Medical Image Analysis 75 (2022) 102263

Contents lists available at ScienceDirect

Medical Image Analysis

journal homepage: www.elsevier.com/locate/media

Vessel-CAPTCHA: An efficient learning framework for vessel annotation and segmentation

Vien Ngoc Dang[a,b,1], Francesco Galati[a,1], Rosa Cortese[c,d], Giuseppe Di Giacomo[a,e], Viola Marconetto[a,e], Prateek Mathur[a], Karim Lekadir[b], Marco Lorenzi[f], Ferran Prados[g,c,h,i], Maria A. Zuluaga[a,*]

[a] *Data Science Department, EURECOM, Sophia Antipolis, France*
[b] *Artificial Intelligence in Medicine Lab, Facultat de Matemátiques I Informática, Universitat de Barcelona, Spain*
[c] *Queen Square MS Centre, Department of Neuroinflammation, UCL Queen Square Institute of Neurology, Faculty of Brain Sciences, University College London, UK*
[d] *Department of Medicine, Surgery and Neuroscience, University of Siena, Italy*
[e] *Politecnico di Torino, Turin, Italy*
[f] *Université Côte d'Azur, Inria Sophia Antipolis, Epione Research Group, Valbonne, France*
[g] *Centre for Medical Image Computing, Department of Medical Physics and Bioengineering, University College London, UK*
[h] *National Institute for Health Research, University College London Hospitals, Biomedical Research Centre, London, UK*
[i] *e-health Center, Universitat Oberta de Catalunya, Barcelona, Spain*

ARTICLE INFO

Article history:
Received 29 January 2021
Revised 4 October 2021
Accepted 6 October 2021
Available online 13 October 2021

Keywords:
Efficient annotation
Weak supervised learning
Segmentation
Deep learning
Cerebrovascular tree

ABSTRACT

Deep learning techniques for 3D brain vessel image segmentation have not been as successful as in the segmentation of other organs and tissues. This can be explained by two factors. First, deep learning techniques tend to show poor performances at the segmentation of relatively small objects compared to the size of the full image. Second, due to the complexity of vascular trees and the small size of vessels, it is challenging to obtain the amount of annotated training data typically needed by deep learning methods. To address these problems, we propose a novel annotation-efficient deep learning vessel segmentation framework. The framework avoids pixel-wise annotations, only requiring weak patch-level labels to discriminate between vessel and non-vessel 2D patches in the training set, in a setup similar to the CAPTCHAs used to differentiate humans from bots in web applications. The user-provided weak annotations are used for two tasks: (1) to synthesize pixel-wise pseudo-labels for vessels and background in each patch, which are used to train a segmentation network, and (2) to train a classifier network. The classifier network allows to generate additional weak patch labels, further reducing the annotation burden, and it acts as a second opinion for poor quality images. We use this framework for the segmentation of the cerebrovascular tree in Time-of-Flight angiography (TOF) and Susceptibility-Weighted Images (SWI). The results show that the framework achieves state-of-the-art accuracy, while reducing the annotation time by ~77% w.r.t. learning-based segmentation methods using pixel-wise labels for training.

1. Introduction

The segmentation of the 3D brain vessel tree is a crucial task to the diagnosis, management, treatment and intervention of a wide range of conditions with a vast population-level impact (World Health Organization, 2020). Due to the high complexity of the cerebrovascular tree, its automatic extraction is a challenging task. Despite decades of research (Lesage et al., 2009; Moccia et al., 2018), the problem remains open.

With the advent of machine learning and, more precisely, deep learning techniques over the last decade (Litjens et al., 2017; Lundervold and Lundervold, 2019), image segmentation of organs, organs substructures, and lesions has reached state-of-the-art performance. This progress, however, has not been as fast in 3D brain vessel segmentation. Differently from the segmentation of other organs, there is no consolidated deep learning method which has reached human performance, and a vast majority of methods (Bernier et al., 2018; Li et al., 2014; 2019; Morrison et al., 2018)

[*] Corresponding author.
E-mail address: maria.zuluaga@eurecom.fr (M.A. Zuluaga).
[1] Joint first auhtorship.

https://doi.org/10.1016/j.media.2021.102263

still rely on more classical techniques. This lag can be explained by two factors. First, deep learning techniques often assume that the object to segment occupies an important part of the image (Deng et al., 2009; Shelhamer et al., 2017). On the opposite, vessels are relatively small objects within a large image volume (Livne et al., 2019; Tetteh et al., 2020). Secondly, deep learning techniques are well-known for being data greedy, as they require large annotated training datasets to avoid poor generalization. Due to the complexity of vascular trees and the small size of vessels, it is challenging to obtain sufficiently large high-quality annotated sets.

This work presents a novel framework to address the challenges faced by deep learning-based 3D vessel segmentation. Taking inspiration from Completely Automated Public Turing Test To Tell Computers and Humans Apart, better known as CAPTCHA (von Ahn and Dabbish, 2004), we initially divide the image volume into 2D image patches and we subsequently request the user to identify the patches containing a vessel or part of it. This task is common on websites to differentiate humans from bots, using image CAPTCHAs (von Ahn and Dabbish, 2004; Elson et al., 2007) of natural images. This procedure, which we denote Vessel-CAPTCHA, simplifies the annotation process by requiring 2D patch tags indicating the presence of a vessel (a part of it, or multiple vessels) and, thus, avoiding pixel-wise annotations. The user-provided patch tags are subsequently used to synthesize a pixel-wise pseudo-labeled training set in a self-supervised manner using a clustering technique. These two sets are used to train the framework.

The proposed framework is composed of two networks: a segmentation network and a classification network. The segmentation network extracts vessels on a patch basis to tackle the limitations of deep nets in the segmentation of small objects. The final volumetric segmentation is obtained by concatenating the 2D segmented patches. The classification network is used for two tasks. First, it allows to enlarge the labeled data without the need for further user-provided annotations. Second, it may act as a second opinion (Leibig et al., 2017; Vrugt and Robinson, 2007) that provides a measure of uncertainty in low quality or complex images. We evaluate the role of the classification network as an expert opinion, where only the segmentations from patches identified as vessel patches are kept and those classified as non-vessel patches are masked out.

1.1. Related work

1.1.1. 3D brain vessel segmentation

A comprehensive collection of methods and techniques for general vascular image segmentation is reviewed in Lesage et al. (2009); Moccia et al. (2018), where they classify different segmentation frameworks according to their characteristic strategies. Classical approaches typically rely on hand-crafted features, with image intensity-derived (Taher et al., 2020), and first (Law and Chung, 2008), second (Frangi et al., 1998; Sato et al., 1997) or higher order (Cetin and Unal, 2015) tensor-derived features among the most common. Feature extraction is followed by a vessel extraction scheme, which performs the final segmentation. Notable extraction schemes include deformable models (Klepaczko et al., 2016; Zhao et al., 2015), voting (Zuluaga et al., 2014b), tracking algorithms (Rempfler et al., 2015; Robben et al., 2016) and statistical approaches (Hassouna et al., 2006). Their main drawbacks are two. First, these methods rely on hand-crafted features that need to be tuned, requiring high expertise to find a good set of parameters. Second, extraction schemes are not fully automatic: many need manual initialization, and the final results typically call for manual correction, specially when images are noisy.

Deep learning techniques have emerged as an alternative to circumvent the difficulties of classical approaches. Existing methods have tried to explicitly address the brain vessel tree complexity by designing shallow convolutional neural networks (CNNs) architectures to avoid possible over-fitting (Phellan et al., 2017), or by partitioning the input image volume, while still relying on deeper and more powerful architectures (Kamnitsas et al., 2017; Ronneberger et al., 2015). Different partitioning strategies include anatomical regions (Kandil et al., 2018), 2D slices (Ni et al., 2020), 3D (Phellan et al., 2017; Tetteh et al., 2020) and 2D patches (Livne et al., 2019). Despite achieving accuracies similar to those of classical approaches, the main limitation towards the broader use of deep learning techniques remains to be the burden linked to pixel-wise data annotation, including multi-plane annotations (Phellan et al., 2017) or further pre-processing (Phellan et al., 2017; Kandil et al., 2018; Livne et al., 2019).

Patch-based approaches (Livne et al., 2019; Tetteh et al., 2020) not only aim at reducing the vessel tree's complexity, but they also try to mitigate the limitations of neural nets in the segmentation of objects occupying small portions of an image. Our work adopts a similar strategy and it builds upon the advantages of 2D patch-based approaches (Livne et al., 2019), thus making vessels cover a significant portion of the patch, while avoiding pixel-wise annotations.

1.1.2. Limited supervision for image segmentation

Different strategies have been explored as an alternative to pixel-wise annotation (Cheplygina et al., 2019; Ørting et al., 2020; Tajbakhsh et al., 2020), a tedious and time consuming task requiring a high level of expertise. These strategies can be roughly classified, according to the type of labels they use, as partial pixel-wise labels, which include incomplete, sparse or noisy pixel-wise labels (Tajbakhsh et al., 2020); or as weak labels, which refer to high-level labels and drawing primitives (Cheplygina et al., 2019).

Partial pixel-wise labels refer to annotations where only a fraction of the pixels of the object of interest are provided (Bai et al., 2018; Çiçek et al., 2016; Liang et al., 2019; Ke et al., 2020). These labels can be provided by the user or generated by simpler methods to produce rough segmentation masks. Semi-supervised methods follow different strategies to exploit partially labeled data under the assumption that it is enough to train a segmentation model. Bai et al. (2018) used image registration to propagate user-provided labels over some image slices containing the aorta. Çiçek et al. (2016) designed the 3D-Unet to account for sparse and incomplete pixel-wise labels. Other methods resort to iterative stages of refinement (Liang et al., 2019; Ke et al., 2020). Although these methods have reported good performances in medical image segmentation (Cheplygina et al., 2019), the complexity of the 3D brain vessel tree makes pixel-wise annotation, even if partial, highly time consuming. As one of our aims is to minimize the annotation effort, our work focuses on the use of weak labels.

1.1.3. Weakly supervised learning

Weak labels for medical image segmentation We consider two forms of weak labels for medical image segmentation tasks: image-level labels and drawing primitives. Image-level labels (Feng et al., 2017; Jia et al., 2017; Raza et al., 2019; Schlegl et al., 2015; Xu et al., 2019; Zhao et al., 2019) assign a tag or rating to an image under the assumption that images contain cluttered scenes with enough information from which a model can learn (Qi et al., 2017). In medical tasks, they have been mainly used with 2D images/slices to segment pathologies, i.e. lung nodules (Feng et al., 2017), damaged retinal tissue (Schlegl et al., 2015), brain tumors (Izadyyazdanabadi et al., 2018) or cancerous tissue (Jia et al., 2017; Kraus et al., 2016; Lerousseau et al., 2020; Xu et al., 2014; 2019). To a lesser extent they have been used for organ structures segmentation, i.e. the optic disc (Zhao et al., 2019). Despite the good reported performances and the annotation time savings they represent, image tags have not been used for 3D vessel segmentation.

Drawing primitives include bounding boxes and contouring shapes (Cheplygina et al., 2016; Gao et al., 2012; Dai et al., 2015; Li et al., 2018; Rajchl et al., 2017; Wang et al., 2018), scribbles and lines (Can et al., 2018; Lin et al., 2016; Matuszewski and Sintorn, 2018; Wang et al., 2015) and clicks (Bruggemann et al., 2018). In 3D vessel segmentation, bounding boxes have been used for aortic segmentation, with the assumption that the aorta is a compact structure, which can be enclosed within a bounding box (Pepe et al., 2020). This assumption does not hold for highly sparse bifurcated trees, as the brain vascular tree, where a 3D bounding box would nearly cover the full brain. Moreover, if an image is analyzed in 2D, the vessel tree appears as a series of disconnected blobs or elongated structures, which challenges the use of 2D contouring shapes. Koziński et al. (2020) address this limitation by using 2D annotations in Maximum Intensity Projections of 3D vascular images. To some extent, these can be considered 2D image scribbles of varying density for the original 3D volume. The framework, however, requires full 2D pixel-wise annotations. Although the scheme significantly reduces the labeling time, more than four hours are needed to generate sufficiently dense 2D annotations that do not compromise performance. Finally, clicks are common in classical 3D vessel segmentation approaches (Benmansour and Cohen, 2009; Moriconi et al., 2019) to provide seed-points, but no works yet integrate them in a weakly supervised learning framework. This may be due to the complexity of the 3D brain vessel tree, where a single click might not carry sufficient information to train a model.

