Tubular structures segmentation of pediatric abdominal-visceral ceCT images with renal tumors: assessment, comparison and improvement

Giammarco La Barbera, Laurence Rouet, Haithem Boussaid, Alexis Lubet, Rania Kassir, Sabine Sarnacki, Pietro Gori, Isabelle Bloch

Tubular structures segmentation of pediatric abdominal-visceral ceCT images with renal tumors: assessment, comparison and improvement

Giammarco La Barbera[a,b,*], Laurence Rouet[c], Haithem Boussaid[f,c,**], Alexis Lubet[b], Rani Kassir[b,d], Sabine Sarnacki[b,d], Pietro Gori[a], Isabelle Bloch[e,a,b]

[a]*LTCI, Télécom Paris, Institut Polytechnique de Paris, France*
[b]*IMAG2, Institut Imagine, Université Paris Cité, France*
[c]*Philips Research Paris, Suresnes, France*
[d]*Université Paris Cité, Department of Pediatric Surgery, Hôpital Necker Enfants-Malades, APHP, France*
[e]*Sorbonne Université, CNRS, LIP6, Paris, France*
[f]*Technology Innovation Institute, Abu Dhabi, UAE*

Abstract

Renal tubular structures, such as ureters, arteries and veins, are very important for building a complete digital 3D anatomical model of a patient. However, they can be challenging to segment from ceCT images due to their elongated shape, diameter variation and intra- and inter-patient contrast hetereogenity. This task is even more difficult in pediatric and pathological subjects, due to high inter-subject anatomical variations, potential presence of tumors, small volume of these structures compared to the surrounding, and small available labeled datasets.

Given the limited literature on methods dedicated to children, and in order to find inspirational approaches, a complete assessment of state-of-the-art methods for the segmentation of renal tubular structures on ceCT images on adults is presented. Then, these methods are tested and compared on a private pediatric and pathological dataset of 79 abdominal-visceral ceCT images with arteriovenous phase acquisitions. To the best of our knowledge, both assessment and comparison in this specific case are novel.

Eventually, we also propose a new loss function which leverages for the first time the use of vesselness functions on the predicted segmentation. We show that the combination of this loss function with state-of-the-art methods improves the topological coherence of the segmentated tubular structures [1].

Keywords: Segmentation, CT, Vessels, renal tumors, pediatric, abdominal, visceral

*Corresponding author: E-mail: giammarco.labarbera@telecom-paris.fr
**This work was done while the co-author was with Philips Research Paris and is now affiliated with Technology Innovation Institute, Abu Dhabi, UAE.

1. Introduction

Building a 3D model of renal tumor anatomy based on image segmentation of contrast-enhanced Computed Tomography (ceCT) images is now becoming popular (Fan et al., 2018; Hyde et al., 2019; Porpiglia et al., 2020). In particular, the usefulness of such 3D individual models lies in facilitating pre-operative planning and per-operative guidance for nephron-sparing surgery (NSS, partial removal of the kidney) in patients affected by nephroblastoma, also called Wilms' tumors. An abdomino-visceral anatomical 3D model should include kidneys, vertebrae, ribs, arteries, veins and tumors; ureters are included if the injection makes them visible, and spleen and liver if the tumor is in contact with them. An example is shown in Figure 1.

Figure 1: Example of a partial 3D model of a pediatric patient affected by nephroblastoma with several tumors. The structures represented are kidneys (in brown), arteries (red), veins (blue), ureters (gray), spleen (dark blue), vertebrae and ribs (light gray), tumors (other colors). Left: Anterior view. Right: Posterior view without ribs.

1.1. Medical motivation

In order to build a renal 3D digital twin of a patient that is as complete as possible, the segmentation of tubular structures, such as ureters, arteries and veins of the renal area, should also be considered.

[1]The code is available at: https://github.com/Giammarco07/DeePRAC_project

Automatic segmentation methods reduce the model creation time and the inter-subject variability in comparison to manual segmentation. While such methods are highly developed and perform well for dense structures, such as kidneys and tumors for adults (Marie et al., 2019; Heller et al., 2021, 2022) and for pediatric subjects (La Barbera et al., 2021), this is not true for renal tubular structures. These structures raise several challenges, such as (i) elongated shape (few pixels in a 2D cross section), (ii) intra- and inter-patient contrast heterogeneity, and (iii) intra-scale changes (i.e., diameter variation across the vascular structure). Some authors (Taha et al., 2018; Li et al., 2018; He et al., 2020) have tried to solve these issues on adults ceCT images without great improvements or limited to specific acquisition CT modalities.

Pediatric pathological patients. The difficulties of segmenting such tubular structures increase in pediatric subjects, due to several factors:

- *Inter-anatomy variation.* Pediatric databases include subjects with ages from 1 day to 16 years, thus the size and position of the vessels vary widely among subjects. Moreover, in case of pathology (e.g. renal tumor) the shape and direction of the vessels may also vary (difficult to choose a 2D cross section to work on).

- *Small volumes.* Pediatric tubular structures can be very small compared to the background, namely everything that is not the structures of interest. In the abdominal-visceral region of children, for example, the diameter of the aorta is in [10,30] *mm* range, [8,25] *mm* for cava vein major axis and [3,18] *mm* for its minor axis, while the renal vessels and ureters are in [3,5] *mm* range (Amir-Khalili et al., 2014). Even using patches, i.e. smaller 3D images extracted from the entire volume, these structures will represent a small number of voxels compared to the surrounding tissues and organs.

- *Small available labeled dataset.* Pediatric databases are limited in number of images (Jia et al., 2017) and therefore usual deep learning strategies might fail or might not give good results (Sun et al., 2017). Moreover, such limited datasets usually do not represent the large variability encountered in clinical practice (age, morphology, pathologies...).

1.2. Contributions

This article proposes three main contributions. First, a complete assessment of state-of-the-art methods for the segmentation of renal tubular structures on ceCT images of adults is presented in Section 2. Secondly, a fair comparison of these methods is performed when applied on pathological and pediatric ceCT images

in Section 3. To the best of our knowledge, both assessment and comparison in this specific case are novel. Eventually, the best identified technique is merged with a proposed oversampling method and improved with the use of the proposed tubular structures loss function based on vesselness. This is the first time that an automatic segmentation approach based on Convolutional Neural Network (CNN) for renal tubular structures is presented for *pediatric* and *pathological* patients. The main features of our loss function are: comparing the eigenvalues of the Hessian matrix of the predicted and reference segmentations to force the network to learn the morphology of the target structure, and combining the deep supervision technique with the use of Gaussian filters with various standard deviations to cope with patches including vessels of different diameters. To the best of our knowledge, this is the first time that the calculation of eigenvalues (and eigenvectors) of the Hessian matrix is used on segmentation masks (i.e. predicted and reference segmentation) in a loss function. For this reason, and in order to make all the steps of the vesselness function differentiable, some modifications to this function are applied. A comprehensive study is detailed in Section 4.

In Section 5 our pediatric database of patients affected by renal tumors is presented as well as the implementation of the experiments performed. Then, in Section 6 the results on both the state-of-the-art comparison and the proposed method are presented and discussed. Finally, in Section 7 our conclusion are drawn.

2. Related Work

Most of the studies on vessel segmentation that are applied to 3D contrast-enhanced imaging modalities (for both CT and MRI) on adults are extensively described by Lesage et al. (2009) for non-machine learning-based methods (which we will refer to as "rule-based" for simplicity) and by Moccia et al. (2018) for machine learning methods, in particular deep learning. In these reviews, it appears that a popular approach is the use of second-order derivative information, captured via Hessian-based filters, to characterize the local image geometry. These techniques can be summarized under the name "vesselness filters" (as done by Lamy et al. (2022)) and because of their importance, as well as inspiration for our proposed method, the first part of this section is dedicated to them. It is important to emphasize that in this work we mainly focus on methods specifically applied to renal tubular structures segmentation on ceCT, although some relevant works from other domains are also introduced.

2.1. *Vesselness filters*

Vessels in contrast-enhanced medical images are characterized by hyper-intensity and specific geometric features. Therefore, they can be seen as bright tubular structures on a dark background, and consequently, for a given function that analyzes these properties, the voxels of the vessels will have a higher score, namely *vesselness*. A fair amount of work was dedicated to the proposition of such a function (see Lesage et al. (2009) and Lamy et al. (2022) for reviews), most of them arising from the analysis by Lorenz et al. (1997). This states that a voxel can be considered belonging to a vessel if the Hessian matrix, computed at each voxel, has a small eigenvalue of either sign (insignificant second derivative along one dimension, grey-values do not vary much along the structure) and the other two eigenvalues are large and negative (significant second derivatives across the structure, grey-value increases rapidly from the structure border to the centerline and decreases again to the opposite border). Let (x, y, z) be the space coordinates and f the intensity function of the image. The Hessian matrix H is defined as:

$$H(f) = \begin{bmatrix} h_{xx} & h_{xy} & h_{xz} \\ h_{yx} & h_{yy} & h_{yz} \\ h_{zx} & h_{zy} & h_{zz} \end{bmatrix} = \begin{bmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial xy} & \frac{\partial^2 f}{\partial xz} \\ \frac{\partial^2 f}{\partial yx} & \frac{\partial^2 f}{\partial y^2} & \frac{\partial^2 f}{\partial yz} \\ \frac{\partial^2 f}{\partial zx} & \frac{\partial^2 f}{\partial zy} & \frac{\partial^2 f}{\partial z^2} \end{bmatrix} \tag{1}$$

and is calculated at each voxel. Let W_1, W_2 and W_3 be the three normalized eigenvectors of $H(f)$, associated with the eigenvalues λ_1, λ_2 and λ_3, respectively, with $|\lambda_1| < |\lambda_2| < |\lambda_3|$. The vesselness is respected if:

$$\begin{aligned} |\lambda_1| &\approx 0 \text{ (namely small)} \\ \lambda_2 &\ll 0 \\ \lambda_3 &\ll 0 \end{aligned} \tag{2}$$

In order to deal with the non-continuity of the digital medical images and with the high sensitivity to noise of second order derivatives, the Hessian matrix is computed on the image convolved with a Gaussian kernel g_σ with standard deviation σ. Moreover, in presence of a vessel, W_1 associated to λ_1 corresponds to the direction of the putative vessel, while W_2 and W_3 form a basis of the vessel cross section where $|\lambda_2|$ and $|\lambda_3|$ are correlated to the sizes of the cross section. The most popular function to score the so-called vesselness is the one proposed by Frangi et al. (1998). Aiming to build a method suited for medical images, these authors developed a filter F based on three measures and three parameters:

$$F = (1 - \exp(\frac{-R_a^2}{2\alpha^2}))\exp(\frac{-R_b^2}{2\beta^2})(1 - \exp(\frac{-S^2}{2\gamma^2})) \tag{3}$$

if $\lambda_2, \lambda_3 < 0$ and $F = 0$ otherwise, with:

$$\begin{aligned} R_b &= |\lambda_1| / \sqrt{|\lambda_2 \lambda_3|} \\ R_a &= |\lambda_2| / |\lambda_3| \\ S &= \sqrt{\lambda_1^2 + \lambda_2^2 + \lambda_3^2} \end{aligned} \tag{4}$$

in which R_b discriminates blobs, R_a distinguishes plate and line structures, S measures the norm of the Hessian matrix to avoid enhancing low contrast structures, and α, β and γ control the importance of each measure. Moreover, to cope with different vessel sizes, the vesselness measure is analyzed at different standard deviations and then the maximum response, i.e. the maximum F, is kept. As stated by Lesage et al. (2009), Hessian-based filters may suffer from sensitivity to local deformations (bifurcations, thrombus or flattened vessels). Moreover, a parameter search is required for both the scale-space parameters (e.g. σ of the Gaussian g_σ) and the intrinsic parameters of the methods (e.g. α, β and γ in Frangi's vesselness). In addition, the extraction of large vessels requires large standard deviations of g_σ that can result in perturbation of the response due to other bright structures in the immediate vicinity of the target vessel. These limitations are confirmed by the study presented by Lamy et al. (2022), where acceptable segmentation results using vesselness filters are obtained only for synthetic images that present no structures other than vessels. However, they can be useful as a first step for segmenting ceCT images, in particular for intra-organ vessels where the neighborhood is more homogeneous (see further details in Section 2.2). Some interesting alternatives to multiscale Hessian-based filters have been developed, such as the optimally oriented flux (OOF) of Law and Chung (2008), the ranking of the orientation responses of path operators (RORPO) of Merveille et al. (2018) and the work of Bauer and Bischof (2008), where the Frangi's vesselness is applied on the local Jacobian of the gradient vector flow (GVF) field. However, OOF and RORPO are also analyzed by Lamy et al. (2022) and no improvement in performance over Hessian-based filters was observed, while the method by Bauer and Bischof (2008) has a high computational cost, also in GPU, due to GVF diffusion calculation.

In the literature, besides segmentation purpose, vesselness functions have also been used as a cost function in lung vessel-tree registration algorithm for ceCT images (Cao et al., 2010; Chen et al., 2012; Du et al., 2013). Small vessels give almost no contribution to intensity-based similarity, thus to further improve the registration accuracy, the sum of squared vesselness measure difference (SSVMD) is employed as geometric-feature similarity. In such applications, the lung region-of-interest (ROI) in ceCTs perfectly fits the idea of bright tubular structures in dark background (i.e. air is black in ceCTs), overcoming some weakness of Hessian-based filters previously exposed. With the same goal, very recently Wang et al. (2022)

proposed the use of vesselness as a loss function in a CNN for the first time. In particular they use a normalized vesselness Jerman's filter (Jerman et al., 2016), that is more robust to weakly constrasted regions than Frangi's one.

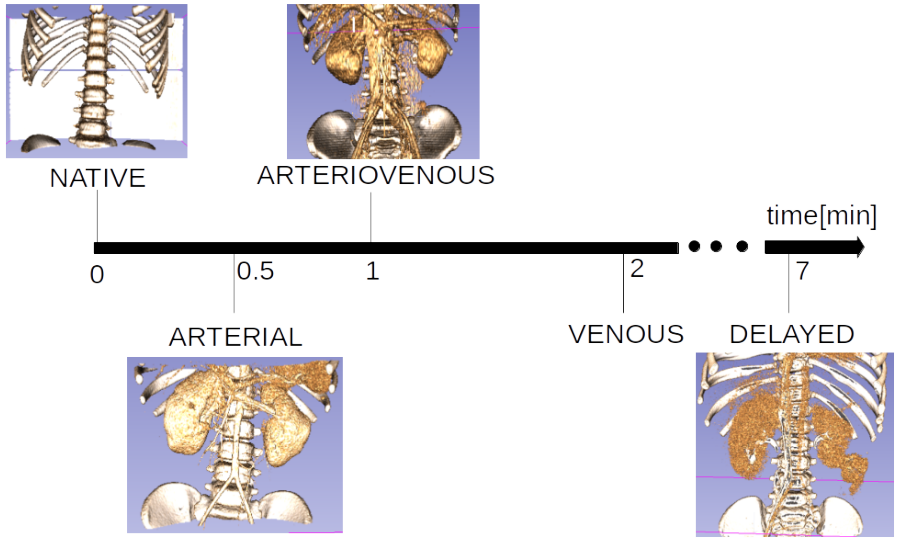


Figure 2: Exemplary timestamp of each phase acquisition in relation to the contrast agent injection. Adapted from (Heryan et al., 2018) using volume-rendered images from Necker hospital dataset (see Section 5.1).

2.2. Rule-based methods for tubular structures segmentation

Works that do not use neural networks (Lesage et al., 2009) usually rely on the fact that vessels are contrasted with respect to other structures, i.e. a bright (resp. dark) structure in a dark (resp. bright) background, as the previously presented vesselness filters. On abdominal-visceral ceCT images, this case is restricted to early arterial phase acquisition for having high contrasted arteries, or to late delayed phase acquisition for having high contrasted excretory pathways (i.e. ureters). Examples of each phase acquisition time in relation to contrast agent injection in CT are displayed in Figure 2.

Nevertheless, due to the very rapid occurrence of the arterial phase (in particular on children) or due to medical choice in order to have both arteries and veins with higher intensity, very often ceCT images are acquired in a late arterial phase, also known as arteriovenous phase (Heryan et al., 2018). Here, both types of vessels are contrasted but the presence of less contrast medium in each one results in lower intensity (and thus less difference with other structures) as well as greater heterogeneity (which may be accentuated by the

presence of tumor or thrombus). Some examples are shown in Figure 3.

