performance may be noted from the whole liver results as OOF suffers from a large number of false detections corresponding to the liver border.

For VascuSynth ($\sigma = 10$), the global results are better since the images present no structures other than vessels. The best behaviour is obtained by Jerman, for MCC (0.612) and Dice (0.589). The second best filter is Frangi, with scores slightly above Jerman. It is worth mentioning that for a better signal-to-noise ratio ($\sigma = 5$, scores not provided here), Frangi behaves slightly better than Jerman with respect to the two scores.

One may also notice that the MCC and Dice scores are strongly correlated for all the filters in both datasets. This happens whenever $tn \gg tp \simeq fp \simeq fn$ which is generally the case for sparse objects.

With no non-vascular structures to mislead the filters, the metrics increase on vessel neighbourhood is less visible for the VascuSynth dataset. In this area of interest, the best filter is Zhang for MCC (0.614) and Dice (0.797). Zhang yields a true positive rate significantly higher than the other methods as observed in Fig. 3(c,d). However its high number of false positives observed on the whole image results in average performance compared to other methods, potentially because the initial K-mean could not separate correctly the vessels from the intensity artefact. Once again, various other filters present good performances in the vesselness neighbourhood, namely Sato, Frangi, Jerman and OOF. Similar conclusions are obtained for $\sigma = 5$ (scores not provided here). Note that for this mask, the MCC and Dice are no longer correlated, since the tn value is significantly decreased compared to the other tp , fp and fn .

When focusing on the mask of bifurcations, Frangi provides the best results on the Ircad dataset for MCC (0.453) and Dice (0.525). In particular, these scores are significantly higher than those of the other filters. Regarding the VascuSynth dataset, Frangi and Zhang provide the best results for MCC (0.650) and Dice (0.679), respectively. However, Sato, Jerman and OOF also provide good results. Similar conclusions are obtained for

$\sigma = 5$ (scores not provided here). Note that RORPO performs poorly on bifurcations as by design it preserves only tube-like structures.

B. Summary, Limits and Conclusion

Overall, in the most difficult—and realistic—context, i.e. on real images and without any kind of prior knowledge or post-processing, carrying out a good filtering is indeed difficult, and the MCC and Dice scores remain lower than 0.4. In this difficult context, RORPO and Frangi appear as the more robust filters when it comes to the metrics involved. In the other cases, i.e. when one knows the approximate location of the vessels where the filtering has to be carried out and/or when the images do not contain complex artifacts, Frangi, Zhang, Jerman and OOF appear as the most robust filters (with an emphasis on Zhang in the case of synthetic images).

These conclusions need, however, to be balanced by the following facts.

First, concerning the performance of Zhang, it is important to keep in mind that the filter uses an ad hoc classification step for better discriminating the different tissues. This induces in particular the plus-value observed compared to Jerman (which broadly relies on the same vesselness function). In other words, context-aware hypotheses can, of course, contribute to improve the filtering results.

Second, the ground-truth provided in the Ircad dataset are not perfect, since the borders of the vessels are sometimes inaccurate, whereas some smaller vessels are sometimes missed. This may induce a bias in the analysis. In particular, it would be important to carry out experiments on other CT images endowed with improved ground-truth, in order to refine the analysis.

Third, as stated in Sec. IV-C, the optimization scheme was carried out per image, leading to a fine tuning of the intrinsic parameters for certain filters. RORPO and Meijering, which are the only two filters without such parameters, often present degraded performances compared to the other five filters (this is obvious for RORPO in the VasuSynth dataset and for Meijering in the Ircad dataset). Completing the current analysis with a second optimization scheme involving parameters learnt per dataset and no longer per image, would be necessary to study the parameters sensitivity of these five filters, and then their actual robustness, in a more realistic scenario.

These different perspectives constitute some of the next steps of our work with respect to the proposed evaluation framework.

Finally, with regard to segmentation, the following points can be highlighted. One might choose a filter with respect to a trade-off between performance and time-cost. For instance Jerman produces close results to Frangi while being significantly easier to parametrize. One might also choose a filter based on his/her segmentation scheme. Using RORPO, a voxel reconnection step will probably be needed whereas a filtering of the overflowing responses will be needed for Hessian-based filters such as Frangi or Sato.