Our work relies on image tags. To cope with the granularity and sparse appearance of vessels, we use 2D patch-level tags, in the form of clicks over a grid. A click selects the patches containing at least one vessel or a part of it. We denote this annotation scheme the Vessel-CAPTCHA.

Weakly supervised learning with image tags Our weakly supervised vessel segmentation framework using image tags can be cast as a multi-instance learning (MIL) problem (Dietterich et al., 1997; Maron and Lozano-Pérez, 1997; Cheplygina et al., 2019), where a bag corresponds to an image patch and the instances are the image pixels. A bag is considered positive (a vessel patch) if at least one instance within the bag is positive (a vessel pixel). The goal is then to infer the key instances (Liu et al., 2012), i.e. the vessel pixels, that activate the bag label.

Standard MIL segmentation approaches, which have been less studied than the classification counterpart (Campanella et al., 2019; Hou et al., 2016; Quellec et al., 2012), follow a multi-stage strategy. In a first stage common to MIL segmentation and classification, they train a model to learn instance-level probabilities of belonging to the positive class. At a second stage, these probabilities are used to obtain pixel-wise labels, which can be considered as the segmentation output (Xu et al., 2014; Kraus et al., 2016) or as pseudo-labels to train a segmentation model in supervised way (Lerousseau et al., 2020; Xu et al., 2019). A main limitation is that the instance-level probabilities are not originally conceived to generate segmentations, but to serve as inputs for bag classification. Therefore, the segmentation results may be poor. Mitigation strategies rely on area constraints (Jia et al., 2017; Lerousseau et al., 2020); robust instance selection operations (Kraus et al., 2016; Xu et al., 2019); post-processing (Kraus et al., 2016); or enriched information, such as supplementary instance-level inputs (Shin et al., 2019) or image landmarks (Schlegl et al., 2015). However, these strategies often come at the cost of further required user inputs (Jia et al., 2017; Schlegl et al., 2015; Shin et al., 2019).

Attention-based MIL (Ilse et al., 2018), an alternative to standard MIL, uses attention mechanisms (Niu et al., 2021), such as class activation maps (CAM) (Zhou et al., 2016), under the assumption that the discriminative regions identified by a network correspond to the key instances, i.e. the pixels to segment (Ahn and Kwak, 2018; Feng et al., 2017; Hong et al., 2017; Izadyyazdanabadi et al., 2018; Ouyang et al., 2019; Shen et al., 2021; Zhao et al., 2019). Since attention mechanisms focus on the localization of the most discriminative regions, they suffer from the same limitations as standard MIL, which lead to inaccurate segmentation masks. For instance, some works (Shen et al., 2021) consider the resulting mask as a localization/detection mask and not as a segmentation one. Others have attempted to refine the attention maps through pixel similarity propagation (Ahn and Kwak, 2018; Zhao et al., 2019), feature assembling (Izadyyazdanabadi et al., 2018) and post-processing stages (Krähenbühl and Koltun, 2011), which all lead to increased model complexity. To avoid the increased complexity, other works propose manual intervention (Feng et al., 2017) or the use of some pixel-wise annotated data (Ouyang et al., 2019; Zou et al., 2021), leading to more user-required inputs.

A last set of methods favors the use of simpler techniques to generate an initial pseudo-labeled set that can be then refined using a learning-based approach. Luo et al. (2020) relied on traditional saliency methods along with a quality control step for object detection from videos. Hou et al. (2016) used a mixture of Gaussians in cancer tissue classification. Lu et al. (2021) used a simple threshold to segment tissue regions, which are refined with a CAM to classify cancerous tissue.

While the cerebrovascular tree is a highly complex structure, the typical available dataset size for training a model to segment it is relatively small. Therefore, avoiding high model complexity is critical in 3D brain vessel segmentation (Phellan et al., 2017). Our work favors simplicity and minimal user interaction. Thus, similarly to (Hou et al., 2016; Luo et al., 2020; Lu et al., 2021), we use a simpler self-supervised technique, such as the K-means, to generate pixel-wise pseudo-labels. As other weakly supervised approaches (Feng et al., 2017; Lerousseau et al., 2020; Luo et al., 2020; Xu et al., 2019), we use the pseudo-labeled set as input of a supervised training phase that learns to segment the brain vessel tree, without the need for any additional user inputs.

1.1.4. Biomedical image classification

Our work explores the use of the Unet (Ronneberger et al., 2015) and the Pnet (Wang et al., 2019), two networks originally conceived for medical image segmentation, for the classification tasks of our framework. These two networks have been originally designed for image segmentation. Their adaptation to a classification task can be considered as a MIL formulation, where instance-level information, i.e. pixels, are used to predict a bag label, i.e. the patch tag. Similar to most biomedical classification tasks, previous MIL-based biomedical image classification works (Campanella et al., 2019; Qi et al., 2017) rely on customized versions of VGG-16 (Simonyan and Zisserman, 2015) and ResNet (He et al., 2016), the most popular architectures for natural image classification. Others (Hou et al., 2016) use task-specific architectures adapted from general purpose networks such as end-to-end CNNs. However, no major performance differences are currently found among them (Lundervold and Lundervold, 2019).

1.2. Contributions

The contributions of this work are four-fold:

1. we introduce an annotation and segmentation scheme, the Vessel-CAPTCHA, to reduce the labeling burden of 3D brain vascular images, consisting of two phases: a first phase where the user provides tags at the 2D image patch-level, and a second stage where pixel-wise pseudo-labels are obtained, in a self-supervised fashion, using only the user-provided patch tags as input.

2. We propose a weakly supervised learning framework on 2D image patches to achieve 3D brain vessel segmentation. To circumvent the problems faced by deep neural networks when segmenting small objects, the framework uses a 2D patch-based segmentation network trained with 2D pixel-wise pseudo-labeled patches synthesized by the Vessel-CAPTCHA annotation scheme using the weak user-provided patch tags as input.
3. We investigate the use of network architectures specifically designed for medical imaging tasks to classify 2D image patches (vessel vs. non-vessel). The classifier networks are used to pseudo-label a potential training set without further user effort, and it may act as a second opinion for segmentation masks obtained from low quality images.
4. Using two different image modalities, we demonstrate that the proposed framework achieves state-of-the-art performance for 3D brain vessel segmentation, while significantly reducing the annotation burden by ~77% compared to the annotation time required in other deep learning-based methods.

To foster reproducibility and encourage other researchers to build upon our results, the source code of our framework is publicly available on a Github repository.[2]

2. Method

The proposed Vessel-CAPTCHA framework algorithm for 3D vessel segmentation is depicted in Fig. 1. In the following, we introduce the Vessel-CAPTCHA annotation scheme and we describe how pixel-wise pseudo-labels are synthesized from the user-provided weak patch labels in a self-supervised way (Section 2.1). In Section 2.2, we present the two networks conforming the proposed framework: a classifier network and a segmentation network. Section 2.3 explains how the classifier network can be used to enlarge the set of weak pixel-wise annotations, allowing to have a larger set to train 2D-WnetSeg. Finally, Section 2.4 briefly explains how to segment unseen images using the proposed framework.

2.1. The vessel-CAPTCHA annotation scheme

We consider a dataset $\mathcal{I}$ of training images. Given an image $\mathbf{I} \in \mathcal{I}$ of size $H \times W \times S$, for each slice X_s, $s \in [1, \ldots, S]$, we consider a partition in P_s non-overlapping patches: $\mathcal{X}_s = \{\hat{X}_k\}_{k=1}^{P_s}$. Each patch is here considered as a function $\hat{X}_k : D_k \to \mathbb{R}$, where D_k is a subset of the slice domain $D_k \subset [1, H] \times [1, W]$.

User annotations on a given patch $\hat{X}_k$ are defined through a function $U_k : D_k \to \{0, 1\}$, assigning a binary label to each coordinate $(i, j) \in D_k$. The set of annotations for a given patch is summarized by an indicator function $f : U_k \to \{0, 1\}$ which takes value 1 if at least one pixel in the patch was labeled with 1:

$$f(U_k) = 1 \iff \exists (i, j) \in D_k \, s.t. \, U_k(i, j) = 1. \tag{1}$$

Fig. 2 illustrates examples of equivalent user annotations. The set of indicators for the slice X_s is denoted by $\mathcal{Y}_s = \{f(U_k)\}_{k=1}^{P_s}$. The training set of patch-level labels for the image $\mathbf{I}$ is defined by the set: $\mathcal{T}_P^{\mathbf{I}} = \{\mathcal{X}_s, \mathcal{Y}_s\}_{s=1}^{S}$. This set is therefore composed by patches and associated indicators/tags of the presence of a vessel according to the user's annotation. Based on the training set $\mathcal{T}_P^{\mathbf{I}}$, we estimate vessel pseudo-labeled masks via a model fitting procedure. For every patch we define a function $M_k : D_k \to \{0, 1\}$, which assigns to each pixel's coordinate a label according to the following scheme:

$$M_k(i, j) = \begin{cases} 0 & \text{if } f(U_k) = 0, \\ KM(\hat{X}_k(i, j)) & \text{otherwise}, \end{cases} \tag{2}$$

where KM is a K-means predictor trained on the intensity values of the patch $\{\hat{X}_k(i, j), \ (i, j) \in D_k\}$. By specifying $K = 2$ clusters we therefore obtain a rough estimate of the low-high intensity partitioning of the patch. The ensemble of estimated partitions across patches is denoted as $\mathcal{M}_s = \{M_k\}_{k=1}^{P_s}$, and we define the pixel-wise pseudo-labeled training set for the image $\mathbf{I}$ as $\mathcal{T}_M^{\mathbf{I}} = \{\mathcal{X}_s, \mathcal{M}_s\}_{s=1}^{S}$.

Finally, for the full image training set $\mathcal{I}$, the user-provided patch-level set and the pixel-wise pseudo-labeled one are denoted by

$$\mathcal{T}_P = \{\mathcal{T}_P^{\mathbf{I}}\}_{\mathbf{I} \in \mathcal{I}}, \tag{3}$$

and

$$\mathcal{T}_M = \{\mathcal{T}_M^{\mathbf{I}}\}_{\mathbf{I} \in \mathcal{I}}, \tag{4}$$

respectively.

2.2. Image segmentation and patch classification networks

2.2.1. Segmentation network

The segmentation network learns from the input training set $\mathcal{T}_M$ how to segment 2D image patches using the Dice similarity coefficient, as proposed by Milletari et al. (2016), which is specifically tailored for segmentation tasks in medical images. The segmented 2D patches are concatenated to reconstruct the original segmented 3D image volume. For this task, we use a segmentation network connecting two 2D-Unets in cascade (Dias et al., 2019). We denote it 2D-WnetSeg (Fig. 3). The network is trained on $\mathcal{T}_M$, the set of 2D image patches with pixel-wise pseudo-labels to tackle the neural networks limitations in the segmentation of objects with a small object-to-image ratio.

The human cerebrovascular system has an intricate shape with large and smaller blood vessels which mainly differ in the spatial scale, but which share similar shapes. The selected self-supervised method, the K-means, favors over-segmentation of larger vessels. Thanks to a set of max pooling layers, the first 2D-Unet allows to learn spatial scaling features from the input training data. Thus, it can recover rough-mask labels from smaller vessels not initially extracted by K-means. This means that the first Unet acts as a refinement module to correct the initial masks by inferring missing vessels based on the structural redundancy of the cerebrovascular tree. The second Unet, with a similar architecture as the first one, receives as input the output of the first Unet with the recovered labels from small vessels. As a result, the 2D-WnetSeg learns vessels even with a pseudo-labeled training set with imperfect labels or noise.

The smaller vessels in the brain vessel tree may disappear in very deep networks due to the subsampling layers. To tackle this, the 2D-WnetSeg has 14 blocks with convolutional layers structured into 4 levels. In this, it differs from previously proposed cascaded networks (Dias et al., 2019) or the Unet-based vessel segmentation from (Livne et al., 2019). This also contributes to reduce the number of trainable parameters. Specifically, the number of trainable parameters in Livne et al. (2019) is about 3.1e7, whereas the WnetSeg has only about 1.6e7 parameters.