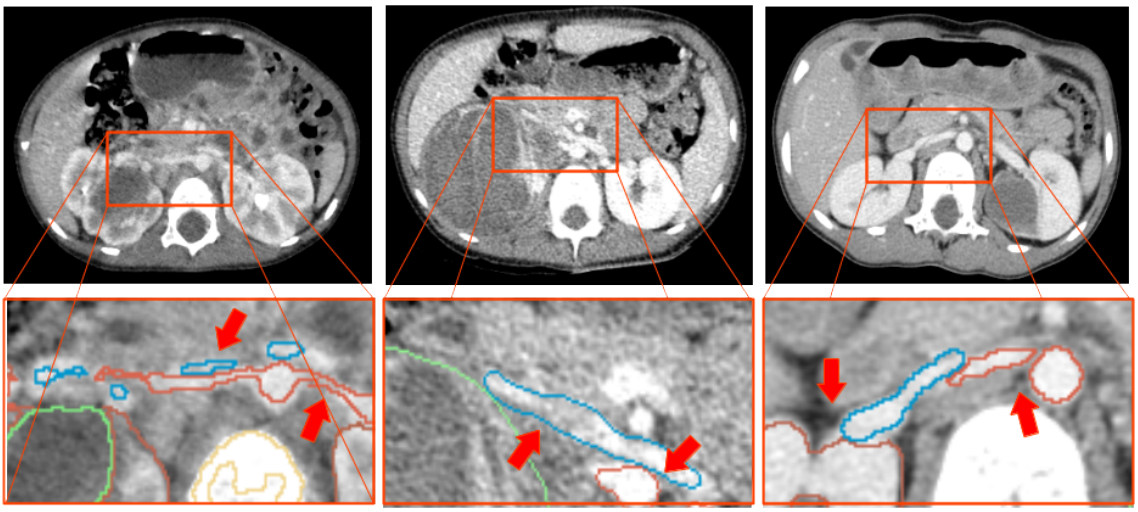


Figure 3: Some examples of arteriovenous phase ceCT images from the pediatric and pathological dataset of Necker hospital. Left figure: low contrast in both veins (left arrow) and arteries (right arrow). Central figure: heterogeneity zone in veins (left arrow) caused by the tumor, and high intensity in arteries (right arrow). Right figure: kidney (left arrow) and blood vessels (right arrow) have very similar intensities, difficult to distinguish by non-medical experts. Color code for contours: kidneys in brown, tumor in green, arteries in red, veins in blue.

To tackle this problem, Bugajska et al. (2015) proposed a semi-automatic approach based on three steps for the segmentation of renal tubular structures in ceCT composed of: (i) a first binarization and erosion, in which the threshold is derived from a ROI (derived by points defined by the user); (ii) a subsequent Locally Adaptive Region Growing (LARG) technique to deal with the lack of homogeneity of voxel intensity values due to an improper contrast propagation; (iii) a final level set method which uses the result of LARG as initialization. Results on 10 patients show that the first two steps are already sufficient to obtain good results for the segmentation of the main and larger abdominal-visceral arteries, while only after the third step the small renal vessels reach 80% of the Dice score. However, the method still works better on ceCT with *non-late* arterial phase, otherwise there is a high demand of user-interaction. The method does not allow for the segmentation of veins, in which the contrast is still too low. A similar approach was used by Heryan et al. (2018), in which the authors point out the difficulties of dealing with a combined arteriovenous phase in ceCT images, and for this reason the main threshold is determined by analyzing different points defined by the user. Then they focus on intra-renal vessels segmentation using a LARG on a manually-delineated

ROI around each kidney in which a vessel enhancement through multiscale vesselness Frangi's filter (Frangi et al., 1998) was first applied. Among unsupervised methods not based on deep learning, it is also important to mention Tensor-cut by Wang et al. (2019b), a novel tensor-based graph-cut method based on the local neighboring Markov random field model. The limitations of these methods lie in the user interaction and in the setting of different initialization parameters, required according to the input image and to the specific tubular structure to be segmented.

2.3. Deep learning-based methods for tubular structures segmentation

Moving on to methods based on deep learning for tubular structures segmentation, we first focus on the method of Virzi et al. (2018) which is not applied to renal ceCT images, but still dedicated to tubular structures segmentation. The authors proposed a semi-automatic patch-based deep learning approach to segment pelvic vessels in 3D MR images of pediatric patients. To consider only relevant patches, the skeleton of the vascular tree is obtained by combining user-selected landmarks and shape-appearance information, and then patches are extracted along this skeleton. This method also allows dealing with small structures, since it can use smaller patches without losing information, and therefore at the same time also decreasing the training time and memory required by the CNN. The method of Virzi et al. (2018) works well for pelvic vessels whose branching is fairly simple and constant, although it always requires some user's interaction. In the case of renal structures, while the ureters tree is quite simple, this is not true for arteries and veins trees, as we can see in the example in Figure 4.

With the same purpose of working with only small and relevant patches, while overcoming the problem given by the skeleton creation via user-selected landmarks, Dang et al. (2022) proposed the use of a PNet-based (Wang et al., 2017) patch classifier called Vessel-CAPTCHA. The difficulty of this method lies precisely in the training of the classifier, which requires a large number of patches to achieve good performance. For these reasons, the authors decided to work in 2D, losing the volumetric information, and to create a user-friendly annotation system on a large unlabeled training set. Although the method seems to work better than the 2D state-of-the-art networks presented in the paper, its application for pediatric and pathological datasets is limited by the small number of available images, even unlabeled ones. Moreover, user interaction can be tedious and difficult, especially for subjects with abnormalities, such as tumors/lesions. For this reason, many authors have proposed fully-automated algorithms. In the following we focus on these methods.

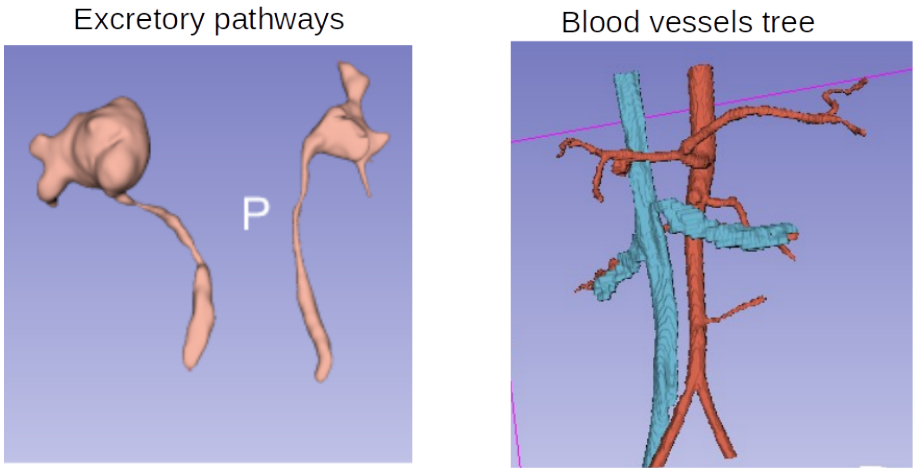


Figure 4: Ureters on the left in pink and renal blood vessels tree on the right with arteries in red and veins in blue. While the ureters tree is quite simple, the arteries and veins trees are very irregular. Building a skeleton of a vascular tree from user-selected landmarks is very complicated.

Fully-automated algorithms. A first fully-automated algorithm is Kid-Net of Taha et al. (2018). It is a method to segment renal arteries, veins and collecting system via CNN on ceCT. The authors point out the importance of using 3D patches instead of 2D slices, because the latter do not have enough information. Moreover, instead of training for individual classes of interest independently, they train the network to detect the three classes together. The first advantage of this approach relies on the use of a single network decision per voxel with a Softmax activation function, where the classes are considered mutually exclusive (useful for facilitating the choice in common edge voxels among multiple structures). Secondly, this approach aligns with Kaiser et al. (2017)'s recommendation that learning tasks with less data benefit largely from joint training on other tasks. Kid-Net is a 3D U-Net and its major contributions rely on the use of a so-called deep supervision, a patch selection method that handles unbalanced data through the use of random sampling and a dynamic weighting based on volume to background ratio. They use cross-entropy (CE) as loss function. Results on a private dataset of arterial phase ceCTs are above 85% of Dice score only for large arteries, degrading to around 60% for small ones, veins and ureters.

For the sake of completeness, the deep supervision is a technique proposed by Dou et al. (2016) in which additional outputs are added in the decoder to all but the lowest two resolutions. This allows the gradient to be further injected into the architecture, facilitating the training of all layers and reducing the vanishing gradient problem (common with restricted datasets). The final loss function L is the sum of the loss functions

L_k computed at different levels k of resolution:

$$L = \sum_{k=0}^{Q} w_k \cdot L_k \tag{5}$$

with:

$$w_k = \frac{1}{2^k} \frac{1}{\sum_{q=0}^{Q} \frac{1}{2^q}} \tag{6}$$

where $Q + 1$ is the number of resolution levels taken into account, starting from 0 as the last output level (e.g. with input image size).

To tackle the difficulties of fine vessels segmentation and of having a small dataset, He et al. (2020) proposed DPA-HRA-DenseBiasNet. The DenseBiasNet takes the standard 3D U-Net as the basic structure, but it compresses and transmits all feature maps in each layer to every forward layer. Moreover, the authors proposed what they call "the deep prior anatomy (DPA) strategy": an autoencoder (AE) is trained in an unsupervised manner with numerous unlabeled data in which noise is added and the AE is optimized to reconstruct the original image. Once the AE is trained, the features are embedded in the supervised model to guide it. Finally, the authors proposed a hard region adaptation (HRA) loss function that samples the loss dynamically according to the segmentation quality of each pixel. They achieved for the first time fine renal artery segmentation with Dice score equal to 88%, showing better results than Kid-Net and 3D U-Net trained and tested on the same database. It is important to underline that with the only use of the Dense 3D U-Net a Dice Score of 86% is already achieved, and the use of HRA and then HRA+DPA both add 1%. We would like to emphasize that no reason is shown why the use of the autoencoder for DPA leads to a priori anatomy modeling. The weakness of this method lies in the computing process that requires a lot of memory and time, due to the dense connections of the network and the additional parameters of DPA strategy. Furthermore, this work has been tested only on arteries, experiments with veins and ureters still need to be done.

A faster method is presented by Li et al. (2018), in which the authors try to segment not only all tubular structures but also renal parenchyma and tumors with a single network. They used a Residual U-Net with a multi-scale weighted cross-entropy loss function that gives greater importance to foreground voxels, edges, and complicated and small structures. Tests confirmed the analysis by other authors with high Dice score for arteries and parenchyma, but a higher difficulty in segmenting veins, ureters and tumors.

Focusing now on methods not developed for the renal area but still applied on ceCT images, special attention should be paid to the well-known nnU-Net (Isensee et al., 2021). This network was tested on 49 segmentation tasks, reaching state-of-art results in 29 of them, otherwise achieving performances on par to the top

leader-board entries, as in the case of aorta, cava vein and celiac artery in abdominal-visceral ceCTs (Isensee et al., 2021). Among the various proposed contributions, similar to the Kid-Net approach (Taha et al., 2018), the authors also proposed the training of multiple "foreground" classes, the use of deep supervision and an oversampling technique. However this is different from Kid-Net's one, here it ensures that one of the classes of interest is contained in at least 33.3% of patches in every mini-batch. The other 66.7% of patches are randomly selected from the entire training set. Another difference with Kid-Net lies in the final loss function, which is a combination of Soft Dice loss and cross-entropy.

Other methods applied on ceCT images exploit distance map to improve performance. Distance maps are generated by computing the distance transform on the segmentation masks. Some authors (Moccia et al., 2018) proposed to use it to leverage the differences in diameter of the vessels. Wang et al. (2019a) used the inverse of the normalized distance map to weight the voxel of the reference segmentation in the Dice loss calculation. Their proposed loss function is called Radial Distance loss, $RDloss$, and successfully improves lung tree segmentation performances of both smaller vessels and voxels at the boundary of thicker ones. A similar approach was proposed by Caliva et al. (2019) to segment tubular bones, such as femur and tibia, but the inverse distance map was applied using cross-entropy calculation. However, in addition to the difficulty in calculating the distance map using GPU, such methods do not deal well with the high inter-anatomy variation present in pediatric databases.

The idea of exploiting the geometry of the vessels was also used by Wang et al. (2020), who proposed to include the tubular geometry information directly in the training of a CNN. The distance map is created for every manual segmentation, and used as reference for a loss function with the second output channel of the network (where the first output channel is the segmentation mask). During inference, the segmentation mask is refined by leveraging the shape prior reconstructed from the distance map (named Geometry-Aware Refinement, GAR). One important proposition is that the distance map D is quantized on values from 0 to K (calculated as the maximum possible distance) and the cross-entropy is calculated between the discrete distance map D and the probability that a voxel belongs the k-th class (softmax of the second output with $K+1$ channels). In this way, Wang et al. (2020) formulate the distance prediction problem as a classification problem. The authors explain that the use of distance map without quantization can make the training difficult because outliers can cause large errors and lead to unstable predictions (resulting in difficulty for the network to converge). Experiments showed better performances compared to the previous state-of-the-art methods on aorta, cava vein and hepatic vessels segmentation in ceCT scans, automatically providing also the distance map for a geometrical measurement of the tubular structures, at the expense of high computational

time and memory. The limitation of this method also lies in the possibility of segmenting one structure per network.

Another method worth mentioning is TopNet by Keshwani et al. (2020), where the authors propose one of the few approaches in the literature for multi-label vascular tree segmentation via deep learning. They present a 3D network architecture that solves three tasks simultaneously using three decoders, which are respectively trained by optimizing the Dice score of the segmented vessels, a centerness score based on the distance of each voxel in the vessel mask to the centerline of the vessel, and a similarity of the learned features of voxels belonging to the same vessel. The results of the segmentation of the whole portal and hepatic vein trees inside the liver from abdominal ceCTs are considerably better than classical segmentation techniques both in quantitative measures (e.g. Dice score) and qualitative results (e.g. topology), even if they lack comparisons with some recent methods, previously detailed in this section. The time and memory required by this approach are acceptable, however an additional network which outputs the location of portal and hepatic vein source (or a user's interaction) is required as well as a post-processing where a Dijkstra's multi-source shortest path tree algorithm is applied (which in addition is very sensitive to misdetection of vascular center-voxels).

Recently some authors are beginning to combine CNN with Graph Neural Network (GNN) with satisfactory results for tasks such as lung vessel-tree semantic segmentation in ceCT images (Garcia-Uceda Juarez et al., 2019; Tan et al., 2021) and head and neck artery semantic segmentation in angiographic ceCT images (Yao et al., 2021). The CNN gives a rough estimate of branch endpoint landmarks and binary branch segmentation locations, and then the GNN refines these rough estimations to produce the final semantic segmentation. However, their utility lies in being able to do semantic segmentation from binary labels, i.e. segmenting lung vessel-tree branches to fine categories of n segments while training the network with the entire lung vessel-tree, which is not our purpose. Moreover, CNN-GNN networks have a very high computational cost and require a high amount of input images to converge.

The same problems affect the transposition for 3D images of methods that focus on the simultaneous optimization of segmentation accuracy and topology coherence developed for 2D imaging of neurons and retinal vessels (Moccia et al., 2018). Some of the most recent methods propose for example the use of a loss function to force the predicted segmentation to have the same Betti number (number of connected components and handles) as the reference one (Hu et al., 2019), or the use of a VAE that learns recurrent patterns to refine the output of the segmentation network (Araujo et al., 2019). However, these datasets have big differences with ours, such as: a high amount of images in the database; the vessels are spread

over the entire image (with very few patches without the structures of interest) with a lot of connected components and handles in recurring patterns; the vessels are always parallel to the acquisition plane; and a single structure has to be segmented. This makes it possible to work with a high number of small 2D patches (instead of small number of large 3D ones) in order to efficiently extract topological information, usually computationally complex, memory- and time-demanding.

2.4. Assessment discussion

According to the analyzed literature, to date, there is no method capable of segmenting all renal tubular structures in arteriovenous ceCT images with high performance. In addition, none of the presented approaches were applied to pediatric ceCT images, which present further difficulties. However, various techniques presented in this section, such as the use of a single network to detect multiple adjacent structures, the deep supervision, and the oversampling and patch selection method, may allow overcoming some limitations related to the segmentation of tubular structures, including the small available labeled dataset, the intra-scale changes, and the small volume to background ratio. The main limitation still seems to be related to the loss functions used. In fact, voxel-wise functions are not sufficient for tubular structures segmentation, and the combined use of distance map does not seem to be effective, inducing instead a redundancy of information for the network. The use of topological measurements seems to lead to good results, however the methods proposed so far are suitable only for specific cases, that do not correspond to the problem addressed here.

Taking inspiration from non-machine learning methods, and in particular from vesselness filters, we propose instead to combine the voxel-wise loss functions with a new function based on the eigenvalues of the Hessian matrix and on the vesselness function itself. Our idea is to exploit both the *a priori* structural morphological knowledge of tubular structures and the information about their neighborhood.

3. Methods selected for comparison

In this section, we first present the approaches for selecting and implementing the methods we will compare. We focus on deep learning-based methods due to the difficulties of applying rule-based methods related to heterogeneity of image intensity in arteriovenous ceCTs, as shown in Section 2.2.

3.1. Methods selection

We decided to compare some of the methods presented in Section 2.3, summarized in Table 1. Our selection criteria are:

(i) fully-automated algorithm,

(ii) presence of a distinctive technique,

(iii) code available or easily reproducible.

The first criterion was chosen as a result of the problems set out for semi-automated algorithms in Section 2.3, related to the high-demand of user-interaction and the restricted database. For the second criterion, we have identified four distinctive techniques: the optimization of topology coherence, the use of the distance transform as a second loss term, the dense connection method useful for segmenting fine structures, and the deep supervision for the loss calculation. Among the methods based on these techniques, almost none has available online code but some are easily reproducible, based on details provided by the authors. We selected TopNet of Keshwani et al. (2020) for topology optimization (noted M1), the method proposed by Wang et al. (2020) for distance transform (M2), the method of He et al. (2020) for the dense connections (M4), and Kid-Net of Taha et al. (2018) for the use of deep supervision (M5). A variant of the Deep Distance Transform method presented by Ma et al. (2020), whose code is available online, has also been analyzed (M3). Lastly we selected the nnU-Net (Isensee et al., 2021), as it is a high performing method in general for the segmentation of medical images with its deep supervision method (such as Kid-Net) optimizing CE combined with Dice score (M6). Moreover, nnU-Net code is available online for the purpose of benchmarking, in contrast to the similar Kid-Net algorithm.