ACKNOWLEDGMENT

This work was funded by the French *Agence Nationale de la Recherche* (R-Vessel-X, grant ANR-18-CE45-0018).

REFERENCES

- [1] O. Tankyevych, H. Talbot, N. Passat, M. Musacchio, and M. Lagneau, "Angiographic Image Analysis," in *Medical Image Processing*, 2011, pp. 115–144.
- [2] 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*, vol. 13, pp. 819–845, 2009.
- [3] S. Moccia, E. De Momi, S. El Hadji, and L. S. Mattos, "Blood vessel segmentation algorithms — Review of methods, datasets and evaluation metrics," *Comput Meth Prog Bio*, vol. 158, pp. 71–91, 2018.
- [4] Y. Sato, S. Nakajima, H. Atsumi, T. Koller, G. Gerig, S. Yoshida, and R. Kikinis, "3D multi-scale line filter for segmentation and visualization of curvilinear structures in medical images," in *CVRMed-MRCAS*, 1997, pp. 213–222.
- [5] A. F. Frangi, W. J. Niessen, K. L. Vincken, and M. A. Viergever, "Multiscale vessel enhancement filtering," in *MICCAI*, 1998, pp. 130–137.
- [6] E. Meijering, M. Jacob, J.-C. Sarria, P. Steiner, H. Hirling, and M. Unser, "Neurite tracing in fluorescence microscopy images using ridge filtering and graph searching: Principles and validation," in *ISBI*, 2004, pp. 1219–1222.
- [7] T. Jerman, F. Pernus, B. Likar, and Z. Spiclin, "Enhancement of vascular structures in 3D and 2D angiographic images," *IEEE T Med Imaging*, vol. 35, pp. 2107–2118, 2016.
- [8] R. Zhang, Z. Zhou, W. Wu, C.-C. Lin, P.-H. Tsui, and S. Wu, "An improved fuzzy connectedness method for automatic three-dimensional liver vessel segmentation in CT images," *J Healthc Eng*, vol. 2018, pp. 1–18, 2018.
- [9] M. W. K. Law and A. C. S. Chung, "Three dimensional curvilinear structure detection using optimally oriented flux," in *ECCV*, 2008, pp. 368–382.
- [10] O. Merveille, H. Talbot, L. Najman, and N. Passat, "Curvilinear structure analysis by ranking the orientation responses of path operators," *IEEE T Pattern Anal*, vol. 40, pp. 304–317, 2018.
- [11] C. Sazak, C. J. Nelson, and B. Obara, "The multiscale bowler-hat transform for vessel enhancement in 3D biomedical images," in *BMVC*, 2018, p. 96.
- [12] B. Obara, M. Fricker, D. Gavaghan, and V. Grau, "Contrast-independent curvilinear structure detection in biomedical images," *IEEE T Image Process*, vol. 21, pp. 2572–2581, 2012.
- [13] D. Li, L. Zhang, C. Sun, T. Yin, C. Liu, and J. Yang, "Robust retinal image enhancement via dual-tree complex wavelet transform and morphology-based method," *IEEE Access*, vol. 7, pp. 47 303–47 316, 2019.
- [14] W. Yu, B. Fang, Y. Liu, M. Gao, S. Zheng, and Y. Wang, "Liver vessels segmentation based on 3d residual U-NET," in *ICIP*, 2019, pp. 250–254.
- [15] G. Tetteh, V. Efremov, N. D. Forkert, M. Schneider, J. Kirschke, B. Weber, C. Zimmer, M. Piraud, and B. H. Menze, "DeepVesselNet: Vessel segmentation, centerline prediction, and bifurcation detection in 3-D angiographic volumes," *CoRR*, no. abs/1803.09340, 2018.
- [16] T. Lindeberg, *Scale-Space Theory in Computer Vision*. Springer, 1994.
- [17] C. Lorenz, I.-C. Carlsen, T. M. Buzug, C. Fassnacht, and J. Weese, "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.
- [18] H. J. A. M. Heijmans, M. Buckley, and H. Talbot, "Path openings and closings," *J Math Imaging Vis*, vol. 22, pp. 107–119, 2005.
- [19] G. Hamarneh and P. Jassi, "VasucSynth: Simulating vascular trees for generating volumetric image data with ground-truth segmentation and tree analysis," *Comput Med Imag Graph*, vol. 34, pp. 605–616, 2010.