In our architecture, the first 7 blocks form the first Unet and the second 7 blocks belong to the second one. Each block consists of 2 convolutional layers with kernel size 3 × 3 pixels, each followed by a rectified linear unit (ReLU). They are both added to the padding to ensure that the output has the same shape as the input. A dropout layer is applied between them. As the input proceeds through different levels along the contracting path, its resolution is reduced by half. This is performed through a 2 × 2 max-pooling operation with stride 2 on 3 levels except for the bottom level. We double the number of feature channels at each level of the contracting path. The right portion of a half-network (Unet), i.e. the expansive path, consists of blocks with concatenation and

[2] https://github.com/robustml-eurecom/Vessel-Captcha

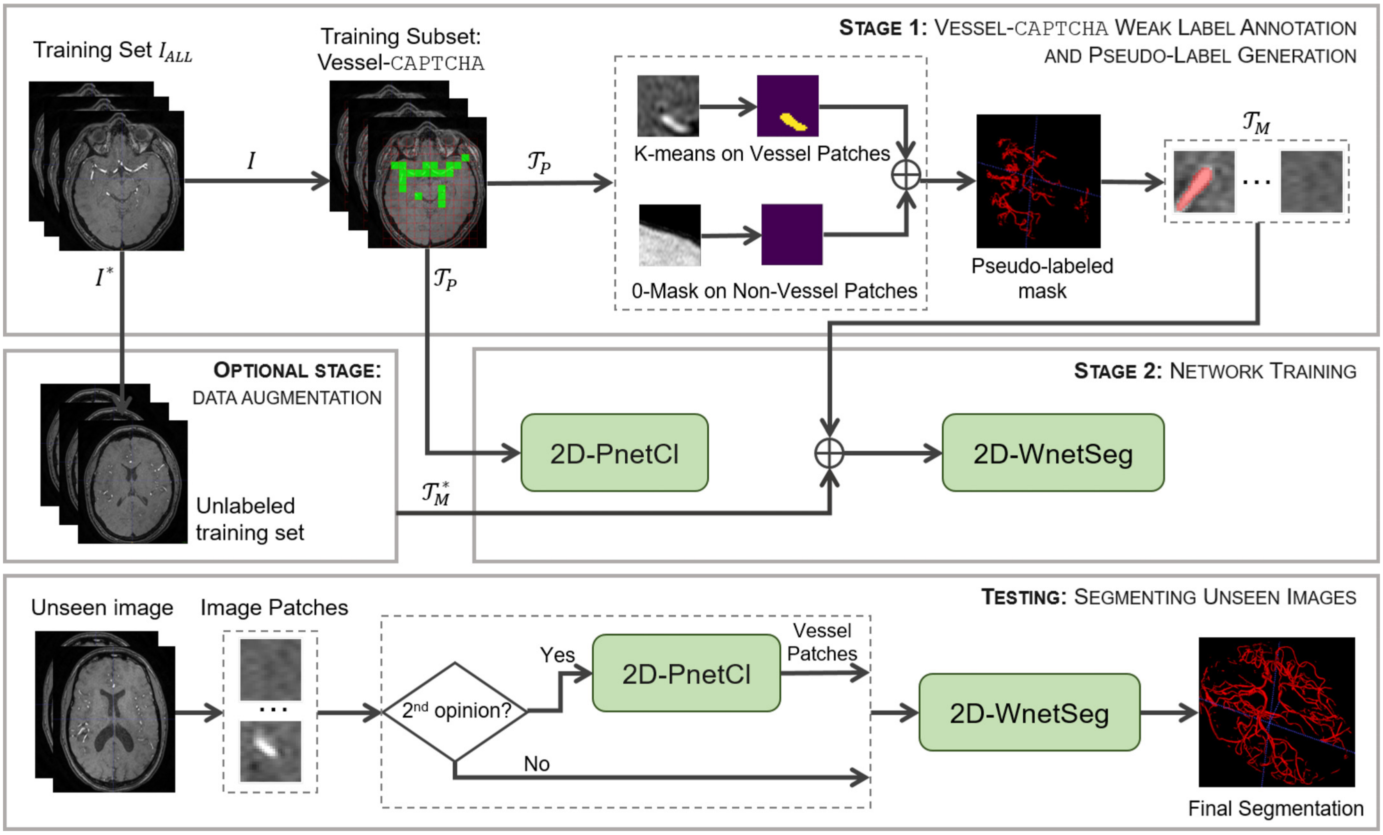
Training Set I_{ALL}
Training Subset: Vessel-CAPTCHA
STAGE 1: VESSEL-CAPTCHA WEAK LABEL ANNOTATION AND PSEUDO-LABEL GENERATION
I
$\mathcal{T}_P$
K-means on Vessel Patches
0-Mask on Non-Vessel Patches
Pseudo-labeled mask
$\mathcal{T}_M$
I^*
$\mathcal{T}_P$
OPTIONAL STAGE: DATA AUGMENTATION
Unlabeled training set
$\mathcal{T}_M^*$
STAGE 2: NETWORK TRAINING
2D-PnetCl
2D-WnetSeg
TESTING: SEGMENTING UNSEEN IMAGES
Unseen image
Image Patches
2nd opinion?
Yes
No
2D-PnetCl
Vessel Patches
2D-WnetSeg
Final Segmentation

Fig. 1. The Vessel-CAPTCHA framework. At Stage 1, an image grid with patch size 32 × 32 covering the brain tissue is presented to the user for annotation. The user selects the patches which contain at least one vessel or a part of it. The process, which we denote the Vessel-CAPTCHA annotation scheme, is done for every axial slice in an image volume. This weakly annotated set $\mathcal{T}_P$ is used to synthetize pixel-wise pseudo-labels for every patch using the K-means algorithm. The resulting pseudo-labeled set is denoted $\mathcal{T}_M$. At stage 2, $\mathcal{T}_P$ is used to train a classification network (2D-PnetCl) and $\mathcal{T}_M$ is used to train a segmentation network (2D-WnetSeg). In the segmentation network training, it is possible to enlarge the set of pseudo-labeled data through an optional data augmentation step. For an unseen image, the final volumetric segmentation is obtained by concatenating the 2D segmentations obtained from 2D-WnetSeg. Optionally, the classification network can be used as a second opinion to refine the segmentation results. In that case, only 2D segmentations from patches classified as vessel ones are considered in the final volume segmentation.

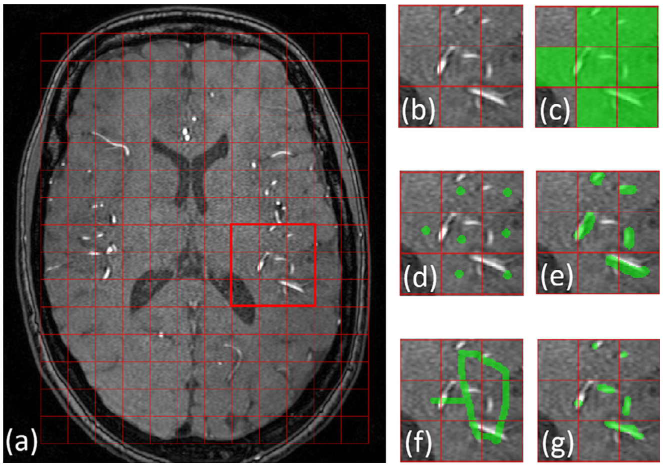

Fig. 2. Example of equivalent CAPTCHA annotations. (a) Image slice $\mathcal{X}_s$ with patch grid, (b) zoomed region corresponding to the highlighted red box in (a), (c) resulting $\mathcal{T}_P$ obtained through equivalent annotations (d-g). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

up-sampling for each level to extract low-features and it expands the spatial support of the lower resolution feature maps to assemble the necessary information and recover the original input size. Finally, we employ skip-connections from the shallow layers to deeper layers between the two 2D-Unets, at the same levels, to ease the training of the network.

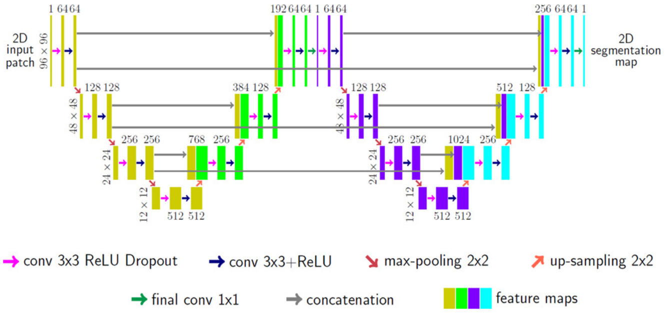

Fig. 3. Illustration of the 2D-WnetSeg architecture.

2.2.2. Networks for vessel vs. non-vessel patch classification

The classification network is trained on $\mathcal{T}_P$ to discriminate between vessel and non-vessel patches in unseen data. This discrimination serves two purposes: (1) to synthesize patch tags without the need of user interventions and (2) to act as a second opinion for segmentations. In the latter case, the segmentation network serves as a first expert predicting pixel-wise labels, whereas the classifier network provides a concept on a per-patch basis. This can be considered an ensemble approach to uncertainty (Vrugt and Robinson, 2007), where a disagreement among the two net-

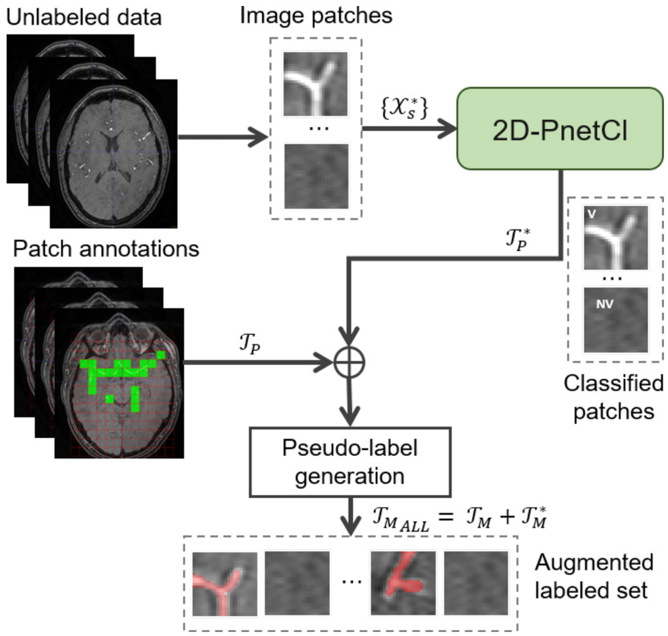

Fig. 4. Data Augmentation procedure. The trained classifier is used as the starting point to enlarge the initial pixel-wise labeled training set $\mathcal{T}_M$ without requiring further user inputs. The resulting training set $\mathcal{T}_{M_{ALL}}$ is a combination of both the pseudo-labels and those obtained via the Vessel-CAPTCHA annotation.

works/opinions indicates uncertainty on the predictions of a given patch.

Most works in the literature rely on customized versions of VGG-16 (Simonyan and Zisserman, 2015) and ResNet (He et al., 2016), the most popular architectures for natural image classification, or on task-specific architectures adapted from general purpose networks (Chen et al., 2016; Setio et al., 2016). In this work, we investigate the use of networks specifically designed for medical imaging applications for our classification task: the Unet (Ronneberger et al., 2015) and the Pnet (Wang et al., 2019). As these two networks have been designed for image segmentation, we hereby describe how they have been modified to achieve classification.

We denote the modified 2D Pnet architecture (Wang et al., 2019) 2D-PnetCl. It consists of 7 convolution layers, 2 dropout layers, and a sigmoid layer. The first 5 convolution layers are concatenated. Each convolutional layer contains 64 filters with 3×3 pixels receptive fields in a 1 pixel stride sliding with different dilation factors. The dilations are 1, 2, 4, 8 and 16, respectively. The last two convolutional layers are the 1×1 convolutions, the output feature map is flattened and fed to a fully connected layer for interpretation with 128 hidden units and the final prediction layer uses a sigmoid function with one unit to classify patches with and without vessels. The adapted 2D-Unet architecture, denoted 2D-UnetCl, uses the network from (Livne et al., 2019) as a starting point. Similarly to the 2D-PnetCl, the output feature map is flattened and fed to a fully connected layer for interpretation with 128 hidden units and a final prediction layer with one unit to classify patches with and without vessels.