Table 1: 3D fully-automated supervised methods to compare.

N°	Method	Backbone	Outputs	Loss functions
M1	TopNet (Keshwani et al., 2020)	U-Net (depth 5)	1: Segmentation 2: CenternessScoreMap 3: Topological Distance	1: Dice 2: Mean of SmoothL1 3: TopLoss
M2	Deep Distance Transform (Wang et al., 2020)	U-Net (depth 5)	1: Segmentation 2: DistanceMap One-Hot Encoder	1: SoftmaxCE 2: SoftmaxCE
M3	Deep Distance Transform (Wang et al., 2020; Ma et al., 2020)	U-Net (depth 5)	1: Segmentation 2: DistanceMap	1: SoftmaxCE 2: Mean of L1
M4	DenseBiasedU-Net (He et al., 2020)	DenseU-Net (depth 4)	1: Segmentation	1: SoftmaxCE+Dice
M5	Kid-Net (Taha et al., 2018)	U-Net (depth 5)	1: Segmentation	1: SoftmaxCE w/ deep supervision
M6	nnU-Net (Isensee et al., 2021)	U-Net (depth 5)	1: Segmentation	1: SoftmaxCE+Dice w/ deep supervision

3.2. Methods implementation

We implemented the selected methods (summarized in Table 1) as follows:

M1. The code of TopNet by Keshwani et al. (2020) is not publicly available. We implemented it from scratch, thanks to the kind help of the authors and to the supplementary material of the original paper. To obtain the full segmentation, we assign a label (arteries or veins) to each voxel of the vessel mask based on the label of nearest "center voxel" (calculated with the Dijkstra's multi-source shortest path tree algorithm). The code would need an additional network for locating the sources for the application of the Dijkstra's algorithm. However, due to the few details provided in the original article (Keshwani et al., 2020) on this part, we decided to proceed as if this localization network was perfect, i.e. using sources points pre-determined by us. In this way, we only evaluate the ability of the TopNet method to provide a good segmentation and a correct topology, which is indeed our final goal.

M2. We implemented this code starting from the available code of M3 which can be seen as a variant of M2. To be able to apply the idea of Deep Distance Transform method of Wang et al. (2020) to formulate the distance prediction problem as a classification problem (as explained in Section 2.3) to two structures (i.e. arteries and vein), we propose that the discrete distance map of veins is added to the discrete distance map of arteries. We add K_A (maximum possible distance of the distance map of arteries) to all values of the distance map of veins where it is different from 0 (the distance map is 0 where the reference segmentation is 0 so this idea cannot produce errors). In this way the final distance map D_{map} values range from 0 to $K_A + K_V$, where K_V is the maximum possible distance of the distance map of veins. By doing so, after quantization, the final distance map D_{map} as well as the second output will have $K_A + K_V + 1$ channels. We coded also the Geometry-Aware Refinement (GAR) of Wang et al. (2020), in order to refine the segmentation output with the quantized distance map output.

M3. This is a variant of the Deep Distance Transform method of Wang et al. (2020) (M2) that was implemented by Ma et al. (2020): here the second channel, i.e. the distance map, is not one-hot encoded. Ma et al. (2020) used an L1 norm as loss function between the reference distance map D_{map} and the second output of the network (Conv3D 1×1×1 with no activation function). This implementation makes it easier to use the distance map for multiple structures but does not respect the idea of the original paper of Wang et al. (2020) (M1). The code is available online: `https://github.com/JunMa11/SegWithDistMap`.

M4. We implemented the DenseBiasedU-Net of He et al. (2020) as in the original paper from scratch. Each dense biased connection compresses, via a convolutional layer, the feature maps in each layer of the U-Net to only four feature maps. Then, these feature maps are transmitted and concatenated to every forward layer. The reduction is done in order to reduce feature redundancy while keeping the integrity of information flow and gradient flow, allowing also fusing multi-scale features. In order to fit the network in a 16 GB GPU, the 3D U-Net has depth of one layer less than the other methods. The HRA and DPA techniques were not implemented due to their low contribution in improving performance (see Section 2.3) and the limited database at our disposal (see Section 5.1).

M5. We implemented this code starting from the avaiable code of M6, using only CE in the loss function and with the different oversampling method presented in Section 2.3.

M6. The original implementation is available online for nnU-Net of Isensee et al. (2021): `https://github.com/MIC-DKFZ/nnUNet`.

4. Proposed Tubular structures Loss Function

In this section, we propose a new loss function based on the vesselness.

4.1. Motivation

The use of Dice score and cross-entropy revealed to be not enough to evaluate the segmentation performance on fine tubular structures. Both are very sensitive to small structures, i.e. changing a few voxels can change the score significantly, in particular the Dice score. This can also affect the training process with patches: we can have high gradient even when the number of wrong pixels is small, and additionally we can have fluctuations due to very different batches. Furthermore, when large and small blood vessels are present in the same patch (e.g. aorta or cava vein and very tiny renal vessels), we will have a good Dice score but the algorithm only segments the aorta or cava vein, while it completely misses the small renal vessels which represent a lower percentage of the foreground voxels in the patch. Eventually, due to the heterogeneity of the vessels, there is a strong uncertainty about some vessels segmentation which may result in prediction of vessels with interruptions.

However, voxel-wise information is necessary to perform voxel classification, such as segmentation. Moreover, for medical image segmentation, cross-entropy proved not to be enough to reach high performance (Isensee et al., 2021), due to the extreme scarcity of foreground voxels in a patch, that will force the

network to have a strong bias to the background (Zhao et al., 2020). The use of the Dice score in the loss function tackles this problem, carrying with it the limitations outlined earlier.

In order to introduce information that is exempt from the number of pixels of the reference segmentation, and that also is not voxel-wise but takes the neighborhood into account, we propose to leverage the use of the eigenvalues of the Hessian matrix and the vesselness function, described in Section 2.1, as a loss function for the segmentation masks. In fact, on the one hand, the eigenvalues allow us to verify that the structural morphology of the predicted segmentation is similar to that of target structure. On the other hand, the vesselness allows us to enhance the segmentation of elongated structures without interruptions. Finally, the different sizes of vessels in a patch can be taken into account by using such loss functions in a multi-scale manner via deep supervision.

Due to the different images, namely the segmentation masks, to which these functions are applied, a new formulation of the steps for calculating eigenvalues and the resulting vesselness score is presented in the next section.

4.2. Formulation

As presented in Sections 2.1 and 2.2, the application of vesselness functions on abdominal ceCT images with arteriovenous phase leads to unsatisfactory results. Taking inspiration from the use of vesselness for registration presented in Section 2.1, we propose for the first time to translate the use of vesselness as loss function for segmentation purpose. Therefore, our proposed vesselness function is not applied on the abdominal ceCT input image but on its segmentation mask and on the predicted one. These images in fact exhibit the sought-after characteristics for a satisfactory vesselness application, similar to those of the lung ROI (bright tubular structures in dark background) used by Cao et al. (2010) and Wang et al. (2022).

One might argue that using such a function on the product between input and segmentation mask would be more appropriate, however; the heterogeneity of ceCT images, particularly of the pediatric ones, makes it complicated to find vesselness parameters that are appropriate for all the structures present in the patch.

Furthermore, since the probability map at the output of the network is not a binary object, an approach directly adapted to binary objects (e.g. moments comparison) would not be appropriate.

Our idea stems from the analysis of probability maps at the network output of a classic 3D U-Net: in these maps, tubular structures were found, but because of the problems exposed in the previous paragraph, a low probability to the correct class was assigned to contour voxels or to the ones belonging to finer portions of the vessel. We noticed that the use of Frangi's vesselness increases the probability assigned to the voxels

of each class that most respect the vesselness. Figure 5 illustrates this observation.

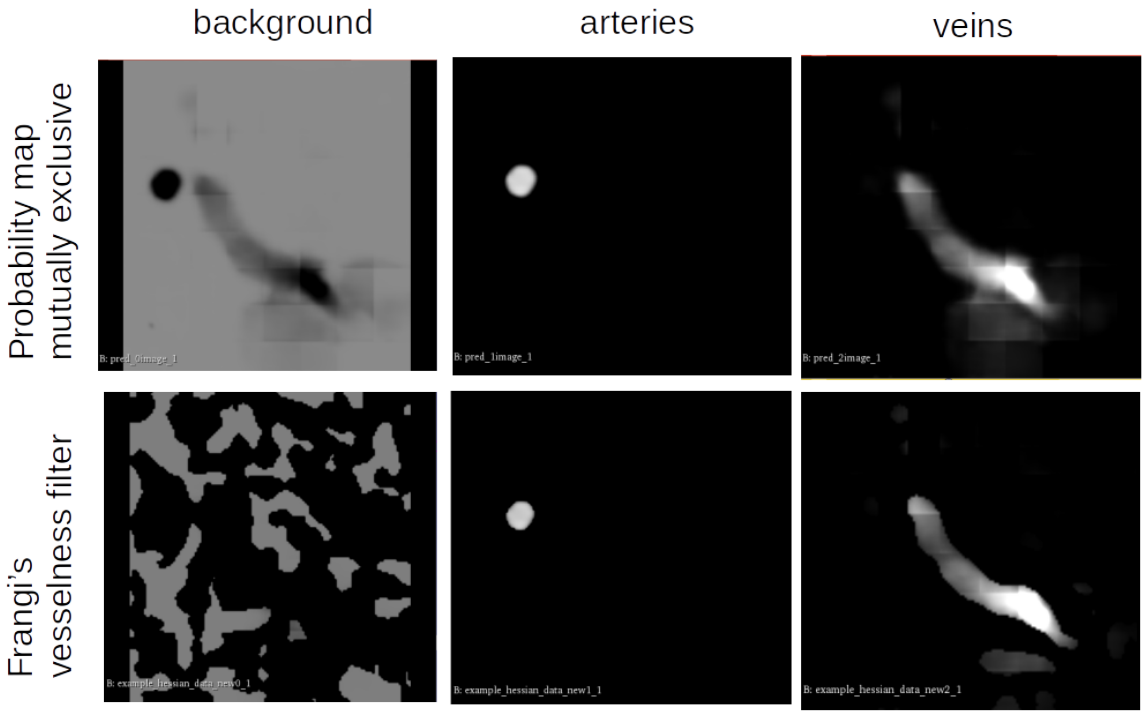


Figure 5: First row: map of probability by class as output of the network after using a Softmax function. Second row: Application of Frangi's vesselness filter. The gray scale corresponds to the probability of a voxel to belong to the considered class, ranging from 0 (black) to 1 (white).

However, applying it as post-processing on probability maps may greatly increase false positives. For this reason, we incorporated this idea directly into the training of the neural network as a loss function. In order to do so, we need to transform the reference segmentation, which is instead a binary object, since applying the Hessian matrix on this will result on a non-zero gradient only on the edges. The proposed pipeline for our Hessian-based vesselness is as follows:

1. **Gaussian kernel application.** We smooth both the binary segmentation of the reference and the probability map of the prediction by applying a Gaussian kernel strong enough to have zero-gradient only in the principal dimension. Nevertheless, a high standard deviation σ might make the small blood vessels completely disappear. As a compromise, we apply five different Gaussian kernels with σ_i, $i \in [1,5]$, ranging from 1 to σ_{max}+1 with a step of $\frac{\sigma_{max}}{4}$, as in (Frangi et al., 1998). The value of σ_{max} is found empirically, in relation to the size of both the selected patch and the structures to

be segmented. The final smoothed predicted P_{g_σ} (or reference R_{g_σ}) segmentation is computed as: $P_{g_\sigma} = \sum_{i=1}^{5}(g_{\sigma_i} * P)$, where P is the predicted segmentation. This way we ensure that we only have zero gradient along the main direction.

2. **Hessian matrix calculation**. We calculate the Hessian matrix as in Equation 1 for every voxel of the filtered segmentation P_{g_σ}, as $H(\sum_{i=1}^{5}(g_{\sigma_i} * P))$. It is important to emphasize that thanks to the convolution with the Gaussian kernel presented in the previous step, we ensure that the second partial derivatives for each voxel of the segmentation masks are all continuous and that each Hessian matrix is a symmetric matrix by Schwarz' theorem. This is fundamental because the eigenvalues of a real symmetric matrix are always real and its eigenvectors can always be taken to be real. Moreover the computation of eigenvalues is always differentiable for real symmetric matrices (Magnus, 1985). Further details on the differentiability are provided in Appendix A.
3. **Ordering of eigenvalues.** Due to the fact that in the case of predictions which include initially no structure or structures with different shapes and directions, using directly the vesselness function could result in a training slowdown or even in worse segmentation performance. This is because such vesselness functions require to sort the eigenvalues by magnitude, that in our case could end in prediction P and reference R having for the same voxel v a very similar vesselness score for structures with different preferential directions. In order to overcome this issue, we order the eigenvalues of predicted and reference voxels (resp. $P(v)$ and $R(v)$) via their associated normalized eigenvectors W. In particular we match the eigenvectors with the smallest angle between them, namely the minimum rotation required to overlap them, finding $\arg\min_{\mathcal{P}} \sum_{i=1}^{3} \|(W_{P(v)}\mathcal{P})_i - (W_{R(v)})_i\|_2$, where i is the index of the column representing the associated normalized eigenvector and $\mathcal{P}$ is a permutation matrix, ranging over all the possible permutation matrices. Figure 6 illustrates this ordering.
4. **Multi-scale supervision.** In order to inject as much information as possible to the network, we do the same for the subsequent three output levels of resolution using the above mentioned Deep Supervision technique in Equation 5. However, given the lower spatial definition of these outputs, the number of σ values used for the Gaussian kernel is set to $5 - q$, i.e. σ_i, $i \in [1, 5 - q]$, where q is the resolution level as in Equation 6 (0 as first level).

As motivated in the previous paragraph, our *vesselness* loss function, named tubular structures loss function and denoted by $TsLoss$, is composed of two parts: a first loss function to check the morphology of the structures, named morphological similarity loss function and denoted by $MsLoss$, by comparing the eigenvalues ordered by the eigenvectors matrix as presented above; a second loss function to force prediction

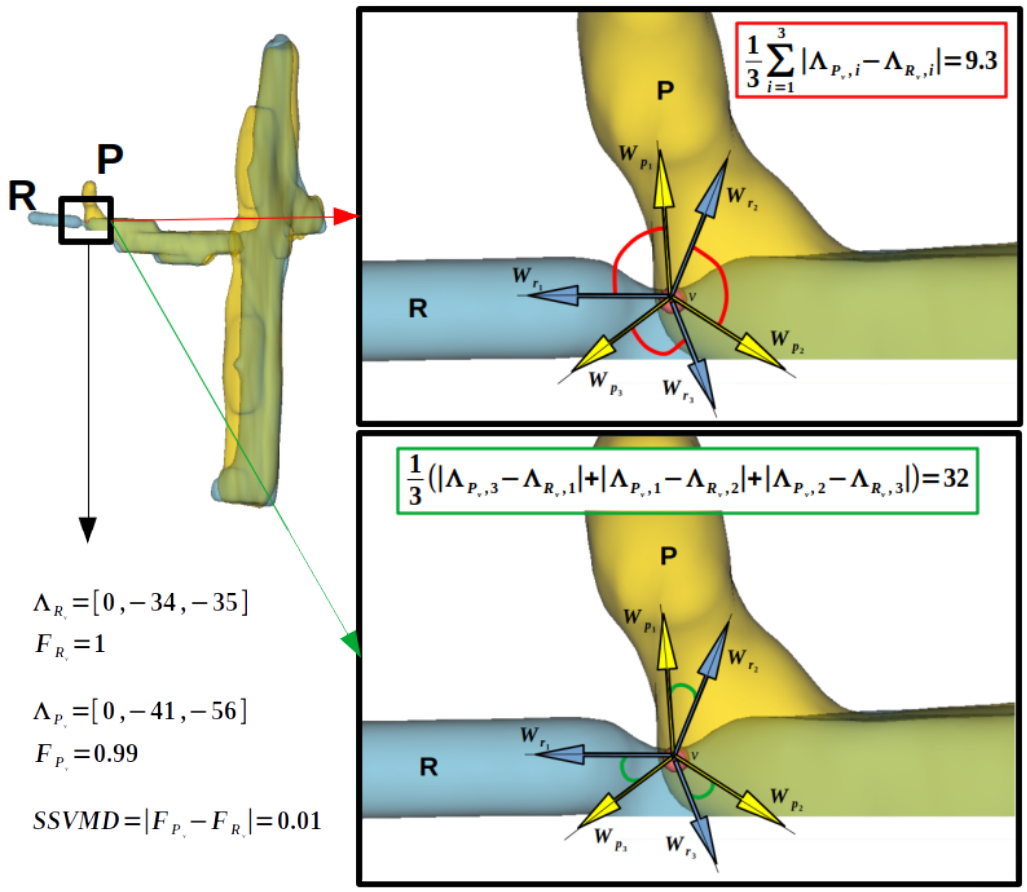


Figure 6: Ordering of eigenvalue values Λ of the same voxel v in order to allow for a fair comparison between them. Using SSVMD to compare vesselness score F results in a very low value even if prediction P (yellow) and reference R (light blue) have different main directions. Moreover, a comparison of eigenvalues ordered by magnitude does not reflect the real dissimilarity between the two segmentations (top box on the right). In order to overcome this issue, we order the eigenvalues of predicted and reference voxels via their associated normalized eigenvectors W, matching the W with smaller angle between them (bottom box on the right). This matching is important in this case, where a voxel-wise loss function would fail to correctly assess this error due to the fineness of the portion of the vessel.

of elongated structures as in Frangi's vesselness function, and thus named Frangi's vesselness loss function *FvLoss*.

The **Morphological similarity loss** function is defined for a single image and a single target structure at

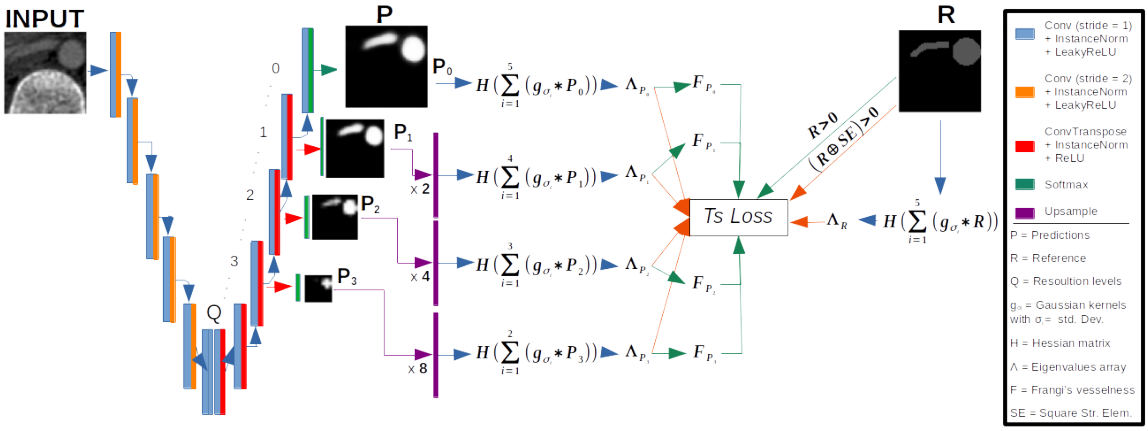


Figure 7: Tubular structures loss function ($TsLoss$) pipeline. The function is composed of two loss terms, one based on the eigenvalues Λ of the Hessian matrix calculated for every voxel of the filtered segmentation masks, and another based on the F Frangi's vesselness function (Frangi et al., 1998) calculated from these eigenvalues. The $TsLoss$ term is applied in deep supervision for the first 4 resolution levels, together with voxel-wise loss functions. See text for details.

resolution level q as:

$$MsLoss_q(P_q, R) = \frac{1}{3M_{\tilde{R}}} \sum_{m=1}^{M_{\tilde{R}}} \sum_{o=1}^{3} (\Lambda_o(H(\sum_{i=1}^{5-q}(g_{\sigma_i} * p_m)) - \Lambda_o(H(\sum_{i=1}^{5-q}(g_{\sigma_i} * r_m)))^2 \tag{7}$$

where g_{σ_i} are the $5-q$ different Gaussian filters applied to the segmentation masks, with standard deviations σ_i (as explained before in step 1), H is the Hessian matrix (step 2), Λ is the array containing the three eigenvalues of H ordered by the associated normalized eigenvectors (via the smallest angle as explained before in step 3 and Figure 6), p_m is the probability of a voxel m of the predicted segmentation P (i.e. the output probability map) at resolution q and r_m is the corresponding target sample of the reference segmentation R, while $M_{\tilde{R}}$ is the number of voxels of the dilation of R with a square structuring element of size $3 \times 3 \times 3$ (calculating eigenvalues over the entire image is expensive in terms of computational time, and the use of dilation revealed to be sufficient for our purpose thanks also to the combined use of voxel-wise loss functions). Moreover, this loss function allows us to take into consideration also flattened and deformed vessels (due to the presence of the tumor), in which instead the direct use of Frangi's vesselness may not be useful due to vesselness scores that can be very close to 0 and thus too similar to the score of *non-found* vessels.

The **Frangi's vesselness loss** function is designed in a non-supervised way for a single image and a

single target structure at resolution level q as:

$$FvLoss_q(P_q, R) = \frac{1}{M_{\hat{R}}} \sum_{m=1}^{M_{\hat{R}}} (1 - \frac{F(H(\sum_{i=1}^{5-q}(g_{\sigma_i} * p_m)))}{F_{max}}) \tag{8}$$

where $M_{\hat{R}}$ is the number of foreground voxels of R, F is the Frangi's vesselness presented in Equation 3, F_{max} is the maximum among the $M_{\hat{R}}$ Frangi's vesselness values. Due to the normalization via F_{max}, this loss function has values ranging between 0 to 1. This loss function forces voxels corresponding to the target structure to have a high vesselness value, avoiding the vanishing gradient problem. The use of the former $MsLoss$ function allows forcing the correct direction of the predicted vessels and thus enables the possibility of using Frangi's method without favoring incorrect predictions.

Finally, the complete **Tubular structures loss** function is:

$$TsLoss = \frac{1}{N} \sum_{n=1}^{N} \sum_{q=0}^{Q} w_q \cdot \frac{1}{N} \sum_{c=1}^{C} (w_{ms} MsLoss_q(P_{n,q,c}, R_{n,c}) + FvLoss_q(P_{n,q,c}, R_{n,c})) \tag{9}$$

where N is the batch size, C the number of structures to be segmented (not counting the background), $P_{n,q,c}$ is the prediction P_q for the class c of the image n of the batch, $R_{n,c}$ is the corresponding reference segmentation, and $Q + 1$ the number of output resolution levels taken into account (as explained before in step 4) with w_q as in Equation 6. This loss function is combined with the voxel-wise functions, such as cross-entropy and Soft Dice loss terms. All loss terms have a similar magnitude, except for the $MsLoss$. For this reason, we multiply the $MsLoss$ term by a scalar factor w_{ms} in the global loss function.

Figure 7 shows the complete pipeline for the proposed tubular structures loss function.

5. Experiments

5.1. Database

We worked on a private pediatric dataset of abdominal-visceral ceCT images gathered at Necker hospital of Paris. Scanners from 79 patients, with different modalities of contrast injection were available. All patients had a renal tumor, and images were acquired pre-operatively. The age ranges from 3 years old to 16, with an average of 2 years old. These exams were performed from 2007 to 2021 in the course of the normal care pathway of the patient and were studied retrospectively after anonymization. Images were acquired using four scanners from three different manufacturers: Siemens (SOMATOM Definition AS), Toshiba (Aquilion PRIME) and GE Medical Systems (LightSpeed and Revolution H).

Reference segmentations were performed by manual annotation under the supervision of medical experts. In particular we labeled:

- arteries and veins on 63 ceCT scanners with vascular/early phase of mono-phasic injection (see Figure 2);
- ureters on 13 ceCT scanners with excretory/delayed phase of mono-phasic injection (see Figure 2);
- arteries, veins and ureters on 3 ceCT scanners of bi-phasic injection (same acquisition with two injections done at different times, in order to have both arteriovenous and delayed phases, see Figure 8).

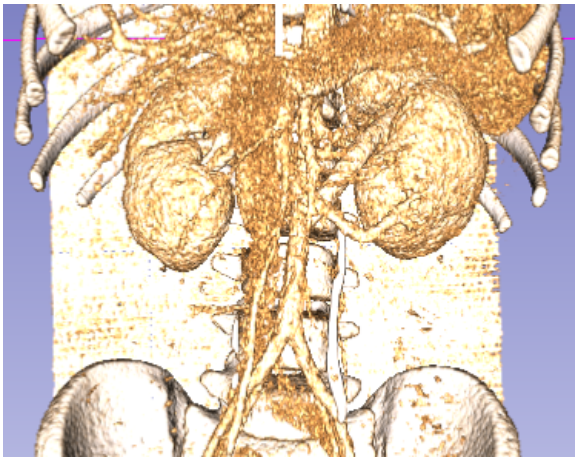


Figure 8: Volume-rendered example of single ceCT acquisition with bi-phasic injection in order to have both arteriovenous and delayed phases.

Examples of ceCT images with vascular phase, excretory phase and bi-phasic injection with respectively blood vessels, ureters and all structures labeled are shown in Figure 9.

All images are pre-processed as for the nnU-Net (Isensee et al., 2021), i.e. (i) a non-zero region cropping, (ii) a resampling of the images to have the same voxel size (0.89×0.46×0.46 mm^3), (iii) a clipping of the intensity values to the 0.5 and 99.5 percentiles of the foreground voxels, and (iv) a Z-scoring normalization. Then, all images are pre-cropped in the abdominal region of interest (ROI) as done in a previous work of ours (La Barbera et al., 2022) where we developed a simple, automatic and on-line method to select the first slice of the lungs and the last slice of the intestinal area as upper and lower landmarks (with number of slices increasing from lower to upper body), which is easy due to the strong presence of black pixels in ceCT acquisitions, as opposed to almost none in the liver-kidney portion. Please refer to the original article and its supplementary material for further details. Eventually, images are divided into 3D patches of size of $96 \times 160 \times 160$.

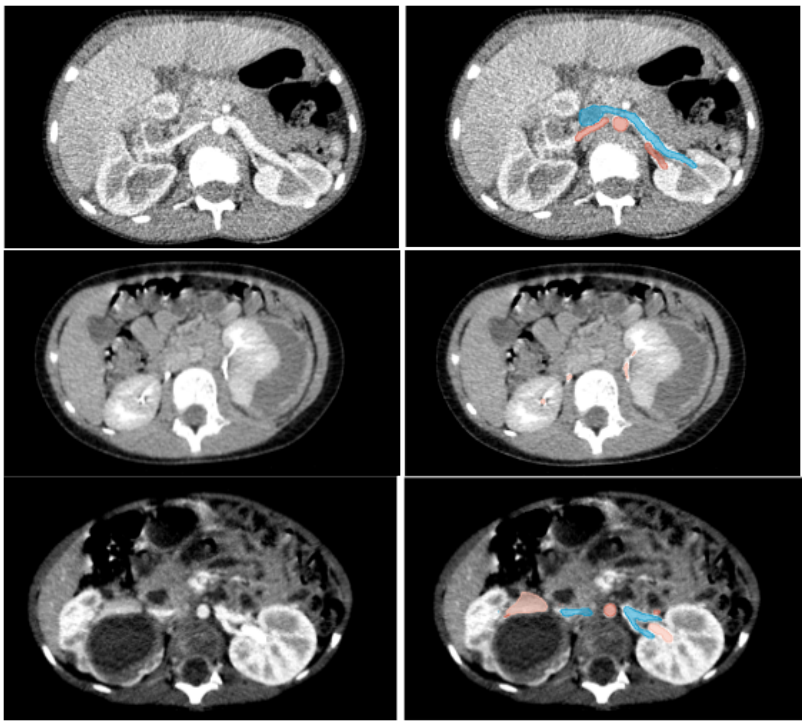


Figure 9: From top to bottom: ceCT images with vascular/early phase of mono-phasic injection, with excretory/delayed phase of mono-phasic injection and with bi-phasic injection. Left: original images. Right: labeled images, with arteries in red, veins in blue and ureters in pink.

5.2. Experiment protocols

In this work we present two sets of experiments, in each of which all methods are tested for a different application case.

The first one aims to segment arteries and veins with a single network for the reasons presented in Section 2.3, i.e. facilitating the choice in common edge voxels among multiple structures and improving learning tasks with less data using joint training. We used 46 vascular/early phase ceCTs for training and 5 for validation. For testing we used the images of 15 patients: 12 with mono-phasic injection and 3 with bi-phasic injection.

The second set of tests aims to train a network dedicated to ureters segmentation using the excretory/delayed phase ceCT of 10 patients, keeping 1 for validation. For inference we used the data from 5 patients: 2 with mono injection and the 3 with bi-phasic injection.

Unfortunately, it is not possible to train a single network capable of segmenting all three structures

because the ureters are not manually segmentable in arteriovenous ceCT nor the blood vessels in delayed ceCT. In addition, a training using both types of ceCT acquisition modalities would be difficult for the network due to the high inter-subject heterogeneity of the surrounding tissues, even after pixel intensities normalization.

5.3. Training implementation details

The number of training epochs is fixed at 1000 with 500 patches seen at each epoch, randomly chosen from the training set. The number of iterations at each epoch depends on the mini-batch size, specifically set to 4 in our experiments. All trainings and tests were run on a GPU NVIDIA® Tesla® P100 with 16 GB of VRAM. Patches are randomly extracted from the abdominal ROI for training and validation images. During inference we used a sliding window to extract patches from the abdominal ROI with overlapping of half of the size of the patch. In addition, to reduce artifacts and reduce the influence of position of patches close to the image borders, a Gaussian importance weighting is applied to the window, slightly increasing the weight of the center voxels in the softmax aggregation, as done in nnU-Net (Isensee et al., 2021).

An on-the-fly data augmentation is applied at each iteration and includes the following transformations, sequentially and with a given probability of application: rotation and scaling, Gaussian noise and Gaussian blur, change of brightness and contrast, simulation of low resolution images, gamma augmentation and mirroring. The data augmentation is implemented with the *batchgenerators* framework (Isensee et al., 2020) and applied to both the entire input images and the target structures alone (only the iconographic changes), in order to better manage heterogeneity of image intensity. This step is critically important in order to make up for the problems associated with a limited database such as our pediatric one.

Stochastic Gradient Descent (SGD) with a Nesterov momentum of 0.99 is used as optimizer for all tests. The initial learning rate lr of 0.01 is reduced following the poly learning rate policy (Cehn et al., 2018), decaying at each epoch e by multiplication of the initial lr by a factor of $(1 - \frac{e}{E})^{0.9}$ where E is the total number of epochs. In our proposed method, we also assessed the performance of Adam (Kingma and Ba, 2015) and Adagrad (Duchi et al., 2011) optimizers with initial lr of 10^{-3}, taking inspiration from the work by Lahlouh et al. (2022). This study is shown in Appendix B.

We empirically found the best value of weight w_{ms} of $MsLoss$ in Equation 9 as 0.05 for arteries and veins, and 0.01 for ureters, and the weights of F (Equation 3) for $FvLoss$ (Equation 8) as $\alpha = 0.1, \beta = 0.1$ and $\gamma = 2$. We also found the best σ_{max} to ensure zero gradient only in the main direction equal to 25. The search of these parameters is shown in Appendix C.

In addition to $TsLoss$, cross-entropy (CE) and Soft Dice score (to which we will refer simply as Dice in tables and figures) were used in the loss function for trainings.

Proposed oversampling method. Due to the strong imbalance between patches with only background and patches with structures, an oversampling technique for selecting at least 50% of patches with a minimal number of voxels per structure, $MinPix$, was adopted for all methods in Table 1 for which no oversampling was done (M1 - M4). This choice was made because early results without such a technique had very poor performance for veins and ureters. Moreover, for networks that segment arteries and veins, an additional oversampling on patches with structures is done due to the lower presence of the latter, ensuring that the previous 50% are equally distributed. A sufficient $MinPix$ was empirically found as 1000 voxels for all the structures. For Kid-Net (M5) and nnU-Net (M6) we used the oversampling technique of these methods, previously presented.

5.4. Evaluation measures

For the quantitative evaluation of the segmentation results, we compute three different categories of measures. The first category includes spatial overlap based measures: Dice score, precision and recall. The second category comprises spatial distances, and the 95th percentile of Hausdorff distance (95HD) is used. Further details on these measures can be found in the work of Taha and Hanbury (2015). However, these measures carry with them the limitations presented in the previous sections, and for this reason, alone they are not sufficient for a proper evaluation of the segmentation results. To overcome these limitations, we decided to use, as a third category, our proposed morphological similarity loss ($MsLoss$ in Equation 7) for the motivations discussed in Section 4.1. In the result tables we refer to this measure as $\Delta\Lambda$.

Moreover, we perform a further analysis for arteries and veins, which we refer to as Recall analysis. In this study, arteries and veins of the reference segmentation are semantically segmented into substructures that differ in diameters and directions. Arteries are divided into aorta, renal arteries and celiac artery, and veins into cava vein and renal veins. Aorta and cava vein are larger and follow approximately a constant direction, renal arteries and veins are very tiny vessels with irregular directions, while the celiac artery has a medium diameter between the previous structures and has a T-shape that branches perpendicularly from the aorta on the coronal plane. An example of this sub-division is shown in Figure 10. The more a vessel is fine and irregular, the more the difficulty in manual segmentation increases. For this reason, the recall measure between the prediction and each of these parts of the vessel tree is calculated. Indeed, from our manual segmentation and correction experience, we believe that having fewer false negatives is really important to

speed up the 3D anatomical modeling process. Moreover, in case of major errors, manual segmentation of missing parts takes longer than false positive removal. These errors can be identified as false negatives on the one hand in a non-segmented branch, a missing portion of branches or even a branch broken into many small portions, and as false positives on the other hand in an additional false branch, blobs in areas not close to the vessel tree or small connected components in a non-debranching area. In this case, such false positives are easily identified and removed thanks to the manual segmentation tools provided by software tools such as Slicer3D (Fedorov et al., 2012), while false negatives require more careful and time-demanding manual segmentation. This is also due to the proximity of arteries and veins and the fineness of the tubular structures being segmented.