2.3. Data augmentation for segmentation network training

The set $\mathcal{T}_M$ consisting of pseudo-labels is used to train the 2D-WnetSeg. To augment its size without increasing the annotation burden, we make use of the classification network to generate a larger set with pixel-wise pseudo-labels. The procedure is depicted in Fig. 4.

Assuming that there is an initial set of unlabeled images I^* that can be used for training, we consider the joint image dataset of labeled and unlabeled images $\mathcal{I}_{ALL} = \mathcal{I} \bigcup \mathcal{I}^*$. The subset $\mathcal{I}$ of these images is used to generate Vessel-CAPTCHAs, which are presented to the user for annotation. This results in the training set $\mathcal{T}_P$ (Eq. (3)), which is used to both train the classification network and to synthesize the pixel-wise pseudo-labeled set $\mathcal{T}_M$ (Eq. (4)).

Using the trained classification network, a set of patches $\{\mathcal{X}_s^*\}$ is obtained in the remaining set of images $\mathcal{I}^*$. Rather than presenting another Vessel-CAPTCHA to the user for annotation, the $\{\mathcal{X}_s^*\}$ are inputted to the classification network to estimate patch labels $\{\mathcal{Y}_s^*\}$. The paired set of patches and estimated labels conform a new set $\mathcal{T}_P^* = \{\mathcal{T}_P^{\mathbf{I}}\}_{\mathbf{I} \in \mathcal{I}^*}$.

The set $\mathcal{T}_P^*$ is used to synthesize pixel-wise pseudo-label masks $\mathcal{M}^*$ following the same procedure applied to $\mathcal{T}_P$ (Section 2.1). This leads to a new pseudo-labeled set $\mathcal{T}_M^*$. The extended set of pixel-wise pseudo-labels is formed by the union of the two sets $\mathcal{T}_{M_{ALL}} = \mathcal{T}_M \bigcup \mathcal{T}_M^*$, and is subsequently used to train the 2D-WnetSeg architecture.

2.4. Inference phase

Unseen 3D images are segmented by extracting 2D image patches that are then segmented by the 2D-WnetSeg and concatenated to build back the original volume (Fig. 1). In low quality or noisy images, the resulting segmentation can often present a large set of pixels erroneously segmented as vessels. To avoid this problem, the trained classifier network may act as an expert providing a second opinion to the results from the segmentation network. In such case, only those patches which have been classified as vessels are taken into account to reconstruct the final volume. All the pixels of the remaining patches are set to zero.

2.5. Implementation details

We used the Keras library to implement 2D-PnetCl, 2D-UnetCl and 2D-WnetSeg. The networks were trained on a GPU workstation with 4-core Intel(R) Xeon(R) CPU @ 2.30 GHz, a NVIDIA Tesla P100-PCIE-16 GB, and 25 GB memory. For both 2D-UnetCl and 2D-PnetCl we optimized the binary cross-entropy loss function with a minibatch stochastic gradient descent and a conservative learning rate of 0.01 and momentum of 0.9. The weights of the 2D-WnetSet were optimized using an Adam optimizer with learning rate $lr = 1e{-}4$, $\beta_1 = 0.9$, and $\beta_2 = 0.999$. All networks were trained from scratch using mini-batches of 64 patches. All input patches were normalized by the mean and standard deviation of the whole training data. A dropout of 0.5 for 2D-PnetCl and 2D-UnetCl, and of 0.1 for 2D-WnetSeg was added to prevent overfitting during the training. For 2D-PnetCl, the dropout is applied before and after the second to last convolutional layer. For 2D-UnetCl and 2D-WnetSeg, the droput is applied after a convolutional layer and the ReLU (Fig. 3). The image input sizes of 2D-PnetCl and 2D-WnetSeg were 32×32 and 96×96, respectively. We implemented a zero-padding technique to preserve output size as input size at each convolution layer in both networks. Therefore, the feature map size at each level in the 2D-PnetCl is 32×32.

3. Experimental setup

In this section, we describe the experimental setup. First, we present the datasets used in our experiments (Section 3.1) and the baselines used for comparison (Section 3.2). Then, we describe the training setup (Section 3.3). Finally, we present the performance evaluation metrics used in our experiments (Section 3.4).

Table 1
Main properties of data used and training and validation test sizes per data type.

	Synthetic	TOF	SWI
Dataset size	136	100	33
Volume dimensions	325 × 304 × 600	560 × 560 × 117 (Set 1) 576 × 768 × 232 (Set 2)	480 × 480 × 288
Voxel spacing	1 × 1 × 1 mm^3	1 × 1 × 1 mm^3 (Set 1) 0.3 × 0.3 × 0.6 mm^3(Set 2)	1 × 1 × 1 mm^3
$\lvert\mathcal{T}_P\rvert$ (patch size 32 × 32)	7.18 M	770 K	30.6 K
$\lvert\mathcal{T}_M\rvert$ (patch size 96 × 96)	1.04 M	110 K	10.2 K

3.1. Data

Three different types of data were used in this study: synthetic, Time-of-Flight (TOF) angiography and Susceptibility-Weighted Images (SWI). The latter two correspond to two magnetic resonance imaging (MRI) sequences commonly used to image and assess the cerebrovascular tree (Radbruch et al., 2013), although blood vessels present different appearances in each modality. In TOF, vessels are hyper-intense structures, whereas they are hypo-intense in SWI. Table 1 summarizes the main properties of each data type and the datasets used.

Synthetic data We use the synthetic data generated and made public by Tetteh et al. (2020).[3] The dataset consists of 136 volumes of size 325 × 304 × 600 with corresponding labels for vessel segmentation, which were generated following the method proposed in Schneider et al. (2012). The vessel labels occupy 2.1% of total intensities, highlighting the problem of vessels being relatively small objects within a large image volume.

TOF data We use 100 TOF scans coming from two different sources. Forty-two TOF subject scans, from retrospective studies previously conducted at the UCL Queen Square Institute of Neurology, were available with volume dimensions 560 × 560 × 117 and isotropic voxel size 1 × 1 × 1 mm^3 (Set 1). The remaining 68 scans were obtained from the OASIS-3 database (LaMontagne et al., 2019) with volume dimensions 576 × 768 × 232 and voxel size 0.3 × 0.3 × 0.6 mm^3 (Set 2).

SWI data We use 33 different subject scans with image dimensions 480 × 480 × 288 and isotropic image resolution 1 × 1 × 1 mm^3, from retrospective studies previously conducted at the UCL Queen Square Institute of Neurology, Queen Square MS Centre, University College London. Due to poor image quality, three SWI scans were discarded for the experiments.

3.2. Baselines

We compare our segmentation framework with several alternatives, including state-of-the art deep learning-based vessel segmentation (Livne et al., 2019; Tetteh et al., 2020) and classical approaches (Frangi et al., 1998; Sato et al., 1997; Zuluaga et al., 2014b), and weakly supervised learning frameworks (Ahn and Kwak, 2018; Lerousseau et al., 2020). Specifically, we evaluate:

1. **Classical 3D Vessel Segmentation Methods:** We consider three classical non-learning based approaches, which use the 3D image volume as input. These are: the Frangi filter (Frangi et al., 1998) (**Frangi**) and the Sato filter (Sato et al., 1997) (**Sato**), two references for vessel segmentation, and a tensor voting framework for 3D brain vessel segmentation (Zuluaga et al., 2014b) (**TV**).
2. **Deep Leaning-based 3D Vessel Segmentation Methods:** We consider the deep learning-based brain vessel segmentation framework from Livne et al. (2019) (**Vessel 2D-Unet**), which relies on the 2D-Unet (Ronneberger et al., 2015) as backbone architecture, and uses 2D patches as input; and **DeepVesselNet**, the framework from Tetteh et al. (2020), which uses the 3D image volume as input, but operates on 3D patches using a fully convolutional architecture to extract the 3D vessel tree.
3. **Weakly Supervised Methods:** We compare our weakly supervised strategy with one standard MIL and a CAM-based approach. Concretely, we use a MIL framework for whole slice (**WS-MIL**) histopathology segmentation (Lerousseau et al., 2020) and the CAM-based approach proposed by Ahn and Kwak (2018) for natural image segmentation (**AffinityNet**). Both methods work with 2D image patches with size 32 × 32 and 96 × 96, respectively.
4. **Other Limited Supervision Strategies:** We consider two semi-supervised strategies using partial labels: the **3D-Unet**, which can be trained using sparsely annotated training data (Çiçek et al., 2016), and a **Pseudo-labeling** strategy, where we use rough masks as labels. The label masks are generated with the Sato filter (Sato et al., 1997) and they are used to train a 2D-Unet network with 2D image slices.

We compare the classification networks, 2D-PnetCl and 2D-UnetCl, with two baselines, **VGG-16** (Simonyan and Zisserman, 2015) and **ResNet** (He et al., 2016), as they are among the most common networks for classification (Litjens et al., 2017). Table 2 summarizes the hyperparameter setup for every baseline network.

3.3. Setup

Pre-processing and annotation We used the available ground truth from the synthetic images to generate Vessel-CAPTCHA annotations. Since the in-plane dimensions of the images are not a multiple of the patch size (Table 1), we overlap the last two rows/columns of patches.

Both TOF and SWI were skull-stripped using a standard tool and we generated the Vessel-CAPTCHA annotation grid only over the brain tissue (Fig. 2). Where the minimum-sized rectangle mask covering the brain tissue was not a multiple of the patch size in a given dimension, we dilated the mask in that dimension until the condition was met and generate the annotation grid. If the minimum-sized rectangle mask touched the image slice borders and the in-plane dimensions of the images were not a multiple of the patch size, we generated the annotation grid by overlapping the last two rows or columns of patches. Three users annotated the images using the Vessel-CAPTCHA annotation scheme: a trainee, an experienced rater and a neurologist. In addition to this, TOF data was pixel-wise annotated. Finally, no pixel-wise labels were obtained for SWI, since it is difficult to obtain a sufficiently robust ground truth. All annotation times were recorded.

For the Vessel 2D-Unet, further data pre-processing for synthetic and TOF data was performed as described in Livne et al. (2019). All datasets where normalized (within modality). For TOF, where two different sources were used, we follow the intensity and spacing normalization strategy from (Full et al., 2021).

[3] https://github.com/giesekow/deepvesselnet/wiki/Datasets

Table 2
Hyper-parameter setup for baseline networks.

Network	Hyper-parameters
Vessel 2D-Unet	batch size: 64, lr: 1e−4, dropout: 0.0
DeepVesselNet	batch size: 10, lr: 1e−3, decay: 0.99, cube size: 64
WS-MIL	batch size: 100, lr: 1e−4, decay: 10e−5, $c_0 = c_1$=1, $\alpha = [1\text{e}-2, \ldots, 0.1]$, β=[0.9, …, 0.99]
AffinityNet	batch size: 16, lr: 1e−1
3D-Unet	lr: 1e−4, reduced by 0.5 every 10 epochs. Stopped at 50 epochs if no improvements in the validation error
VGG-16	batch size: 64, lr: 1e−4
ResNet	batch size: 64, lr: 1e−3

Training setup Table 1 displays the number of available 2D patches for training and validation per dataset. For every dataset, we performed data splitting at the image volume level, using a split ratio 70/10/20% for training, validation and testing, respectively. The training sets were augmented through the use of different random rotations, flips and shears at every epoch for every 2D patch. Models are chosen based on the best performance in the validation set.

Two different rules are used to synthesize pseudo-labels for the annotated training set $\mathcal{T}_M$ with the K-means algorithm. In synthetic data and TOF, vessels are associated to the cluster with the highest mean value, whereas the vessel class is associated to the cluster with the lowest mean value in SWI. The training sets, $\mathcal{T}_P$ and $\mathcal{T}_M$, are used to separately train a classification and a segmentation network per modality.