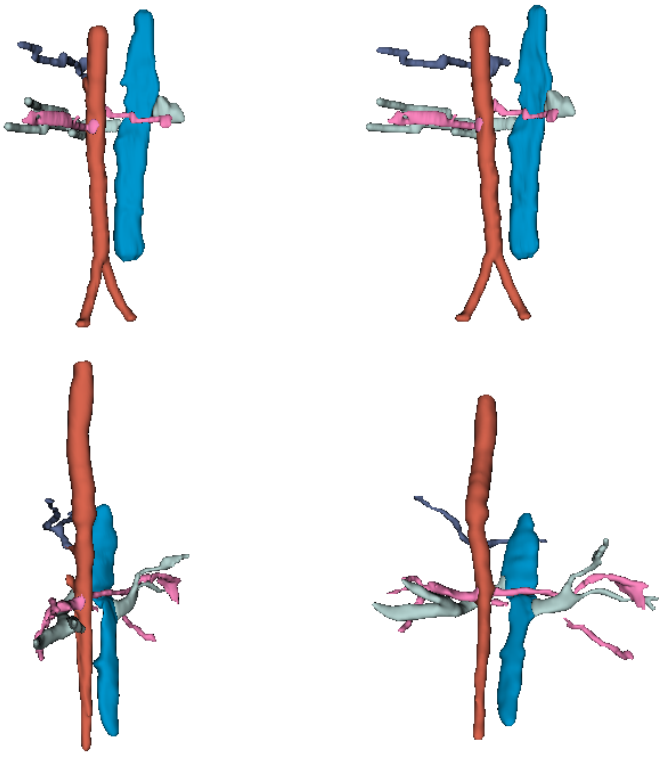


Figure 10: Examples of subdivision of the reference segmentation for the Recall analysis. Arteries are divided into aorta (in red), renal arteries (fuchsia) and celiac artery (purple), while veins are divided into cava vein (blue) and renal veins (light blue). Left: lateral view. Right: front view.

6. Results and Discussion

In this section we show and discuss the results obtained with the methods in Table 1 and with our proposed method. A further study on different implementations of the proposed vesselness loss functions in both the best method found in the comparison and the proposed method is detailed in Appendix D.

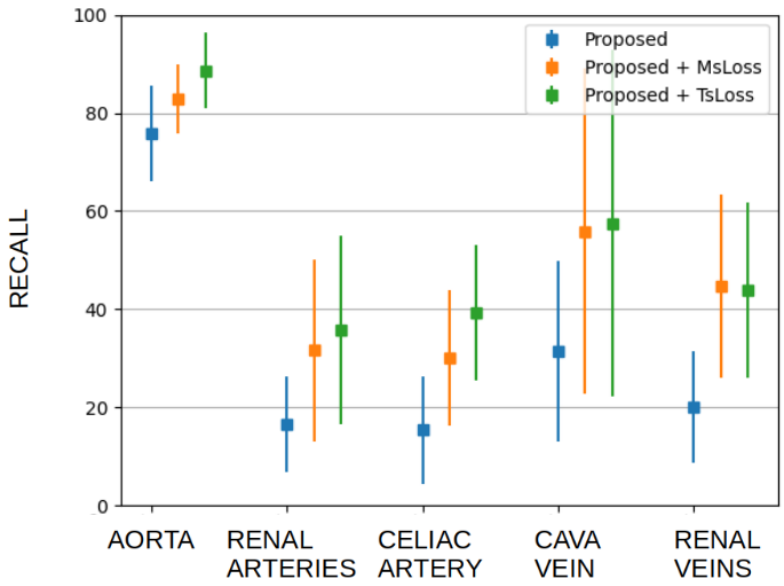


Figure 11: Recall analysis on the last three rows of experiments in Table D.6 for the different structures. Arteries are divided into aorta, renal arteries and celiac artery, while veins are divided into cava vein and renal veins.

6.1. Arteries and veins segmentation

Table 2 shows the quantitative results for the comparison of state-of-the-art methods for arteries and veins using the evaluation measures presented in the previous section. The Deep Distance Transform method of Wang et al. (2020) (M2) outperforms the other state-of-the-art methods in all the measures (precision and recall need to be read together). Nevertheless, once the GAR post-processing is removed (which takes a long time in inference), the performances drop. The DenseBiasedU-Net of He et al. (2020) (M4) reduces false positives, as we can see from the values of precision and 95th of Hausdorff Distance (95HD). The problem with this technique lies in the lower depth of U-Net, due to the large amount of memory required by dense connections, which limits the capability of the network to extract information. Regarding Kid-Net method of Taha et al. (2018) (M5), we cannot say if the worse results are due to the oversampling technique

Table 2: Results on 15 patients using patches of size 96×160×160 obtained with the methods in Table 1 and with our proposed method for arteries (A) and veins (V) segmentation. *pre-selected location sources. **without GAR. ***without HRA and DPA. Mean and standard deviation of the results are given. Other quantitative results on arteries and veins segmentation using patches of smaller size and different loss function implementations are shown in Appendix D.

Method	Oversampling	S	Dice Score [%] (↑)	Precision [%] (↑)	Recall [%] (↑)	95HD [mm] (↓)	ΔΛ (↓)
M1 - TopNet*	Proposed	A	50.23 (19.46)	80.13 (23.31)	38.27 (18.06)	23.98 (8.63)	22.22 (2.27)
(Keshwani et al., 2020)	*MinPix*	V	46.05 (30.64)	64.07 (28.88)	41.10 (29.72)	22.48 (13.27)	20.34 (3.59)
M2 - Deep Distance Transform**	Proposed	A	69.99 (3.44)	85.08 (10.63)	60.42 (5.28)	16.65 (13.29)	21.62 (1.99)
(Wang et al., 2020)	*MinPix*	V	37.29 (23.06)	82.59 (19.62)	26.77 (19.69)	27.28 (26.41)	22.51 (3.11)
M2 - Deep Distance Transform	Proposed	A	71.91 (3.85)	81.39 (10.44)	65.47 (5.42)	16.40 (13.45)	20.19 (1.95)
(Wang et al., 2020)	*MinPix*	V	41.98 (23.76)	78.55 (18.95)	32.32 (22.41)	27.57 (29.59)	21.78 (2.95)
M3 - Deep Distance Transform	Proposed	A	63.73 (5.63)	78.52 (13.59)	54.56 (4.75)	22.75 (13.95)	22.44 (1.79)
(Ma et al., 2020; Wang et al., 2020)	*MinPix*	V	32.26 (22.60)	82.26 (26.92)	22.21 (16.81)	36.29 (20.49)	22.28 (3.25)
M4 - DenseBiasedU-Net***	Proposed	A	65.76 (4.15)	86.95 (11.79)	53.71 (4.64)	18.26 (15.17)	23.42 (1.95)
(He et al., 2020)	*MinPix*	V	34.89 (22.20)	85.86 (10.37)	24.66 (18.93)	19.62 (10.10)	23.07 (3.09)
M5 - Kid-Net	Random and	A	65.70 (3.23)	88.93 (9.88)	52.82 (5.29)	19.48 (7.88)	23.19 (1.92)
(Taha et al., 2018)	dynamic weighting	V	28.36 (23.56)	87.01 (16.84)	19.78 (18.87)	28.52 (21.23)	22.02 (3.89)
M6 - nnU-Net	Foreground in	A	68.05 (5.26)	84.43 (11.05)	57.85 (6.32)	15.55 (9.13)	22.12 (1.99)
(Isensee et al., 2021)	33.3% of mini-batch	V	39.78 (16.80)	79.36 (12.87)	28.15 (14.27)	25.12 (28.97)	23.18 (2.60)
Proposed 1:	Proposed	A	63.45 (5.67)	71.73 (9.99)	57.87 (7.31)	17.46 (9.65)	21.15 (1.93)
Deep Sup. Dice + CE	*MinPix*	V	42.64 (20.12)	76.67 (13.17)	31.84 (17.12)	23.55 (17.00)	21.38 (3.27)
Proposed 2:	Proposed	A	75.88 (3.03)	87.92 (4.64)	67.09 (6.15)	**9.79 (4.58)**	19.67 (2.39)
Deep Sup. Dice + CE + MsLoss	*MinPix*	V	**60.33 (25.63)**	81.76 (9.33)	53.26 (27.28)	**18.65 (21.29)**	19.47 (3.65)
Proposed 3:	Proposed	A	**76.77 (3.93)**	80.41 (10.17)	75.04 (7.66)	10.02 (5.80)	**17.73 (2.49)**
Deep Sup. Dice + CE + TsLoss	*MinPix*	V	58.35 (26.79)	75.83 (13.01)	54.09 (28.87)	18.84 (21.31)	**19.37 (3.29)**

or to the only use of cross-entropy as loss function. Given the high presence of false negatives (low recall in Table 2) we think it may be due more to the latter. The TopNet method (M1) shows balanced results between arteries and veins. However, the final segmentation strongly depends on the output of the first decoder, i.e. the vessel mask, which is only optimized via the Dice loss function (without deep supervision). For this reason, the algorithm has difficulties finding blood vessels with a smaller diameter than the main ones, as explained in Section 4.1. Indeed, this method is conceived for blood vessels inside the liver that have the same size and the same intensity. Results are thus not surprising. Overall, nnU-Net of Isensee et al. (2021) (M6) can be identified as the best technique, where deep supervision with Dice and CE leads to good results without any heavy post-processing. For this reason and because of the ease in building on this method, we decided to apply our oversampling method within nnU-Net and then add our proposed loss

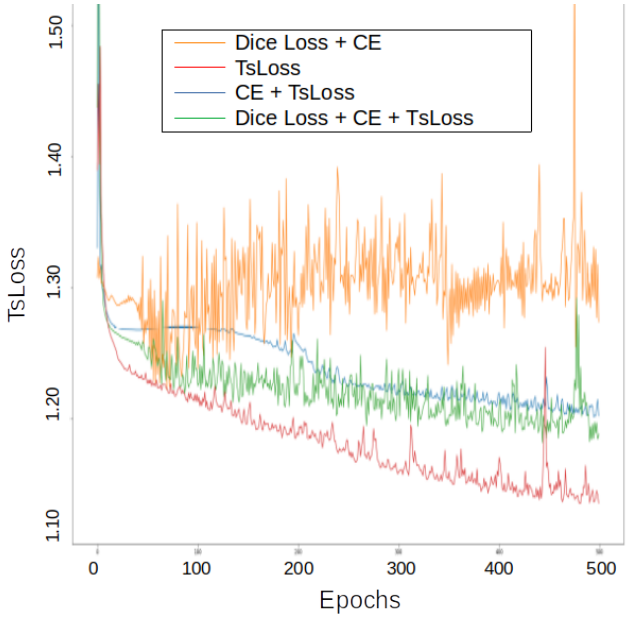
functions. The proposed oversampling method is effective and comparable to the one presented in nnU-Net, nevertheless it seems to better balance the number of voxels examined for each class in the case of multiple structures, reducing the difference in performance among classes. The use of $MsLoss$ greatly improves segmentation results, highly reducing false negatives and both spatial distance and morphological similarity between prediction and reference. The addition of $FvLoss$ to build the final $TsLoss$ decreases even more the number of false negatives, at the expense of increasing false positives, but with significant morphological similarity improvement.

This is best seen from the qualitative results in Figure 13 (worst, average and best results for each method) and from the Recall analysis in Figure 11. In the latter, we can infer that the use of the proposed loss functions allows the network to identify better vessels of smaller diameter and with different directions (and thus morphology) from each other. Moreover, better results for both vein sub-structures underline that such vesselness loss functions may overtake heterogeneity problems.

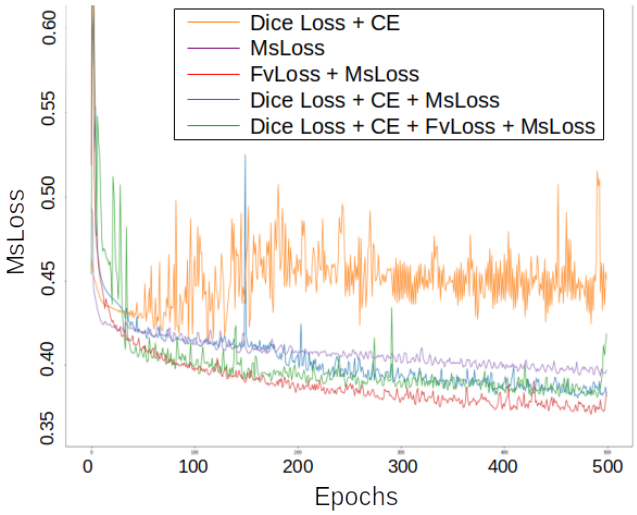
We underline that the use of Dice and cross-entropy (CE) alone does not consider the minimization of morphological differences, as can be seen from the graphs in Figures 12a and 12b. In these experiments we selected 100 patches with high presence of the target structures and trained with different combinations of loss functions for 500 epochs. The $TsLoss$ and $MsLoss$ values are plotted for each training, including one in which it is not used as a loss function during training (in orange). In Figure 12a, the proposed loss function is considered in its entirety as $TsLoss$ and we can see that it is not minimized during Dice+CE training. In Figure 12b, the attention is on the morphological loss term ($MsLoss$), and we can notice that the use of the specific vesselness loss term ($FvLoss$) helps in better optimizing the $MsLoss$ (red and green curve).

6.2. Ureters segmentation

Similar considerations can also be made for ureters segmentation. However, the database used was very limited and the test set of only 5 subjects may not be representative. Quantitative results are shown in Table 3. Here the most performing state-of-the-art method, from a spatial overlap and morphological point of view, is DenseBiasedU-Net of He et al. (2020) (M4). This may be due to the very small thickness of the ureters (comparable to renal blood vessels, see Section 1.1), whose information is better propagated thanks to the use of the dense connections. Nevertheless, memory usage is already up to the limit with such a network and adding another loss term such as the one we propose is not possible. Moreover, the results of DenseBiasedU-Net have a high number of false positives that produce significant errors as we can see from the high 95HD and precison, in contrast to what observed in Table 2. The other networks show similar be-



(a) Evolution of $TsLoss$



(b) Evolution of $MsLoss$

Figure 12: Evolution of proposed losses during training for different loss functions combinations.

haviors to those for blood vessel segmentation, with nnU-Net of Isensee et al. (2021) (M6) performing better when not considering the post-processing step (namely GAR) of the Deep Distance Transform of Wang et al.

Table 3: Results on 5 patients using patches of size 96×160×160 obtained with the methods in Table 1 and with our proposed method for ureters (U) segmentation. *pre-selected location sources. **without GAR. ***without HRA and DPA. Mean and standard deviation of the results are provided.

Method	Oversampling	S	Dice Score [%] (↑)	Precision [%] (↑)	Recall [%] (↑)	95HD [mm] (↓)	ΔA (↓)
M1 - TopNet* (Keshwani et al., 2020)	Proposed *MinPix*	U	52.82 (29.45)	82.29 (12.29)	47.59 (12.29)	18.53 (30.84)	23.07 (4.91)
M2 - Deep Distance Transform** (Wang et al., 2020)	Proposed *MinPix*	U	54.47 (28.49)	79.94 (12.03)	49.82 (29.72)	18.45 (27.12)	23.27 (4.46)
M2 - Deep Distance Transform (Wang et al., 2020)	Proposed *MinPix*	U	55.50 (28.17)	83.83 (9.98)	48.50 (27.93)	17.71 (25.62)	22.95 (4.36)
M3 - Deep Distance Transform (Ma et al., 2020; Wang et al., 2020)	Proposed *MinPix*	U	46.69 (28.77)	81.22 (13.56)	43.03 (34.25)	18.01 (24.62)	24.69 (6.28)
M4 - DenseBiasedU-Net*** (He et al., 2020)	Proposed *MinPix*	U	57.31 (28.14)	70.25 (8.96)	57.45 (31.05)	23.42 (24.87)	21.45 (3.91)
M5 - Kid-Net (Taha et al., 2018)	Random and dynamic weighting	U	50.38 (26.66)	79.97 (14.84)	46.07 (31.99)	17.71 (24.35)	23.95 (5.97)
M6 - nnU-Net (Isensee et al., 2021)	Foreground in 33.3% of mini-batch	U	54.55 (24.82)	78.95 (13.20)	49.28 (29.35)	19.02 (24.76)	23.21 (4.97)
Proposed 1: Deep Sup. Dice + CE	Proposed *MinPix*	U	53.58 (24.03)	80.55 (10.40)	45.15 (29.95)	18.17 (24.29)	22.66 (4.32)
Proposed 2: Deep Sup. Dice + CE + MsLoss	Proposed *MinPix*	U	**59.51 (25.85)**	79.89 (6.53)	54.41 (26.85)	**8.15 (12.47)**	**20.90 (5.65)**
Proposed 3: Deep Sup. Dice + CE + TsLoss	Proposed *MinPix*	U	53.30 (29.49)	84.03 (12.04)	49.64 (33.79)	17.09 (25.78)	21.15 (3.02)

(2020) method (M2). When applying our oversampling method to nnU-Net we get worse results for overlapping measures (lower Dice score and combination of precision and recall), and better results for spatial and morphological measures (lower Hausdorff distance and morphological similarity). The use of $MsLoss$ improves all the measures, with a particular decrease in false negatives. The use of $FvLoss$ in combination with $MsLoss$ worsens these results, and this may be due to an inappropriate choice of parameters or to the not tubular shape of the renal calyces (i.e. the beginning of excretory pathways that is usually segmented as ureters, as in Figure 4). Finally, it is important to note that, unlike what was observed for blood vessels, there are generally few differences in quantitative results among the different techniques. The high standard deviation for all measures as well as a very limited test set, as mentioned earlier, make it difficult to draw conclusions with confidence.

Differences are barely visible also for qualitative results. In Figure 14, worst, average and best results from each method are displayed. Some ceCT images have a high contrast heterogeneity in these structures, particularly in the case of biphasic injection, such as for the patient shown on the top of Figure 14. Further-

more, the low number of voxels of the ureters makes overlap measurements not very reliable, as we can see from the last two patients in Figure 14 (predictions almost complete but Dice score under 80%). Eventually, thanks to this figure it is easier to understand the fineness of these structures on the tubular section and the presence of the renal calyces discussed above, which both make segmentations even more complicated.

7. Conclusions

In this work, we proposed for the first time an assessment and comparison of state-of-the-art methods for segmentation of renal tubular structures (arteries, veins and ureters) on ceCT images from pediatric patients with renal pathologies. The assessment focused on prior art addressing these images and structures, but other interesting approaches were also analyzed. Although the analysis may not be fully extensive, especially for ruled-based methods or *non-deep* machine learning approaches, preliminary tests resulted in great difficulty in segmenting the renal tubular structures in pediatric ceCT images acquired on arteriovenous phase. Regarding comparisons, although more methods would have been interesting to compare, the lack of codes and the few details available in the articles did not allow us to reproduce them with confidence.

We also proposed a new loss function designed from the so-called vesselness function to improve state-of-the-art results. This loss function is based on the comparison of eigenvalues of the Hessian matrix of segmentation masks and Frangi's vesselness (Frangi et al., 1998) enhancement on target voxels in a multi-scale deep supervision way. The combination of this tubular structures loss with voxel-wise loss functions allows overcoming some problems of the latter, such as the difficulty in correctly optimizing tubular structures with elongated shape, intra-scale changes and inter-anatomy variation. The results demonstrated great improvements from a morphological point of view, with segmentation results showing fewer interruptions, at the expense of a slight increase in false positives. This confirms that the use of voxel-wise loss functions and overlapping measures is not sufficient for the evaluation of such structures. The use of the second loss term, related to Frangi's vesselness, appears to be of no benefit in cases where the structure has non-tubular regions. In addition, several hyperparameters are introduced with the use of this loss function.

Future work points to an automation of the choice of loss function parameters, as well as to new experiments on different target tubular structures such as liver or brain vessels on ceCT images but also on other acquisition modalities. Moreover, for ureters segmentation, new experiments will be performed once more data will be collected, and a semantic division between renal calyces and ureters should be performed in the manual reference segmentation.

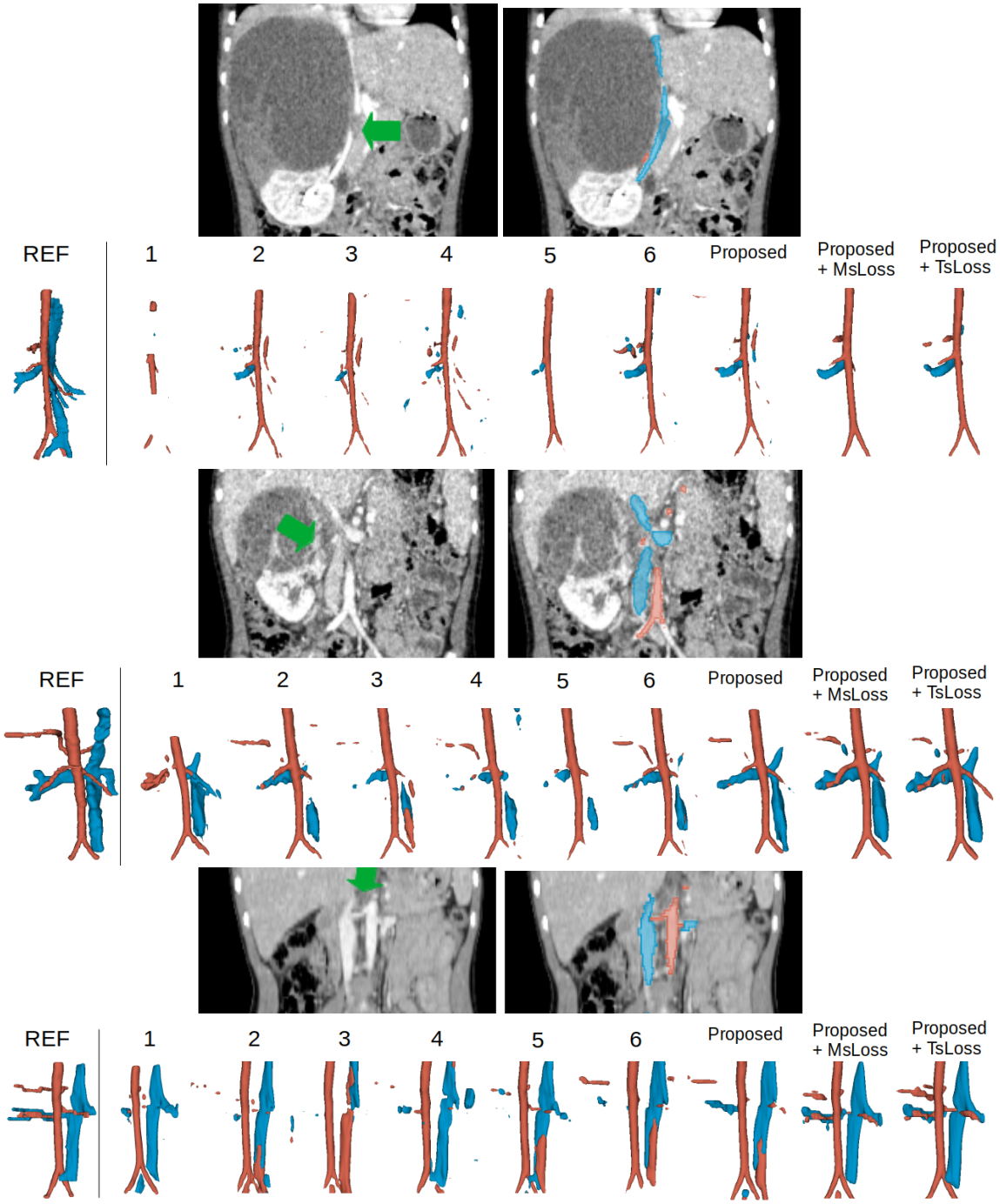


Figure 13: Worst (top, $DiceScore$: $A = 78.30$; $V = 15.01$), average (middle, $A = 77.94$; $V = 62.07$) and best (bottom, $A = 83.56$; $V = 84.29$) segmentation results (averaged between arteries and veins) on single patients for our method using the proposed $TsLoss$, and the results for the same patients with the other methods. For each patient we present also one coronal slice highlighting the most peculiar and difficult areas with green arrows (left: input ceCT; right: reference segmentation). The number-method correspondence is shown in Table 1. The order of the methods correspond to that shown in Table 2. 3D models are back-front to make renal arteries visible, otherwise are covered by renal veins. Arteries in red and veins in blue.

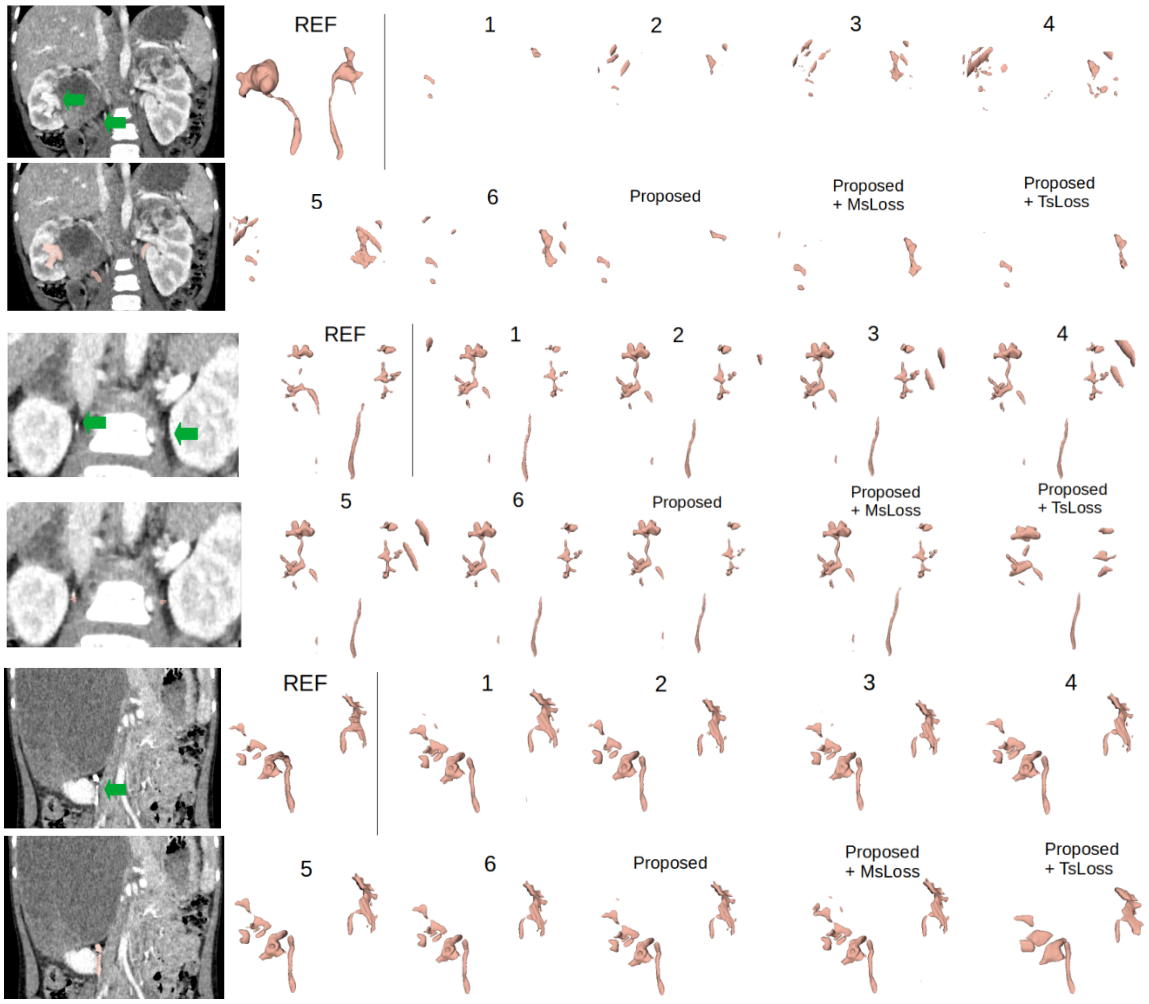


Figure 14: Worst (top, $DiceScore : U = 14.92$), average (middle, $U = 67.80$) and best (bottom, $U = 79.05$) segmentation results on ureters (in color pink) on single patients for our method using the proposed $MsLoss$, and the results for the same patients with the other methods. For each patient we show also one coronal slice highlighting the most peculiar and difficult regions with green arrows (left: input ceCT; right: reference segmentation). The number-method correspondence is shown in Table 1, while the order corresponds to that shown in Table 3. The differences are barely visible and do not result in major differences from the point of view of subsequent manual correction.

Appendix A. Differentiability of Morphological similarity loss function

According to the first Magnus' theorem (Magnus, 1985):

"Let H_o be a real symmetric $n \times n$ matrix. Let W_o be a normalized eigenvector associated with a simple eigenvalue λ_o of H_o. Then a real-valued function λ and a vector function W are defined for all H in some neighborhood $N(H_o) \subset \mathbb{R}^{n \times n}$ of H_o, such that: $\lambda(H_o) = \lambda_o$, $W(H_o) = W_o$, and $HW = \lambda W$, $W'W = 1$, $H \in N(H_o)$, where W' denote the transpose of W. Moreover, the functions λ and W are ∞ times differentiable on $N(H_o)$, and the differentials at H_o are: $d\lambda = W_o'(dH)W_o$ and $dW = (\lambda_o I_n - W_o)^+(dH)W_o$ (where A^+ is the Moore–Penrose inverse of A, i.e. the pseudo-inverse). Equivalently, the derivative at H_o for λ is given by: $\frac{\partial \lambda}{\partial (vecH)'} = W_o' \otimes W_o'$ or $\frac{\partial \lambda}{\partial H} = W_o W_o'$, where $vecH$ denotes the column vector that stacks the columns of H one underneath the other, and $\otimes$ denotes the Kronecker product."

This means that if the Hessian matrix H is real symmetric, as in our case, the eigenvalues function λ is differentiable in the neighborhood of H via the product of the normalized eigenvectors W_o (associated with the simple eigenvalue λ_o of H) with its transpose.

To give further details, the derivative of the cost function $MsLoss$ for a single predicted voxel P_m with respect to a parameter Z_p of the network Z can be written as:

$$\frac{\partial MsLoss}{\partial Z_p} = \frac{\partial MsLoss}{\partial \Lambda} \frac{\partial \Lambda}{\partial H} \frac{\partial H}{\partial g_\sigma} \frac{\partial g_\sigma}{\partial P_m} \frac{\partial P_m}{\partial Z_p} \tag{A.1}$$

where $P_m = (Z(I))(m)$, m is a voxel at position (x_m, y_m, z_m) of the input image I, $\Lambda = (\lambda_1, \lambda_2, \lambda_3)$ and $MsLoss = \frac{1}{3M} \sum_{m=1}^{M} [(\lambda 1_{P_m} - \lambda 1_{R_m})^2 + (\lambda 2_{P_m} - \lambda 2_{R_m})^2 + (\lambda 3_{P_m} - \lambda 3_{R_m})^2]$.

For the sake of simplicity we define $V_m = g_\sigma * P_m$, and using the Magnus's theorem with an eigenvector $W_o = (w_{o1}, w_{o2}, w_{o3})'$ for each $\lambda_{o \in [1,3]}$, we have for a single voxel m (with M equal to the number of voxels)

with respect to a parameter Z_p:

$$\frac{\partial MsLoss}{\partial Z_p} =$$

$$= \frac{\partial MsLoss}{\partial \Lambda}\frac{\partial \Lambda}{\partial H}\frac{\partial H}{\partial V_m}\frac{\partial V_m}{\partial Z_p} =$$

$$= \frac{\partial MsLoss}{\partial \Lambda}\frac{\partial \Lambda}{\partial (vecH)'}\frac{\partial (vecH)'}{\partial V_m}\frac{\partial V_m}{\partial Z_p} =$$

$$= \begin{bmatrix} \frac{\partial MsLoss}{\partial \lambda_1} & \frac{\partial MsLoss}{\partial \lambda_2} & \frac{\partial MsLoss}{\partial \lambda_3} \end{bmatrix} \times \begin{bmatrix} \frac{\partial \lambda_1}{\partial (vecH)'} \\ \frac{\partial \lambda_2}{\partial (vecH)'} \\ \frac{\partial \lambda_3}{\partial (vecH)'} \end{bmatrix}$$

$$\times \begin{bmatrix} \frac{\partial h_{x2}}{\partial (V_m)_1} & \cdots & \frac{\partial h_{x2}}{\partial (V_m)_M} \\ \cdots & \cdots & \cdots \\ \frac{\partial h_{z2}}{\partial (V_m)_1} & \cdots & \frac{\partial h_{z2}}{\partial (V_m)_M} \end{bmatrix} \times \begin{bmatrix} \frac{\partial (V_m)_1}{\partial Z_p} \\ \cdots \\ \frac{\partial (V_m)_M}{\partial Z_p} \end{bmatrix} =$$

$$= \begin{bmatrix} \frac{\partial MsLoss}{\partial \lambda_1} & \frac{\partial MsLoss}{\partial \lambda_2} & \frac{\partial MsLoss}{\partial \lambda_3} \end{bmatrix} \times \begin{bmatrix} W_1' \otimes W_1' \\ W_2' \otimes W_2' \\ W_3' \otimes W_3' \end{bmatrix}$$

$$\times \begin{bmatrix} \frac{\partial h_{x2}}{\partial (V_m)_1} & \cdots & \frac{\partial h_{x2}}{\partial (V_m)_M} \\ \cdots & \cdots & \cdots \\ \frac{\partial h_{z2}}{\partial (V_m)_1} & \cdots & \frac{\partial h_{z2}}{\partial (V_m)_M} \end{bmatrix} \times \begin{bmatrix} \frac{(V_m)_1}{Z_p} \\ \cdots \\ \frac{(V_m)_M}{Z_p} \end{bmatrix}$$

where $W_o' \otimes W_o' = \begin{bmatrix} w_{o1} & w_{o2} & w_{o3} \end{bmatrix} \otimes \begin{bmatrix} w_{o1} & w_{o2} & w_{o3} \end{bmatrix} =$

$$= \begin{bmatrix} w_{o1}w_{o1} & w_{o1}w_{o2} & w_{o1}w_{o3} & \ldots & w_{o3}w_{o1} & w_{o3}w_{o2} & w_{o3}w_{o3} \end{bmatrix}$$

This matrix multiplication returns a single value, i.e. the gradient of $MsLoss$ with respect to the parameter Z_p, which will be multiplied by lr in order to update the parameter (e.g. weight) Z_p.

Appendix B. Comparison of different optimizers

In our method, we assessed also the performance of Adam (Kingma and Ba, 2015) and Adagrad (Duchi et al., 2011) optimizers with initial lr of 10^{-3}, taking inspiration from Lahlouh et al. (2022). Results are shown in Table B.4, which show that the use of SGD with lr=0.01 leads to better performance.

Table B.4: Benchmarks using different optimizers and different *lr* on our proposed method using the baseline loss function: CE+Soft Dice with deep supervision. Patch size 32×64×64. A: Arteries; V:Veins. Mean and standard deviation of the results are shown.