3.4. Evaluation metrics

Vessel segmentation We estimate the Dice Similarity Coefficient (DSC), the Hausdorff Distance (HD), the 95% Hausdorff Distance (95HD) and the mean surface distance error (μD) between the segmentation and the annotated ground truth to quantitatively assess the segmentation accuracy in TOF and the synthetic dataset. We measure HD, 95HD and μD in voxels.

In SWI, the segmentations are assessed qualitatively. Based on a visual inspection by two raters (an expert rater and a neurologist), the segmented images are classified as good (3), average (2) or low quality (1). A segmented image is considered good, if it segments the large and medium vessels, and avoids the segmentation of noisy regions, with an elongated appearance similar to a vessel, and sulci. It might miss some small vessels. A segmented image is considered of average quality if it segments large and medium vessels, it misses small ones, it may segment noisy areas in a small proportion (less than 50%), specially in the anterior part of the brain, and often segments sulci. All other cases are considered as low quality ones. We use the Cohen's Kappa coefficient (κ) to measure the level of agreement among raters.

Patch classification We measured precision (P), recall (R) and the F-score (F_1), using a vessel patch as the positive class to assess the quality of the classification results obtained by the classifier networks.

4. Experiments and results

We assess the performance of the Vessel-CAPTCHA in terms of vessel segmentation accuracy and required annotation time (Section 4.1). In Section 4.2, we compare our weak learning strategy with other limited supervision techniques. Section 4.3 studies the proposed classification networks and their performance as a data augmentation strategy. Next, we perform an ablation study to understand how the different components of the framework contribute to performance (Section 4.4) and we present a brief summary of all the obtained results in Section 4.5.

4.1. 3D brain vessel segmentation performance

We evaluate the performance of the Vessel-CAPTCHA framework in terms of segmentation accuracy and required annotation time using all available datasets. We compare it against the 3D brain vessel segmentation, i.e. the deep learning vessel segmentation frameworks and the classical techniques.

Synthetic data We use the synthetic data to provide a controlled setup, where the ground truth is fully reliable, to assess the learning-based vessel segmentation strategies. In addition to the required fully supervised training, Vessel 2D-Unet and DeepVesselNet are trained using weak labels from the Vessel-CAPTCHA annotation scheme.

Fig. 5 summarizes the segmentation accuracy results from the different networks. The Vessel 2D-Unet and DeepVesselNet present the best performances when they are trained using fully labeled and reliable ground truth data. DeepVesselNet reports a minor drop in performance (1 and 2%) w.r.t. the values reported in Tetteh et al. (2020), which we consider related to implementation details. As it could be expected, the Vessel-CAPTCHA has a slightly lower performance than Vessel 2D-Unet and DeepVesselNet trained with full precision labels. However, it surpasses the performance of both architectures trained with weak labels, indicating that Vessel-CAPTCHA is better suited for the weak learning setup.

TOF images We use real clinical data from the TOF images to evaluate the Vessel-CAPTCHA and to compare it against the 3D vessel segmentation baselines in terms of segmentation accuracy and training set annotation time.

Among classical 3D vessel segmentation methods, the Frangi (Frangi et al., 1998) and Sato (Sato et al., 1997) filters produce real-valued maps that need to be thresholded to get a binary segmentation. The TV (Zuluaga et al., 2014b) provides a probability map, which may produce small spurious segmentations that need to be filtered out. The three methods allow to identify vessels at different spatial resolutions. In our experiments, we set 10 scales in the range [0.5,2] mm. We obtain final binary segmentations for the classical methods in two ways:

1. **No post-processing (NP):** the real-valued masks obtained with the Frangi and Sato filter are normalized to the range [0,1]. We set a fixed threshold ($t > 0.6$) to binarize the three maps, and we do no filter out potential small spurious objects.
2. **Post-processing (PP):** Every (real-valued and probability) map is inspected by overlaying it on the original testing image, to define and apply a per-image threshold. The resulting binary maps are filtered by masking out any connected component with a size equal or smaller than 4. Through visual inspection of every binary segmentation overlaid in the original image, the minimum connected component size could be modified. Where the results are yet not satisfactory, the base method can be rerun using a different set of scales, followed by a new round of post-processing operations. We record the time required to obtain a visually satisfactory segmentation.

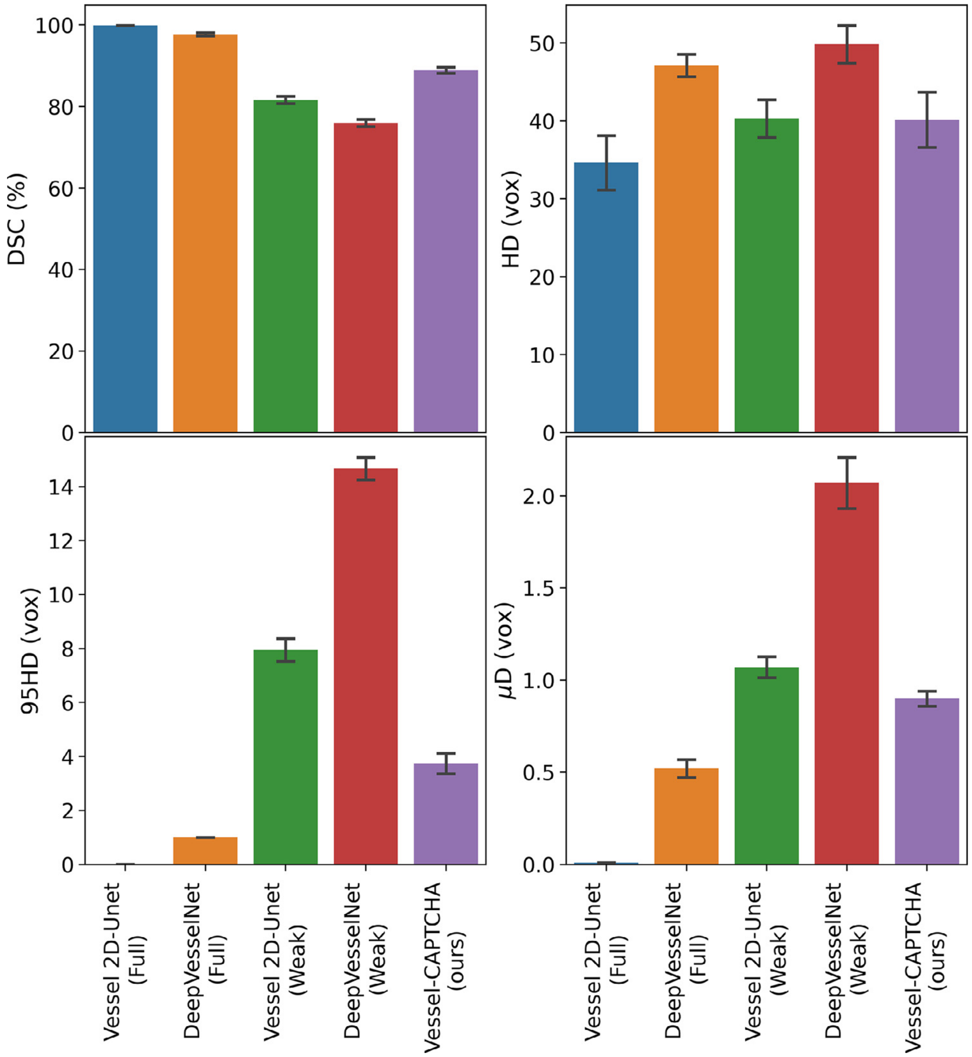

Fig. 5. Segmentation performance in synthetic data. Vessel 2D-Unet and DeepVesselNet are trained with full pixel-wise annotations (Full) and with weak labels (Weak). A higher value is better for DSC, lower is better for HD, 95HD and μD, indicatin that our Vessel-CAPTCHA is the best method among the weakly supervised ones.

Table 3
3D brain vessel segmentation methods accuracy in TOF. The bold font denotes best value, with underlined values not significantly different from it ($\alpha = 0.05$). Classical methods and DeepVesselNet use 3D volumes as input. Vessel 2D-Unet and our framework use 2D patches as inputs. HD, 95HD and μD are reported in voxels.

	Method	DSC (↑)	HD (↓)	95HD (↓)	μD (↓)
NL	Frangi-NP	54.16 ± 8.81	81.04 ± 18.48	14.78 ± 13.83	2.47 ± 2.22
	Sato-NP	55.75 ± 7.15	78.60 ± 16.37	11.53 ± 12.01	2.17 ± 1.07
	TV-NP	68.41 ± 5.01	60.23 ± 10.08	10.97 ± 11.72	2.10 ± 1.00
	Frangi-PP	68.44 ± 3.15	20.60 ± 10.91	9.01 ± 10.38	2.36 ± 2.01
	Sato-PP	69.01 ± 3.67	21.53 ± 9.11	8.86 ± 10.09	2.10 ± 1.01
	TV-PP	70.74 ± 3.38	**20.11 ± 8.45**	8.31 ± 8.23	2.07 ± 1.02
FS	Vessel 2D-Unet	77.66 ± 4.32	74.78 ± 16.73	12.60 ± 18.16	0.60 ± 0.11
	DeepVesselNet	76.13 ± 5.51	75.32 ± 12.94	4.32 ± 1.16	1.65 ± 0.26
	Vessel-CAPTCHA (ours)	**79.32 ± 3.02**	51.70 ± 5.92	**4.06 ± 1.50**	**0.50 ± 0.09**

NL, No labels; FS, Fully supervised; NP, No post-processing; PP, Post-processing.

Table 3 summarizes the segmentation performance. Classical vessel segmentation methods show a poor performance when no manual post-processing is done. This is expected, as it is a well-known limitation of such approaches. The manual post-processing step allows an important jump in performance. In particular, it allows to remove spurious and disconnected false positives, which is reflected on their low HD, the best among all methods, and an important drop of the 95HD, while maintaining μD relatively constant. However, post-processing requires high level of expertise and it is time consuming.

With the exception of the HD, learning-based methods consistently show a better performance across measures, with no statistical differences among them, and the Vessel-CAPTCHA reporting the best results among all methods. This demonstrates that the proposed framework can reach state-of-the-art performance despite the use of less accurate annotations (Fig. 6). We bring at-

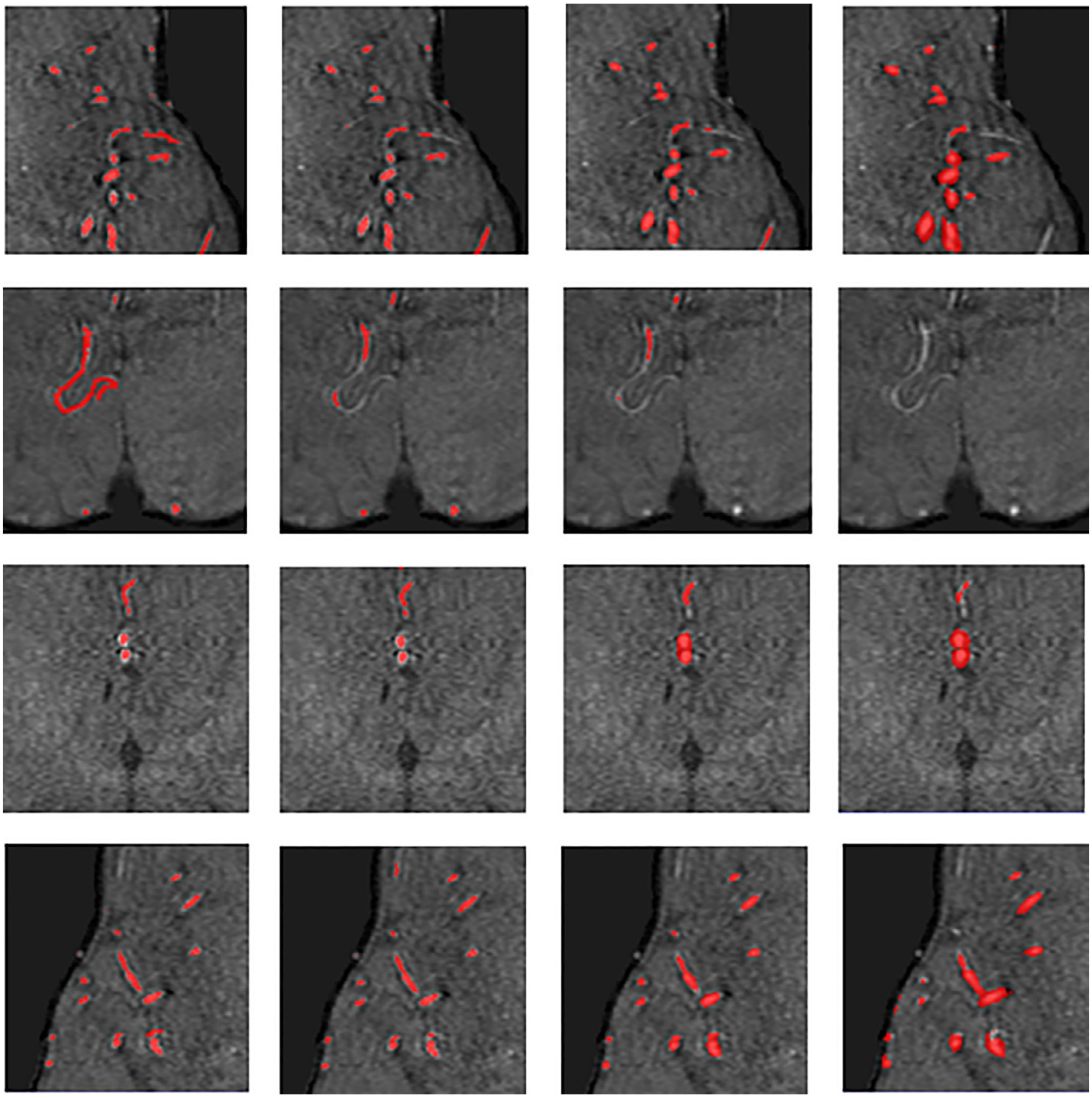

Fig. 6. Segmentation results in TOF images. From left to right: ground truth, Vessel-CAPTCHA (ours), Vessel 2D-Unet and DeepVesselNet.

tention to the fact that Vessel 2D-Unet and DeepVesselNet report lower DSC (77.66 vs. 89.0 and 76.13 vs. 81.0, respectively) than the reported in Livne et al. (2019); Tetteh et al. (2020). However, for Vessel 2D-Unet our results show a better 95HD (12.6 vs. 47.27) and a comparable sub-voxel μD (0.60 vs. 0.38). The better distance-based measures suggest that the differences in the DSC might come from the ground truth annotation protocol, in which our data might include more distal, hence thinner vessels that are more prone to be unsegmented. This is confirmed by DeepVesselNet's DSC on synthetic data. In the controlled setup, the reported results are comparable to (Tetteh et al., 2020).

Fig. 7 presents segmentation accuracy measured with the DSC as a function of the required average user intervention time per image. For the proposed framework, the user intervention time corresponds to the average time required to obtain weak labels using the Vessel-CAPTCHA annotation scheme. We report the average from the time measurements from the three raters (75.5 ± 12.5 min). For 2D Vessel-Unet and DeepVesselNet, the user intervention time corresponds to the average time to fully pixel-wise annotate TOF images (327.5 ± 20.5 min). The 2D-Unet framework (Livne et al., 2019) requires additional data pre-processing to obtain patches with vessels located at the center of the patch, which is not considered in the reported numbers. While this operation could represent a further increase in the time needed to prepare the training set, we consider it marginal in comparison with the time required to do the pixel-wise annotation. Finally, for the classical methods, the user intervention time corresponds to the average time required to segment and post-process one image. We observe that, on average, the Vessel-CAPTCHA reduces the annotation time by 77%, w.r.t. pixel-wise annotations in the same image, while achieving a higher segmentation accuracy.

Susceptibility-weighted images (SWI) We study the capacity of the Vessel-CAPTCHA to segment different image modalities by qualitatively assessing the segmentation results obtained in SWI. The framework was trained and visually assessed on the validation set. The model visually judged as best was used to segment the test set.

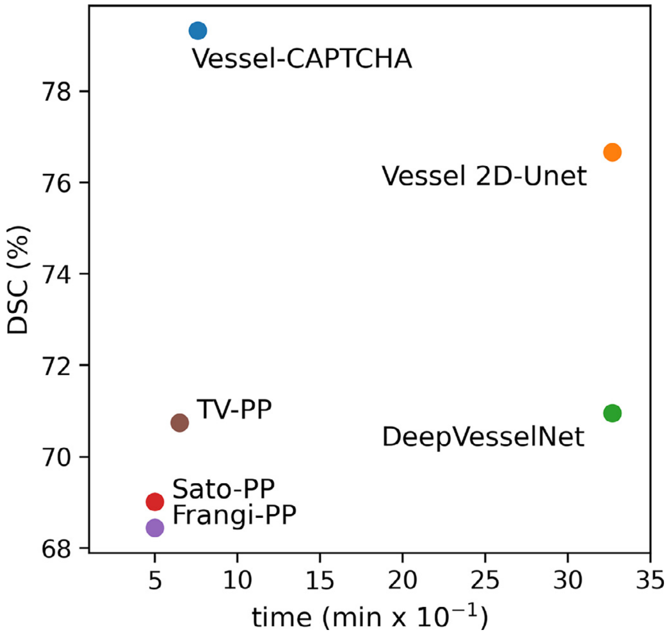

Fig. 7. Segmentation accuracy (DSC) vs. User intervention time.

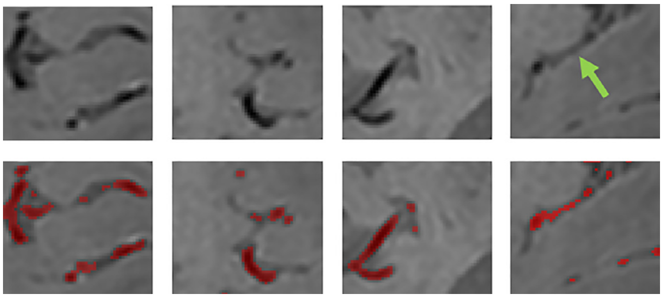

Fig. 8. Segmentation results in SWI images. Top: Original image. Bottom: Overlaid segmentation. From left to right the first three cases present good segmentation results. The rightmost example shows a sulci that has been segmented as if it was a vessel (green arrow). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Fig. 8 illustrates some segmentation results. Overall, SWI is more complex than TOF, thus further errors are observed. As a general pattern, the SWI segmentations tend to miss small vessels, while there is also a high incidence of false positives due to erroneously segmented sulci and noise. Nevertheless, the raters judged more that 50% of the segmentations as good and only one image was considered poor by one of them. Their visual judgment an average rating score of 2.57 with an agreement $\kappa = 0.75$.

SWI Vessel-CAPTCHA annotation requires 38% more time than in TOF (94.5 ± 11.5). This is expected given the increased complexity of SWI scans: small vessels require more effort to be identified and vessels often present an appearance similar to sulci (Fig. 8). These factors have a direct incidence in the time needed by a rater to discriminate vessel from non-vessel patches. Nevertheless, SWI Vessel-CAPTCHA accounts for 71% less time than the pixel-wise annotation baseline (327.5 ± 20.5 min, see Fig. 7).

4.2. Alternative limited supervision strategies

Using the TOF dataset, we choose to do a separate comparison of the Vessel-CAPTCHA and other limited supervision strategies, which excludes fully supervised 3D brain vessel segmentation approaches. As there are no works using limited supervision addressing 3D brain vessel segmentation we consider that a direct comparison between the two families of methods (i.e. limited vs. full supervision) is advantageous towards the fully supervised techniques.

Table 4
Comparison with partial labeling methods using TOF images. The bold font denotes best value. Our framework uses 2D patches, Pseudo-labeling uses image slices and 3D-Unet image volumes as input.

	3D-Unet	Pseudo-labeling	Vessel-CAPTCHA (ours)
DSC (↑)	68.50 ± 3.37	54.99 ± 5.86	$\mathbf{79.32 \pm 3.02}$
HD (↓)	76.12 ± 8.47	68.50 ± 9.58	$\mathbf{51.70 \pm 5.92}$
95HD (↓)	15.72 ± 2.23	24.19 ± 5.25	$\mathbf{4.06 \pm 1.50}$
μD (↓)	2.56 ± 1.44	4.48 ± 1.67	$\mathbf{0.50 \pm 0.09}$

Partial labeling techniques Table 4 compares our framework with the partial labeling techniques, 3D-Unet, and Pseudo-labeling. The 3D-Unet is trained with the pixel-wise annotations. Given that 3D pixel-wise vessel annotations are highly prone to error, given the difficulties that the brain vessel tree poses, the resulting annotated dataset is likely to present missing labels (i.e. sparsity), which the 3D-Unet handles seamlessly. Pseudo-labeling uses rough segmentation masks obtained using the Sato filter (Sato et al., 1997) to the image volumes, thus avoiding user annotations. Despite being designed to handle sparse pixel-wise annotations and being the only method directly processing the image volume, the 3D-Unet does not achieve the best performance. The results are lower than those reported by other frameworks requiring precise pixel-wise annotations, i.e. Vessel 2D-Unet and DeepVesselNet (Table 3). These results are consistent with other works in the literature (Livne et al., 2019; Koziński et al., 2020; Ni et al., 2020; Phellan et al., 2017; Tetteh et al., 2020), which avoid the use of end-to-end 3D networks and favor the use of networks relying on smaller input spaces, e.g. 3D subvolumes (Phellan et al., 2017; Tetteh et al., 2020), 2D images (Koziński et al., 2020; Ni et al., 2020) or patches (Livne et al., 2019). Pseudo-labeling results suggest that, in isolation, this approach cannot reach a good accuracy, which explains why it is often coupled with a refinement stage (Liang et al., 2019; Ke et al., 2020).

Weakly supervised strategies In our experiments, we were not able to achieve sufficiently good results with WS-MIL and AffinityNet that could allow a quantitative comparison with the other baselines. In this section, we perform a qualitative analysis of the obtained results to gain understanding about the limitations of standard MIL- and CAM-based segmentation techniques for brain vessel tree segmentation.

We adapt WS-MIL to address 3D brain vessel segmentation by using the Vessel-CAPTCHA patches as input rather than an image slice (Lerousseau et al., 2020). WS-MIL splits its input into sub-patches and it ranks them according to their predicted probability of containing a vessel. We consider two sub-patch sizes, 16×16 and 8×8. The final sub-patch labeling is achieved by using the ranked patches along with two hyper-parameters, α and β, which control the minimum number of pixels belonging to the foreground (α) and the background class (β) (Table 2). We observe two limitations in the obtained results (Fig. 9). First, the resulting masks correspond to vessel localization masks, not segmentations, due to the granularity of the patches. The original WS-MIL formulation (Lerousseau et al., 2020) has been conceived for super resolution histology images, where the resulting labeled sub-patches can be considered a segmentation mask. Standard brain images have a much lower resolution. Therefore, the final result lacks the necessary specificity to be considered a segmentation. Second, we observe that it is difficult to set a value for α and β that works well for all the slices in an image volume. As shown in Fig. 9, while a low α value works well in image slices with larger vessels, the

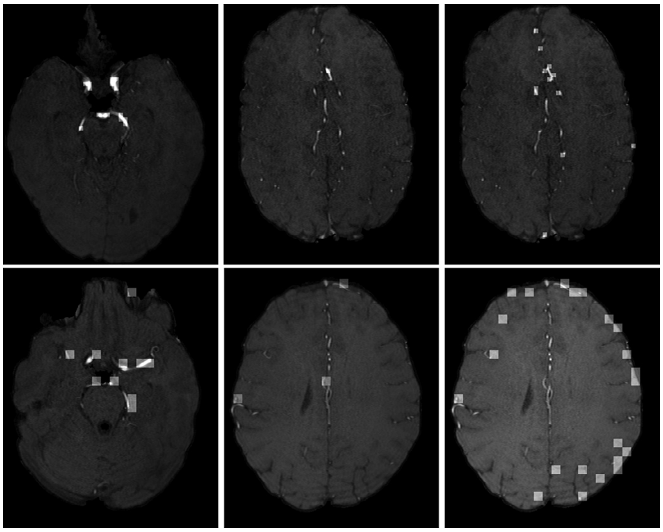

Fig. 9. Vessel localization results with WS-MIL using sub-patch resolution 8×8 (top) and 16×16 (bottom). The first two columns use $\alpha = 0.01$, $\beta = 0.99$. The rightmost column uses $\alpha = 0.07$, $\beta = 0.93$ on the middle column images.