Technique		Structures					
Optimizer	*lr*	S	Dice Score [%] (↑)	Precision [%] (↑)	Recall [%] (↑)	HD95 [mm] (↓)	ΔΛ (↓)
SGD	0.01	A	**74.25 (4.24)**	85.80 (11.25)	67.02 (7.45)	13.76 (17.15)	**19.65 (2.04)**
		V	**58.67 (27.06)**	82.73 (9.89)	51.10 (27.69)	16.17 (12.86)	**19.77 (3.36)**
SGD	0.001	A	73.39 (3.25)	86.54 (6.65)	64.48 (7.11)	**9.23 (3.56)**	20.35 (2.01)
		V	56.89 (28.44)	83.03 (8.45)	50.47 (28.78)	**14.59 (11.67)**	19.99 (3.40)
ADAM	0.01	A	69.97 (5.17)	79.07 (13.14)	64.89 (8.08)	20.21 (17.96)	19.91 (2.39)
		V	51.84 (24.70)	61.55 (17.61)	52.53 (28.82)	24.71 (24.36)	20.16 (3.36)
ADAM	0.001	A	74.32 (2.48)	85.88 (8.34)	66.52 (6.99)	9.48 (4.58)	19.92 (1.87)
		V	58.12 (26.24)	80.69 (12.85)	50.21 (27.59)	16.23 (17.87)	20.05 (3.63)
ADAGRAD	0.01	A	73.54 (4.83)	85.86 (11.79)	66.02 (7.64)	14.46 (17.13)	19.90 (2.38)
		V	57.40 (27.07)	83.69 (8.17)	50.31 (27.81)	16.05 (10.15)	19.76 (3.59)
ADAGRAD	0.001	A	60.34 (6.90)	53.75 (10.66)	70.77 (5.97)	16.11 (1.87)	18.67 (1.45)
		V	20.95 (14.76)	51.22 (27.09)	13.55 (10.11)	13.87 (3.56)	22.55 (2.98)

Appendix C. Parameters research

We conducted a parameters search for α, β and γ of Frangi's vesselness (Frangi et al., 1998) and the best σ_{max} for ensuring zero gradient only in the main direction. The goal was to find the parameters for which the vesselness score F was greater than zero for all voxels of the target structures. Results are shown in the tables in Figure C.15.

Appendix D. Implementation study of the proposed tubular structures loss functions

Two other empirically studies were done using patches of size $32 \times 64 \times 64$: the first one to find the best implementation for Gaussian kernel application, while the second one to find the best w_{ms} of $MsLoss$. The choice of using small patches was made in order to make training faster and in order to verify the effectiveness of the proposed loss functions even when the ratio of foreground to background voxels is higher. However, even with the use of a method for ROI cropping, the use of such small patches greatly slows down the inference phase and it is not recommended for the implementation of pipelines for the creation of anatomical 3D models. Here, we made use of reference segmentations to extract the ROI where target structures are present. Results are respectively shown in Table D.5 and Table D.6, where the best methods are those later used as presented in the main text. The application of several Gaussian kernels in all

BLOOD VESSELS %(F>0)	σ_{max}				
α, β, γ	2	5	10	20	25
[0.1,0.1,2]	27%	60%	87%	97%	100%
[0.1,0.1,5]	27%	60%	87%	97%	100%
[0.1,0.1,15]	27%	60%	87%	97%	100%
[0.1,0.5,2]	27%	60%	87%	97%	100%
[0.1,0.5,5]	27%	60%	87%	97%	100%
[0.1,0.5,15]	27%	60%	87%	97%	100%
[0.5,0.1,2]	26%	59%	86%	96%	99%
[0.5,0.1,5]	26%	59%	86%	96%	99%
[0.5,0.1,15]	26%	59%	86%	96%	99%
[0.5,0.5,2]	26%	59%	86%	96%	99%
[0.5,0.5,5]	26%	59%	86%	96%	99%
[0.5,0.5,15]	26%	59%	86%	96%	99%

URETERS %(F>0)	σ_{max}		
α, β, γ	5	15	25
[0.05,0.05,2]	41%	92%	100%
[0.1,0.1,2]	40%	91%	100%
[0.5,0.5,2]	39%	90%	99%

Figure C.15: Search of best parameters for σ_{max} of Gaussian filter and α, β and γ for Frangi's vesselness filter: we counted the percentage of voxels of target structures that have a Frangi vesselness greater than 0. The use of σ_{max}=25 and α=0.1, β=0.1 and γ=2 allows values greater than zero for all voxels and thus distinguishing them from having no segmentation or from blob or plate structures. From the tables we can deduce that the most important value to set is σ_{max} to have a strong enough gradient given the larger cross sectional size of some vessels (e.g. aorta and cava vein) and of calyces (attachment of ureters to the kidneys). For the same reason, a smaller value of α also allows us to consider structures with a blob-like shape. As there are no plates or structures with little contrast, the weights of β and γ are irrelevant. Color code: from light to dark green as the percentange is better.

levels of resolution allows for a better extraction of morphological information and thus better comparison between the predicted and reference structures. The first w_{ms} to weight $MsLoss$ was calculated in order to have values around 1, as Dice loss and $Fvloss$ values are between 0 and 1. Due to the fact that $\Delta\Lambda$ for the

baseline is approximately 20, we chose w_{ms} equal to 0.05. This value was then confirmed to be the best suited.

Table D.5: Results using 3D patches of size 32×64×64 for different Gaussian kernel implementations. The values of $\sigma_{i\in[1,5]}$ are in the range [1,σ_{max}+1] with a step of $\frac{\sigma_{max}}{4}$. We used as method for this study the nnU-Net (Isensee et al., 2021) with U-Net as backbone and CE+Soft Dice with deep supervision as baseline loss function. Here w_{ms} is fixed at 0.05, σ_{max} at 25. A: Arteries: V: Veins. Mean and standard deviation of the results are shown.

Vesselness			Structures					
Loss	Deep Sup.	σ_i	S	Dice Score [%] (↑)	Precision [%] (↑)	Recall [%] (↑)	HD95 [mm] (↓)	ΔΛ (↓)
No	-	-	A	73.49 (3.87)	85.15 (10.80)	66.10 (7.21)	15.78 (18.19)	20.05 (2.19)
			V	57.07 (27.75)	84.73 (8.31)	48.79 (27.54)	17.19 (13.93)	20.01 (3.18)
MsLoss	No	[1,5]	A	74.17 (5.49)	83.73 (10.86)	68.11 (8.05)	14.21 (17.59)	19.12 (2.19)
			V	58.55 (29.16)	82.58 (6.47)	53.12 (30.26)	16.02 (12.77)	19.24 (3.57)
MsLoss	Yes	5	A	76.16 (5.16)	86.04 (10.88)	69.81 (7.79)	17.09 (16.86)	18.16 (1.89)
			V	57.78 (27.29)	82.81 (7.85)	51.43 (28.42)	21.15 (17.92)	19.63 (3.54)
MsLoss	Yes	[q+1]	A	74.14 (5.35)	80.78 (11.68)	70.23 (7.08)	14.13 (17.50)	18.71 (1.86)
		as Deep	V	58.93 (26.89)	81.65 (8.12)	52.86 (28.39)	15.01 (11.23)	19.50 (3.44)
MsLoss	Yes	[1,(5-q)]	A	**75.30 (4.92)**	84.42 (11.34)	69.55 (7.51)	**13.49 (17.97)**	**18.83 (2.09)**
		as Deep	V	**60.68 (26.45)**	80.94 (7.89)	55.16 (28.68)	**16.18 (15.31)**	**18.99 (3.56)**
MsLoss + FvLoss	No	[1,5]	A	76.84 (5.69)	80.04 (11.44)	75.60 (7.25)	15.69 (17.73)	17.23 (1.89)
			V	58.71 (27.05)	76.89 (9.16)	55.97 (30.28)	14.77 (10.91)	19.11 (3.26)
MsLoss + FvLoss	Yes	5	A	74.23 (5.60)	84.05 (12.91)	68.31 (7.10)	15.85 (17.53)	19.23 (1.71)
			V	58.41 (26.59)	82.44 (9.79)	50.39 (26.92)	16.63 (18.12)	20.17 (3.06)
MsLoss + FvLoss	Yes	[q+1]	A	75.07 (6.12)	83.96 (13.06)	69.89 (7.67)	**15.66 (17.59)**	18.73 (1.84)
		as Deep	V	58.70 (25.83)	81.47 (11.19)	50.54 (26.26)	**13.62 (11.61)**	20.22 (3.10)
MsLoss + FvLoss	Yes	[1,(5-q)]	A	**77.04 (7.18)**	84.49 (13.47)	72.82 (7.65)	15.77 (17.58)	**17.72 (1.97)**
		as Deep	V	**58.04 (24.35)**	81.30 (10.95)	50.54 (24.47)	15.44 (8.98)	**17.83 (4.65)**

Using the same small patches of size as before we tested the vesselness implementation used by Cao et al. (2010) and Wang et al. (2022) where eigenvalues are ordered by magnitude as in the original vesselness functions. Results are shown in Table D.7. As discussed in Section 4, a direct comparison of vesselness scores leads to worsening of results because two voxels belonging to vessels with different preferential directions (the predicted and reference ones) can have very similar vesselness scores. Furthermore, the Recall analysis using patches of size 32×64×64 is also shown in Figure D.16, which confirms what was stated in Section 6.

Other qualitative results are shown in Figure D.17 from the experiments done using smaller patches of size 32×64×64. Note that the proposed loss function is capable to deal with the strong heterogeneity of veins, and that the segmentation of fine renal arteries is improved.

Table D.6: Results using 3D patches of size 32×64×64 with different w_{ms} to weight $MsLoss$. For these tests we used our proposed method that differs on the oversampling technique used compared to nnU-Net (Isensee et al., 2021). We used U-Net as backbone and CE+Soft Dice+TsLoss with deep supervision as loss function. The proposed vesselness loss functions are applied also with deep supervision with the Gaussian kernel applied with $\sigma_{i\in[1,5-q]}$ from 1 to σ_{max}+1 with a step of $\frac{\sigma_{max}}{4}$, where q is the output resolution level (0 is the output at the same size of the input image). A: Arteries: V: Veins. Mean and standard deviation of the results are shown.

w_{ms}	S	Dice Score [%] (↑)	Precision [%] (↑)	Recall [%] (↑)	HD95 [mm] (↓)	ΔΛ (↓)
0.01	A	77.19 (7.79)	79.56 (13.16)	76.86 (6.62)	15.08 (17.68)	17.07 (1.98)
	V	56.22 (26.92)	81.25 (8.84)	50.47 (29.15)	16.65 (12.79)	19.76 (3.63)
0.05	A	**77.31 (4.42)**	78.10 (12.48)	77.19 (6.64)	**13.83 (17.39)**	17.15 (1.81)
	V	**60.20 (25.15)**	76.38 (9.08)	57.22 (29.36)	**14.58 (11.33)**	19.11 (3.57)
0.1	A	72.79 (7.26)	70.61 (13.06)	77.72 (6.59)	15.69 (17.77)	19.92 (2.27)
	V	58.57 (25.08)	78.92 (8.07)	53.11 (27.85)	16.78 (15.32)	19.99 (3.50)
0.5	A	68.85 (6.83)	62.91 (12.19)	79.03 (6.77)	21.57 (19.77)	17.17 (1.99)
	V	58.87 (23.73)	72.76 (9.19)	55.58 (28.31)	19.36 (26.81)	18.98 (3.53)
1	A	51.72 (8.85)	40.64 (11.42)	75.87 (5.99)	31.72 (18.23)	**16.16 (1.42)**
	V	44.42 (15.31)	34.67 (11.52)	67.57 (26.91)	41.74 (25.13)	**17.06 (3.62)**

Table D.7: Results using 3D patches of size 32×64×64 on combination of different loss functions. For these tests we used our proposed method that differs on the oversampling technique used compared to nnU-Net (Isensee et al., 2021). We used U-Net as backbone and CE+Soft Dice with deep supervision as baseline loss function. Here w_{ms} is fixed at 0.05, the proposed loss functions are applied also with deep supervision with the Gaussian kernel applied with $\sigma_{i\in[1,5-q]}$ from 1 to σ_{max}+1 with a step of $\frac{\sigma_{max}}{4}$, where q is the output resolution level (0 is the output at the same size of the input image). A: Arteries: V: Veins. Mean and standard deviation of the results are given. *The eigenvalues are ordered by magnitude as in the original vesselness functions.

Loss function used	S	Dice Score [%] (↑)	Precision [%] (↑)	Recall [%] (↑)	HD95 [mm] (↓)	ΔΛ (↓)
No	A	74.25 (4.24)	85.80 (11.25)	67.02 (7.45)	13.76 (17.15)	19.65 (2.04)
	V	58.67 (27.06)	82.73 (9.89)	51.10 (27.69)	16.17 (12.86)	19.77 (3.36)
FvLoss* as Frangi SSVMD	A	73.65 (5.16)	86.71 (11.72)	65.53 (7.10)	15.10 (17.03)	19.56 (1.96)
(Cao et al., 2010)	V	58.81 (27.93)	80.41 (9.63)	53.40 (29.25)	**14.42 (10.42)**	19.42 (3.36)
FvLoss* as Jerman SSVMD	A	73.13 (5.56)	85.37 (12.14)	65.63 (7.66)	14.71 (17.24)	20.03 (1.86)
(Wang et al., 2022)	V	56.06 (27.40)	83.33 (8.05)	48.72 (27.74)	15.26 (11.83)	20.23 (3.32)
MsLoss	A	75.63 (4.57)	81.07 (11.64)	72.74 (7.55)	**13.26 (17.54)**	18.21 (1.88)
(Λ in Ms ordered via W)	V	59.73 (26.19)	79.81 (9.65)	53.73 (28.23)	14.71 (12.07)	19.23 (3.25)
MsLoss + FvLoss	A	**77.31 (4.42)**	78.10 (12.48)	77.19 (6.64)	13.83 (17.39)	**17.15 (1.81)**
(Λ in Ms ordered via W)	V	**60.20 (25.15)**	76.38 (9.08)	57.22 (29.36)	14.58 (11.33)	**19.11 (3.57)**

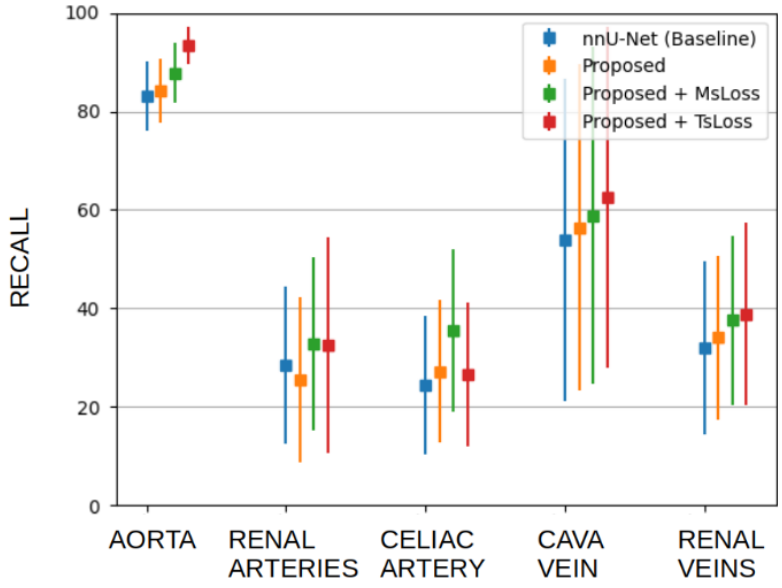


Figure D.16: Recall of the first row and last four rows of experiments in Table D.5 with patches of size 32×64×64 for the different structures. Arteries are divided into aorta, renal arteries and celiac artery, and veins are divided into cava vein and renal veins.

Compliance with Ethical Standards

This research study was conducted retrospectively using human subject data made available by the Necker Hospital (Paris, France). Approval was granted by the Ethics Committee of the Hospital.

Acknowledgments

This work has been partially funded and supported by Region Ile de France (DIM RFSI) and Agence Nationale de la Recherche (ANR-17-CE23-0021). This work was performed using HPC resources from GENCI–IDRIS (Grant 2022-AD011013561). We would like to thank the radiology service of the Necker hospital of Paris, especially Dr. Cécile Lozach, for the help in creating the database presented in this article. We would like to thank also the authors of TopNet (Keshwani et al., 2020) for their help, which allowed us to reproduce their code with sufficient consistency.

Declaration of Competing Interest

The authors have no relevant financial or non-financial interests to disclose.