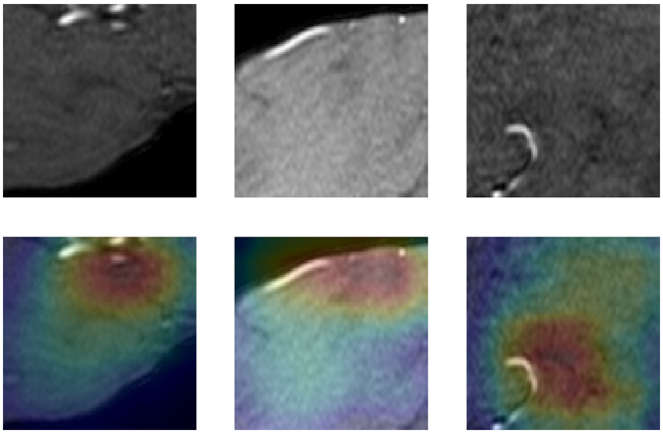

Fig. 10. Vessel patches of size 96×96 (top) with overlaid CAMs (bottom) from the AffinityNet framework.

same value fails to detect smaller vessels, hence it is necessary to train a new model with different α, β values.

The architecture of AffinityNet does not allow images below a certain size to be fed into it. Therefore, we had to enlarge the patch used from 32×32 to 96×96, similar to the one we use as input of 2D-WnetSeg. The larger patches were obtained by grouping 32×32 patches. A vessel label was assigned if at least one sub-patch was originally labeled as a vessel patch. Otherwise, the patch was labeled as non-vessel.

Despite the larger field of view of the new input patches, our experiments did not achieve good results with AffinityNet. A visual inspection of the CAMs showed that, although they activate consequently with the class associated to the patch, these did not contain discriminative information about vessels (Fig. 10). Let us recall that AffinityNet (Ahn and Kwak, 2018) uses the input image and the CAMs (Zhou et al., 2016) to synthesize pseudo-labels, which are then used to train a segmentation model. However, CAMs are rough approximations of the object of interest (Ahn and Kwak, 2018; Bae et al., 2020; Zou et al., 2021). In the past, CAM-based methods have been used to segment relatively large objects in natural scenes (Ahn and Kwak, 2018; Hong et al., 2017; Zou et al., 2021), damaged tissue (Izadyyazdanabadi et al., 2018) or blob-like structures occupying an important part of the image, such as the optic disc (Zhao et al., 2019). In our case, as vessels are relatively small objects, it seems that the network requires to use much more information from the scene to discriminate between vessel and non-vessel patches, as reflected by the CAMs (Fig. 10). The information, however, is to broad to locate the vessels and thus AffinityNet fails.

4.3. Classification networks

Classification networks performance We study the performance of the two classification networks, 2D-UnetCl and 2D-PnetCl, to determine if they are well-suited as discriminators within our framework. Table 5 compares the classification performance of 2D-UnetCl and 2D-PnetCl in TOF and SWI images with VGG-16 and the ResNet. For each network, two models were trained, one for TOF and one for SWI. Results are reported on the best performing model in the validation set.

The two proposed networks, derived from medical imaging task-specific networks, present a higher overall performance (F-score) than VGG-16 and the ResNet, suggesting that the networks specifically designed for medical imaging tasks can contribute to an increased performance. All methods report a drop in performance from TOF to SWI, which is expected given that SWIs are more challenging to classify and segment due to several factors. First, vessels in SWI are hypo-intense, being similar in appearance to the image background. As such, vessels close to the brain surface are prone to misclassification. Second, SWI is capable of imaging very small vessels that can be difficult to identify within a patch, as they can have an appearance similar to the one of brain tissue inhomogeneities or sulci, this leading to misclassification.

Among the proposed networks, 2D-PnetCl presents the highest performance in both modalities. This reflects a good balance in the network's capability to discriminate among vessel and non-vessel patches, which is key for its use within the Vessel-CAPTCHA framework. In the remaining, we rely on 2D-PnetCl as a classification network.

Classification network as a weak Pseudo-label generator We use a percentage (25%, 50% and 100%) of the weakly annotated training set $\mathcal{T}_M$. Where applicable, we enlarge it with a fixed set of 10 images automatically labeled through the data augmentation process, i.e. $|\mathcal{T}_M^*|$=10, (Fig. 4). Fig. 11 reports DSC in the different scenarios. The results show that the data augmentation step improves performance w.r.t. using the same annotated training set with no augmentation, while reaching a comparable performance to that one of using a dataset entirely annotated by the user. The comparable performances come as a result of the high classification accuracy of the 2D-PnetCl (F-score=94.71%), which sits close to the performance of a human rater.

Classification network as a second opinion The results obtained by post-processed classical methods (Table 3) suggest that a revision of the segmentation results and their refinement through post-processing can lead to a significant improvement in performance. We investigate if the classification network can act as an expert providing a second opinion on the segmentation results obtained by the 2D-WnetSeg, on a per-patch basis. If the classification network labels a patch as vessel patch, the segmented pixels in the patch will be preserved. Instead, if the classification network classifies the patch as a non-vessel one, any segmented pixels are masked out. To this end, we calibrate the 2D-PnetCl output by choosing the classification threshold of the final prediction layer, which maximizes the DSC (Fig. 12).

Fig. 13 reports vessel segmentation DSC, using Set 1 of the TOF images, in the following scenarios: (1) on all the testing set (ALL); (2) on 4 images identified as of low quality by the raters (LQ); (3) using a second opinion on the testing set (Cl(ALL)); (4) using a second opinion on the low quality data (Cl(LQ)); and (5) in all

Table 5
Classification network comparison in TOF and SWI. For each row, bold font denotes the best value, with underlined values not significantly different from it ($\alpha = 0.05$). An asterisk (*) denotes a network proposed in this work.

		VGG-16	ResNet	2D-UnetCl*	2D-PnetCl*
TOF	Precision	92.48 ± 1.54	93.66 ± 1.48	94.82 ± 0.48	**94.91 ± 1.04**
	Recall	87.39 ± 4.60	93.27 ± 1.73	94.04 ± 0.65	**94.94 ± 1.09**
	F-score	88.68 ± 3.81	93.34 ± 1.62	94.27 ± 0.54	**94.71 ± 1.23**
SWI	Precision	82.34 ± 1.15	80.14 ± 1.13	82.44 ± 1.18	**82.97 ± 1.55**
	Recall	77.45 ± 4.17	**79.39 ± 3.35**	74.35 ± 5.35	79.30 ± 4.07
	F-score	78.76 ± 3.39	79.17 ± 2.31	76.42 ± 4.63	**80.31 ± 3.31**

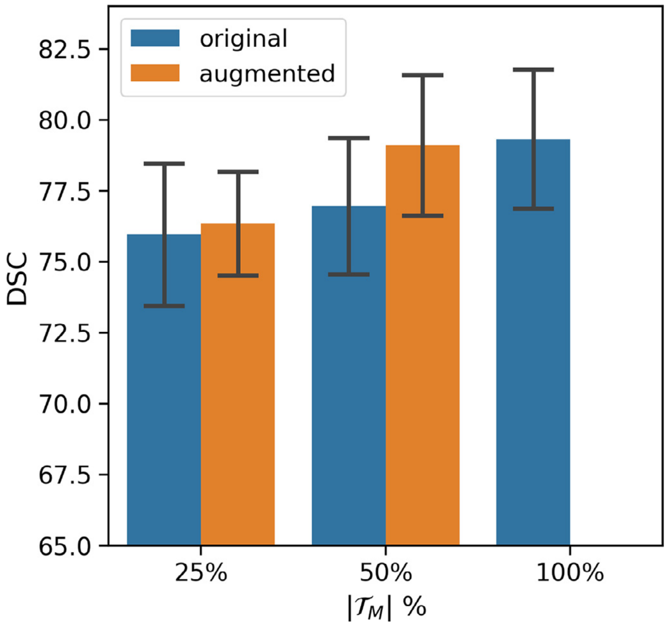

Fig. 11. Segmentation performance with varying training set size with (augmented) and without (original) data augmentation.

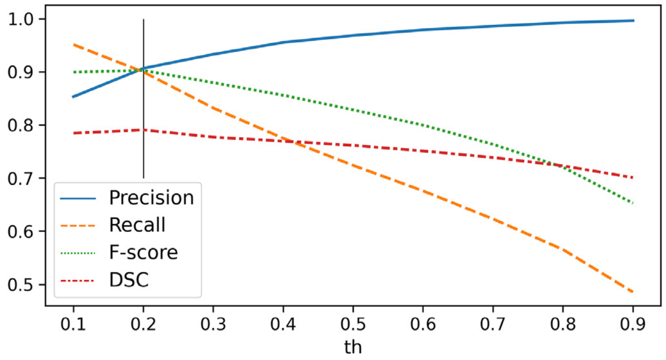

Fig. 12. Threshold (th) calibration of the 2D-PnetCl output. Precision, recall and F-score measure patch classification accuracy, wehereas DSC measures pixel-wise segmentation performance.

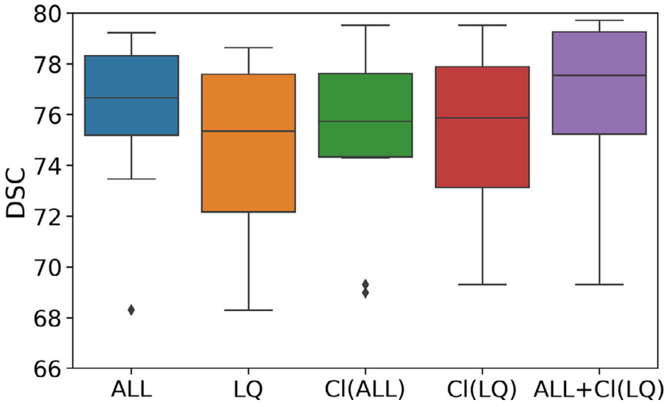

Fig. 13. Classification network as a second opinion in TOF. Vessel segmentation DSC for all the test set (ALL), low quality test images (LQ), full test set after second opinion (Cl(ALL)), low quality images after second opinion (Cl(LQ)) and full test with only the low quality subject to a second opinion (ALL+Cl(LQ)) using 2D-WnetSeg trained on original training set 1.

the testing set with the a second opinion only on the low quality data (ALL + Cl(LQ)). The results suggest that using the classifier network as a second opinion has a significant impact in the segmentations' accuracy and variability for low quality (LQ) images (p-value$<$0.05), although when applied to the full test set there is a slight drop in accuracy ($\sim$ 1.9%), indicating a negative impact on the segmentation accuracy in high quality images. As a result, one could consider the classifier as a second opinion and not the main expert. In images were there is a discrepancy between the segmentation network and the classifier, the user may inspect them and decide what to do. As an example, the second opinion could be used only on those images identified as of low quality by the raters. The results from Fig 13 indicate that, in such scenario, a higher overall performance is achieved.

We follow the same procedure using SWI segmentations and present the revised segmentation masks to the raters for visual judgement. The average rating score achieved was 2.30 with an agreement $\kappa = 0.57$, which is lower than that one achieved without using a second opinion (i.e. 2.57, see Section 4.1). This lower rating score is explained by the fact the classification network allows to correct segmentations containing large regions of false positives caused by noise in the image, mostly in the boundaries of the brain tissue, at the cost of removing true positives (Fig. 14). One rater considered this as less critical than the other, which explains the lower agreement among them. The results suggest that the classifier network should not be considered as an expert, i.e. it acts as a mask, but as a second opinion providing a heuristic measure of uncertainty on patches where the two networks disagree. The mismatching and uncertain regions should be thus validated by an external user.

4.4. Ablation study

We study the properties of the different components of the proposed annotation and segmentation framework through a set of ablation studies. We investigate the incidence of the K-means as and we investigate the role of the 2D-WnetSeg network.