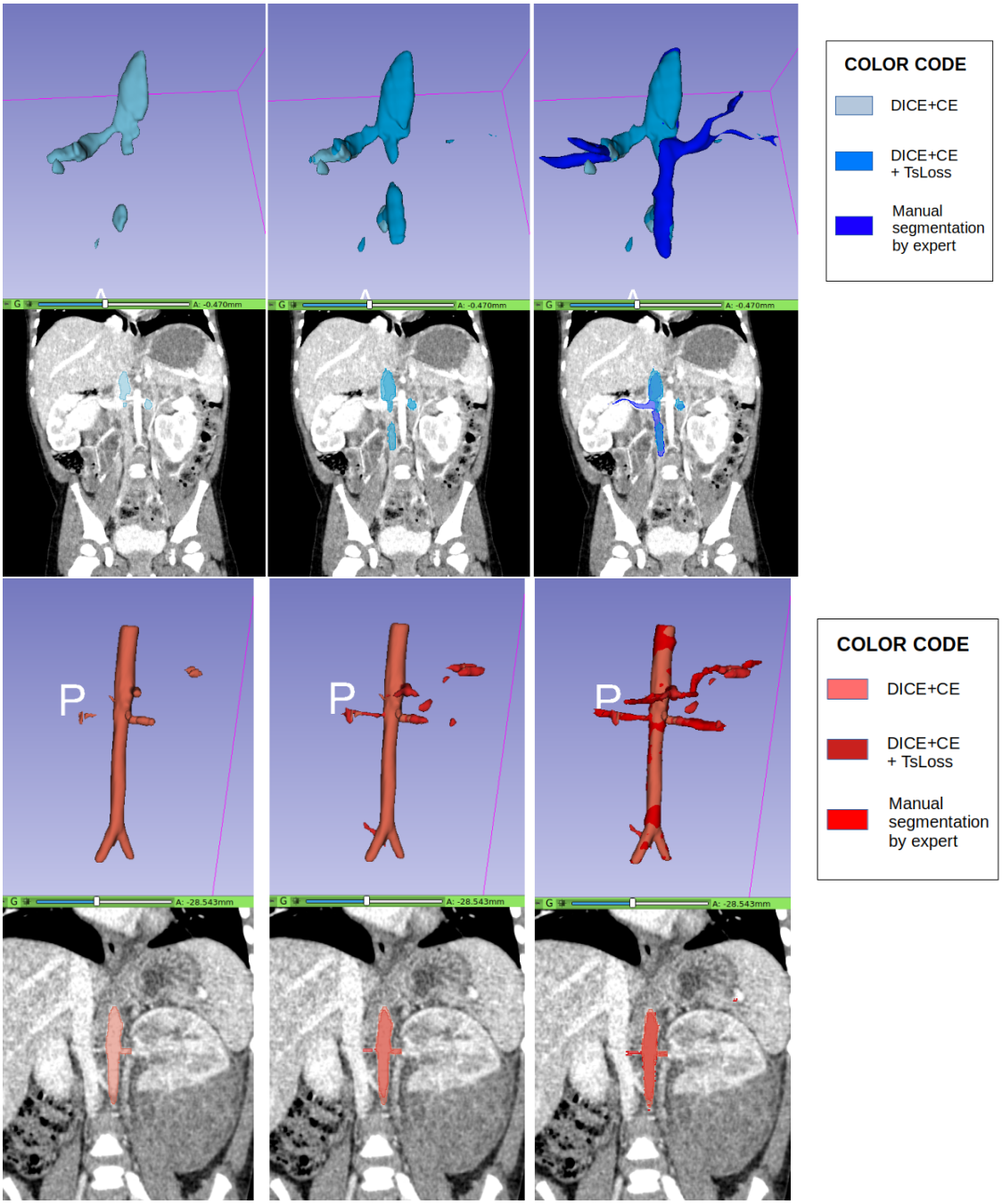


Figure D.17: Example of segmentation for veins (top) and arteries (bottom) on two difficult cases. The top ceCT image presents a strong heterogeneity in the cava vein due to the tumor presence. The bottom ceCT image presents renal arteries with a very few voxels.

References

Amir-Khalili, A., Peyrat, J.M., Abinahed, J., Al-Alao, O., Al-Ansari, A., Hamarneh, G., Abugharbieh, R., 2014. Auto localization and segmentation of occluded vessels in robot-assisted partial nephrectomy, in: Medical Image Computing and Computer Assisted Intervention (MICCAI), pp. 407–414.

Araujo, R.J., Jaime, S.C., Hélder, P.O., 2019. A deep learning design for improving topology coherence in blood vessel segmentation, in: Medical Image Computing and Computer Assisted Interventio (MICCAI), pp. 93–101.

Bauer, C., Bischof, H., 2008. A novel approach for detection of tubular objects and its application to medical image analysis, in: DAGM Symp. Pattern Recognition, pp. 163–172.

Bugajska, K., Skalski, A., Gajda, J., Drewniak, T., 2015. The renal vessel segmentation for facilitation of partial nephrectomy. Signal Processing: Algorithms, Architectures, Arrangements, and Applications (SPA) , 50–55.

Caliva, F., Iriondo, C., Martinez, A.M., Majumdar, S., Pedoia, V., 2019. Distance map loss penalty term for semantic segmentation, in: Medical Imaging with Deep Learning (MIDL).

Cao, K., Du, K., Ding, K., Reinhardt, J., Christensen, G., 2010. Regularized nonrigid registration of lung CT images by preserving tissue volume and vesselness measure, in: Medical Image Analysis for the Clinic: A Grand Challenge, pp. 43–54.

Cehn, L.C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A.L., 2018. DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. IEEE Transactions on Pattern Analysis and Machine Intelligence 40, 834–848.

Chen, M., Cao, K., Zheng, Y., Siochi, R.A.C., 2012. Motion-compensated mega-voltage cone beam CT using the deformation derived directly from 2D projection images. IEEE Transactions on Medical Imaging 32.

Dang, V.N., Galati, F., Cortese, R., Di Giacomo, G., Marconetto, V., Mathur, P., Lekadir, K., Lorenzi, M., Prados, F., Zuluaga, M.A., 2022. Vessel-CAPTCHA: an efficient learning framework for vessel annotation and segmentation. Medical Image Analysis 75, 102263.

Dou, Q., Chen, H., Jin, Y., Yu, L., Qin, J., Heng, P.A., 2016. 3D deeply supervised network for automatic liver segmentation from CT volumes, in: Medical Image Computing and Computer-Assisted Intervention (MICCAI), pp. 149–157.

Du, K., Reinhardt, J.M., Christensen, G.E., Ding, K., E, B.J., 2013. Respiratory effort correction strategies to improve the reproducibility of lung expansion measurements. Medical Physics 40, 123504.

Duchi, J., Hazan, E., Singer, Y., 2011. Adaptive subgradient methods for online learning and stochastic optimization. Machine Learning Research 12, 2121–2159.

Fan, G., Li, J., Li, M., Ye, M., Pei, X., Li, F., Zhu, S., Weiqin, H., Zhou, X., Xie, Y., 2018. Three-dimensional physical model-assisted planning and navigation for laparoscopic partial nephrectomy in patients with endophytic renal tumors. Scientific Reports 8, 582–587.

Fedorov, A., Beichel, R., Kalpathy-Cramer, J., Finet, J., 2012. 3D Slicer as an image computing platform for the quantitative imaging network. Magnetic Resonance Imaging , 1323–1341.

Frangi, R., Niessen, W., Vincken, K., Viergever, M., 1998. Multiscale vessel enhancement filtering, in: Medical Image Computing and Computer-Assisted Intervention (MICCAI), pp. 130–137.

Garcia-Uceda Juarez, A., Selvan, R., Saghir, Z., de Bruijne, M., 2019. A joint 3D UNet-Graph Neural Network-based method for airway segmentation from chest CTs., in: Medical Image Computing and Computer Assisted Intervention (MICCAI) MLMI Challenge, pp. 583–591.

He, Y., Yang, G., Yang, J., Chen, Y., Kong, Y., Wu, J., Tang, L., Zhu, X., Dillenseger, J.L., Shao, P., Zhang, S., Shu, H., Coatrieux, J.L., Li, S., 2020. Dense biased networks with deep priori anatomy and hard region adaptation: Semi-supervised learning for fine renal artery segmentation. Medical Image Analysis 63, 101722.

Heller, N., Isensee, F., Maier-Hein, K., Hou, X., Xie, C., Li, F., Nan, Y., Mu, G., Lin, Z., Han, M., Yao, G., Gao, Y., Zhang, Y., Wang, Y., Hou, F., Yang, J., Xiong, G., Tian, J., Zhong, C., Weight, C., 2021. The state of the art in kidney and kidney tumor segmentation in contrast-enhanced CT imaging: Results of the KiTS19 challenge. Medical Image Analysis 67, 101821.

Heller, N., Isensee, F., Trofimova, D., Tejpaul, R., Papanikolopoulos, N., Weight, C., 2022. Kidney and Kidney Tumor Segmentation: MICCAI 2021 Challenge, KiTS 2021, Held in Conjunction with MIC-

CAI 2021, Strasbourg, France, September 27, 2021, Proceedings. Lecture Notes in Computer Science, Springer International Publishing.

Heryan, K., Choragwicki, D., Sandheim, M., Jakubowski, J., Drewniak, T., 2018. Renal vessels segmentation for preoperative planning in percutaneous nephrolithotomy, in: Imaging Systems and Techniques (IST), pp. 1–6.

Hu, X., Fuxin, L., Samaras, D., Chen, C., 2019. Topology-preserving deep image segmentation, in: Advances in Neural Information Processing Systems (NeurIPS).

Hyde, E., Berger, L., Ramachandran, N., Hughes-Hallett, A., Pavithran, N., Tran, M., Ourselin, S., Bex, A., Mumtaz, F., 2019. Interactive virtual 3D models of renal cancer patient anatomies alter partial nephrectomy surgical planning decisions and increase surgeon confidence compared to volume-rendered images. International Journal for Computer Assisted Radiology and Surgery (IJCARS) 14, 723–732.

Isensee, F., Jaeger, P., Kohl, S., Petersen, J., Maier-Hein, K.H., 2021. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nature Methods 18, 1–9.

Isensee, F., Jäger, P., Wasserthal, J., Zimmerer, D., Petersen, J., Kohl, S., Schock, J., Klein, A., Roß, T., Wirkert, S., Neher, P., Dinkelacker, S., Köhler, G., Maier-Hein, K., 2020. Batchgenerators - a python framework for data augmentation. https://doi.org/10.5281/zenodo.3632567, Zenodo version 0.19.6.

Jerman, T., Pernus, F., Likar, B., Spiclin, Z., 2016. Enhancement of vascular structures in 3D and 2D angiographic images. IEEE Transaction of Medical Imaging 35, 2107–2118.

Jia, J., An, Z., Ming, Y., Guo, Y., Li, W., Li, X., Liang, Y., Guo, D., Tai, J., Chen, G., Jin, Y., Liu, Z., Ni, X., Shi, T., 2017. PedAM: a database for pediatric disease annotation and medicine. Nucleic Acids Research 46, D977–D983.

Kaiser, L., Gomez, A., Shazeer, N., Vaswani, A., Parmar, N., Jones, L., Uszkoreit, J., 2017. One model to learn them all. arXiv preprint arXiv:1706.05137 .

Keshwani, D., Kitamura, Y., Ihara, S., Iizuka, S., Simo-Serra, E., 2020. TopNet: Topology preserving metric learning for vessel tree reconstruction and labelling, in: Medical Image Computing and Computer Assisted Interventio (MICCAI), pp. 14–23.

Kingma, D.P., Ba, J., 2015. Adam: A method for stochastic optimization, in: International Conference on Learning Representations (ICLR).

La Barbera, G., Boussaid, H., Maso, F., Sarnacki, S., Rouet, L., Gori, P., Bloch, I., 2022. Anatomically constrained CT image translation for heterogeneous blood vessel segmentation, in: British Machine Vision Conference (BMVC).

La Barbera, G., Gori, P., Boussaid, H., Belucci, B., Delmonte, A., Goulin, J., Sarnacki, S., Rouet, L., Bloch, I., 2021. Automatic size and pose homogenization with spatial transformer network to improve and accelerate pediatric segmentation, in: IEEE International Symposium on Biomedical Imaging (ISBI), pp. 1773–1776.

Lahlouh, M., Chenoune, Y., Blanc, R., Szewczyk, J., Passat, N., 2022. Aortic arch anatomy characterization from MRA: A CNN-based segmentation approach, in: IEEE International Symposium on Biomedical Imaging (ISBI), pp. 1–5.

Lamy, J., Merveille, O., Kerautret, B., Passat, N., 2022. A benchmark framework for multi-region analysis of vesselness filters. IEEE Transactions on Medical Imaging 41, 3649–3662.

Law, M.W.K., Chung, A.C.S., 2008. Three dimensional curvilinear structure detection using optimally oriented flux, in: European Conference on Computer Vision (ECCV), pp. 368–382.

Lesage, D., Angelini, E., Bloch, I., Funka-Lea, G., 2009. A review of 3D vessel lumen segmentation techniques: Models, features and extraction schemes. Medical Image Analysis 13, 819–845.

Li, J., Lo, P., Taha, A., Wu, H., Zhao, T., 2018. Segmentation of renal structures for image-guided surgery, in: Medical Image Computing and Computer Assisted Interventions (MICCAI), pp. 454–462.

Lorenz, C., Carlsen, I., Buzug, T., Fassnacht, C., Weese, J., 1997. Multi-scale line segmentation with automatic estimation of width, contrast and tangential direction in 2D and 3D medical images, in: CVRMed-MRCAS, pp. 233–242.

Ma, J., Wei, Z., Zhang, Y., Wang, Y., Lv, R., Zhu, C., Chen, G., Liu, J., Peng, C., Wang, L., Wang, Y., Chen, J., 2020. How distance transform maps boost segmentation CNNs: An empirical study, in: Medical Imaging with Deep Learning (MIDL), pp. 479–492.

Magnus, J.R., 1985. Econometric Theory. Cambridge University Press. volume 1. chapter On Differentiating Eigenvalues and Eigenvectors. pp. 179–191.

Marie, F., Corbat, L., Chaussy, Y., Delavelle, T., Henriet, J., Lapayre, J.C., 2019. Segmentation of deformed kidneys and nephroblastoma using Case-Based Reasoning and Convolutional Neural Network. Expert Systems with Applications 127, 282–294.

Merveille, O., Talbot, H., Najman, L., Passat, N., 2018. Curvilinear structure analysis by ranking the orientation responses of path operators. IEEE Transaction on Pattern Analysis and Machine Intelligence 40, 304–317.

Moccia, S., De Momi, E., Mattos, L., El Hadji, S., 2018. Blood vessel segmentation algorithms — Review of methods, datasets and evaluation metrics. Computer Methods and Programs in Biomedicine 158, 71–91.

Porpiglia, F., Checcucci, E., Amparore, D., Piramide, F., Volpi, G., Granato, S., Verri, P., Manfredi, M., Bellin, A., Piazzola, P., Autorino, R., Morra, I., Fiori, C., Mottrie, A., 2020. Three-dimensional augmented reality robot-assisted partial nephrectomy in case of complex tumours (PADUA $\geq$ 10): A new intraoperative tool overcoming the ultrasound guidance. European Urology 78, 229 – 238.

Sun, C., Shrivastava, A., Singh, S., Gupta, A., 2017. Revisiting unreasonable effectiveness of data in deep learning era, in: International Conference on Computer Vision (ICCV), pp. 843–852.

Taha, A., Lo, P., Li, J., Zhao, T., 2018. Kid-Net: Convolution networks for kidney vessels segmentation from CT-volumes, in: Medical Image Computing and Computer-Assisted Intervention (MICCAI), pp. 463–471.

Taha, A.A., Hanbury, A., 2015. Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool. BMC Medical Imaging 15.

Tan, Z., Feng, J., Zhou, J., 2021. SGNet: Structure-aware graph-based network for airway semantic segmentation, in: Medical Image Computing and Computer Assisted Intervention (MICCAI), pp. 153–163.

Virzi, A., Gori, P., Muller, C., Mille, E., Peyrot, Q., Berteloot, L., Boddaert, N., Sarnacki, S., Bloch, I., 2018. Segmentation of Pelvic Vessels in Pediatric MRI Using a Patch-Based Deep Learning Approach, in: Medical Image Computing and Computer-Assisted Intervention (MICCAI) PIPPI Workshop, pp. 97–106.

Wang, C., Hayashi, Y., Oda, M., Itoh, H., Kitasaka, T., Frangi, A., Mori, K., 2019a. Tubular structure segmentation using spatial fully connected network with radial distance loss for 3D medical images, in: Medical Image Computing and Computer Assisted Intervent (MICCAI), pp. 348–356.

Wang, C., Oda, M., Hayashi, Y., Yoshino, Y., Yamamoto, T., Frangi, A., Mori, K., 2019b. Tensor-cut: A tensor-based graph-cut blood vessel segmentation method and its application to renal artery segmentation. Medical Image Analysis 60, 101623.

Wang, D., Pan, Y., Durumeric, O., Reinhardt, J., Hoffman, E., Schroeder, J., Christensen, G., 2022. PLOSL: Population learning followed by one shot learning pulmonary image registration using tissue volume preserving and vesselness constraints. Medical Image Analysis 79, 102434.

Wang, G., Zuluaga, M., Li, W., Aughwane, R., Patel, P., Aertsen, M., Doel, T., David, A., Deprest, J., Ourselin, S., Vercauteren, T., 2017. DeepIGeoS: A deep interactive geodesic framework for medical image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence 41, 1559–1572.

Wang, Y., Wei, X., Liu, F., Chen, J., Zhou, Y., Shen, W., Fishman, E., Yuille, A., 2020. Deep distance transform for tubular structure segmentation in CT scans, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3832–3841.

Yao, L., Jiang, P., Xue, J., Zhan, Y., Wu, D., Zhang, L., Wang, Q., Shi, F., 2021. Graph convolutional network based point cloud for head and neck vessel labeling, in: Medical Image Computing and Computer Assisted Intervention (MICCAI) MLMI Challenge, pp. 474–483.

Zhao, R., Qian, B., Xianli, Z., Li, Y., Wei, R., Liu, Y., Pan, Y., 2020. Rethinking Dice loss for medical image segmentation, in: IEEE International Conference on Data Mining (ICDM), pp. 851–860.