4.4.1. K-means as a Pseudo-label generation strategy

We study how the pixel-wise pseudo-labeled dataset $\mathcal{T}_M$ synthesized from user-provided weak patch tags affects the framework's performance in TOF. We achieve this in two ways. First,

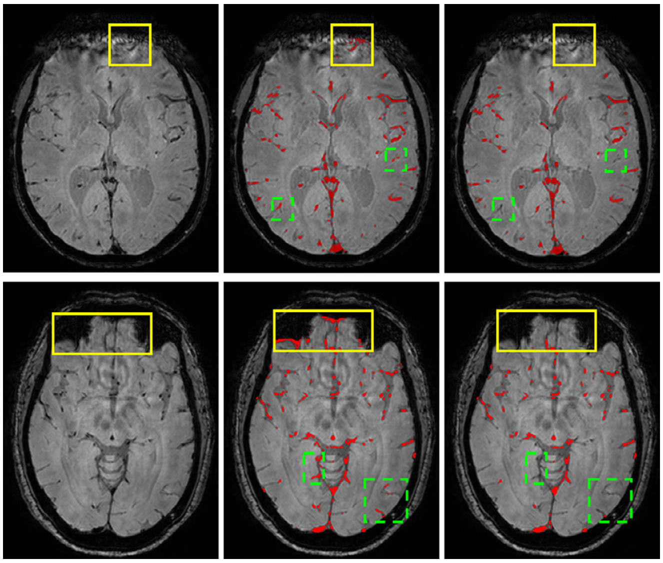

Fig. 14. Classification network as a second expert opinion in two SWI slices. From left to right, original image, segmentation from 2D-WnetSeg, segmentation after filtering. The yellow boxes highlight areas with image noise that are first segmented as vessel, but corrected with the filter. The green dashed boxes, highlight areas with segmented vessels that are removed. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

we investigate if the pixel-wise pseudo-labels synthesized by K-means represent a good rough approximation of pixel-wise user-annotated labels. Second, we assess how the size of the patches used as input of the segmentation network influences the latter's performance. In our experiments, we compare with Gaussian mixture models (GMM), an alternative self-supervised approach to obtain pixel-wise pseudo-labels from image tags (Luo et al., 2020). Two components (vessel and background) are used for the GMM to be comparable with K-means. For both cases, patches with more than 30% pixels marked as vessel are fully masked out and considered as non-vessel. These correspond to highly noisy patches containing only brain tissue.

The role of the self-supervised method, i.e. the K-means in our case, is to synthesize pixel-wise pseudo-label masks $\{\mathcal{M}_s\}_{s=1}^{S}$ which are sufficiently good to train the segmentation network. In other words, the pseudo-labels should be as close as possible to hypothetically pixel-wise annotations provided by a user. We thus measure the similarity between the pixel-wise pseudo-labeled masks $\{\mathcal{M}_s\}_{s=1}^{S}$ and the available pixel-wise annotations of the TOF training set. The K-means (and GMM) are applied on different input sizes, namely directly on the full image volume, or on subsets of it that are then concatenated. For this we use image slices and patches of varying sizes: 96, 64 and 32. For the patches, K-means and GMM are only applied to vessel patches. We set 32 as the smallest patch size, which corresponds to the size set for the Vessel-CAPTCHA, i.e. the user-input. Larger patches are obtained by concatenating the user input into a 2 × 2 and 3 × 3 grid.

Smaller patches are best for Pseudo-label generation Fig. 15(top) shows the similarity between the training set pixel-wise annotations and the weak pixel-wise label masks measured with the DSC. The performance of both methods is inverse to the size of the input sample. As it would be expected, when applied to large extents of the image volume, i.e. the full image volume (FV) or on a per image slice basis (IS), the DSC is very low ($< 40\%$), with GMM reporting slightly higher values. As the extent of the input sample decreases, i.e using patches, K-means performs better, which could be justified by the fact that smaller regions tend to be more homogeneous. Two aspects should be highlighted from the obtained results. Firstly, we observe that GMMs lead to thinner vessel masks than those synthesized by K-means (Fig. 16), which is consistent with the higher DSC, as over-segmentations tend to be less penalized than mis-segmentations. Given the way that the 2D-WnetSeg learns, it is better to have overestimated masks from K-means than the finer ones. However, being K-means a simpler algorithm, the patch size used as the input plays an important role. Our results suggest that smaller patch sizes lead to better results. Secondly, we shall recall that both self-supervised methods are only applied to vessel patches. This is a necessary condition to obtain pseudo-labels of a minimum quality using these two algorithms. The condition is guaranteed by the patch tags discriminating vessel from non-vessel patches, which are obtained through the Vessel-

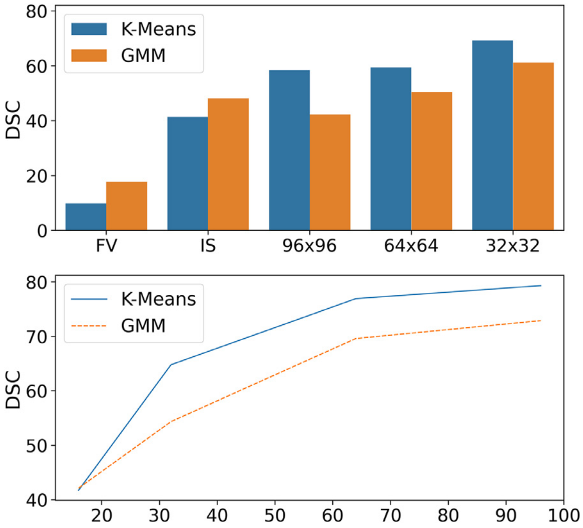

Fig. 15. Top: Similarity between user-provided pixel-wise annotations and weak pixel-wise labels obtained through K-means and GMM, measured through the DSC in TOF. K-means and GMM are applied on the full volume (FV), on a per slice basis (IS) and on different patch sizes. Bottom: 2D-WnetSeg performance using pixel-wise pseudo-labels by K-means and GMM for different input patch sizes (16, 32, 64 and 96).

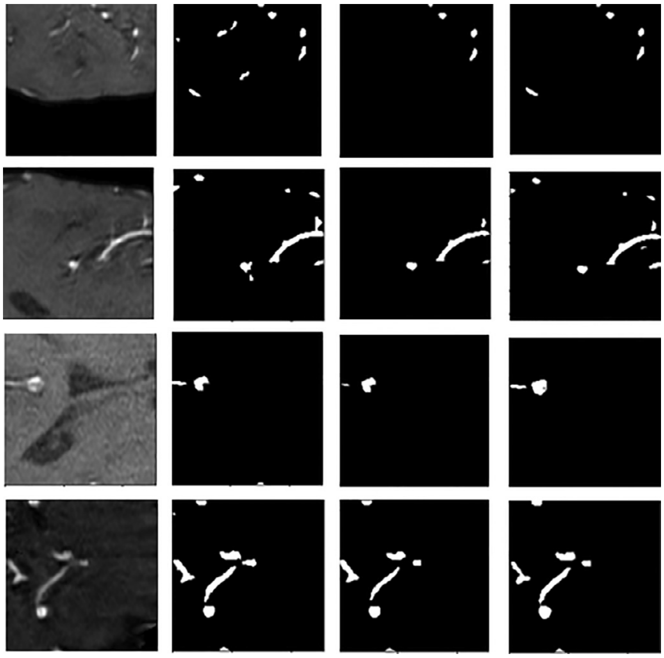

Fig. 16. Examples of the generated training set $\mathcal{T}_M$. From left to right original TOF image, ground truth, GMM pseudo-labels and K-means pseudo-labels.

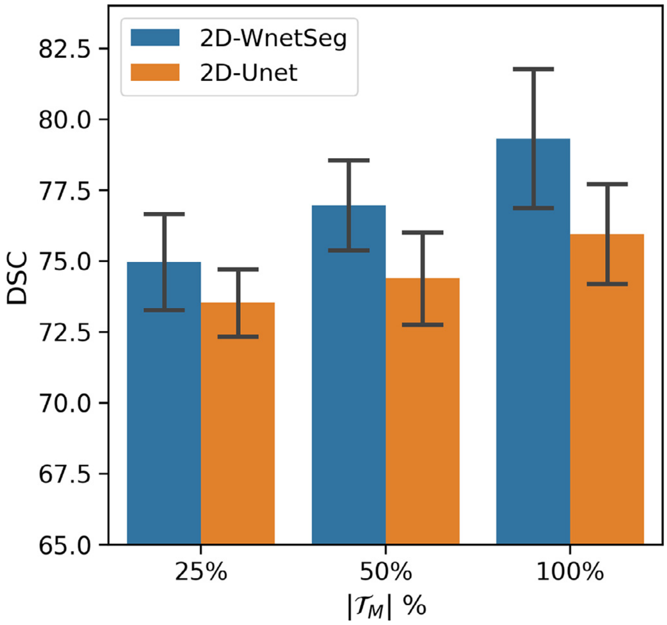

Fig. 17. 2D-WnetSeg (ours) vs. single Unet performance (DSC) for varying training set size, $|\mathcal{T}_M|$.

CAPTCHA. Based on these results, for the remaining experiments we set the patch size input to the K-means to 32 × 32, which corresponds to the same value used in the Vessel-CAPTCHA.

Larger patches are best for segmentation Fig. 15 (bottom) shows the 2D-WnetSeg accuracy with varying input patch sizes over the validation set. The patches are obtained by rebuilding the rough mask volume from the 32 × 32 patches and re-cropping the volume into different patch sizes. It should be noted that the segmentation network's input patch size does not have to match that one of the Vessel-CAPTCHA. Coherently with the previous results showing that K-means pseudo-labels are more similar to true annotations, their use consistently leads to higher DSCs. The Vessel-CAPTCHA patch size, 32 × 32, seems too small for the 2D-WnetSeg to capture the features that allow to discriminate vessel pixels from non-vessel ones. Instead, larger patches lead to higher DSCs. However, we avoid the use of larger patch sizes to avoid the problem of vessels becoming a small portion of the full image/patch, leading to drops in performance. For instance, we set the segmentation network's input patch size to 96 × 96.

4.4.2. The role of the segmentation network

We perform an ablation study to explore the effectiveness of the 2D-WnetSeg. Fig. 17 compares the performance of 2D-WnetSeg with its ablated version consisting its first Unet (2D-Unet), while varying the size of the training set. The 2D-WnetSeg reports a higher DSC across datasets. The better performance of the 2D-WnetSeg is explained by the fact that the deep networks are trained on rough segmentation maps. The first Unet works as a refinement module to correct the mask by inferring potentially missing vessels based on the structural redundancy of the cerebrovascular tree. The second Unet can learn from the raw brain image and the previously improved segmentation mask, leading to an increased segmentation performance. The single Unet, instead, is faced directly with the rough masks. We further investigate this behavior using the synthetic dataset, which provides a controlled setup for comparison (Table 6). The higher reported DSC of 2D-WnetSeg indicates it is better at detecting vessel pixels. Moreover, the lower 95HD and μD are a sign of the more refined results that the 2D-WnetSeg can achieve w.r.t. its ablated version.

Table 6
2D-WnetSeg (ours) vs. single Unet performance using synthetic data.

Measure	2D-WnetSeg	One 2D-Unet
DSC (↑)	88.77 ± 0.90	86.61 ± 1.05
HD (↓)	40.31 ± 2.95	41.18 ± 4.32
95HD (↓)	6.74 ± 0.48	7.96 ± 0.52
μD (↓)	0.91 ± 0.06	1.08 ± 0.07

4.5. Summary

Table 7 summarizes the performance of the different baselines compared in this work, along with their computational costs, in terms of model size, FLOPs, training and inference time, and user intervention time. Training time denotes the time required to train a learning-based model, except for Pseudo-labeling, where it refers to the time to train the model and to obtain sequentially the pseudo-labels for each image in our training and validation sets using the Sato filter. User intervention time represents the time to annotate the training set in learning-based approaches, or to post-process the segmentation results for classical methods. It should be noted that for the latter user intervention occurs every time an image is segmented, whereas for learning-based methods this only happens once during training. In addition to the considered baselines, we include two further methods for reference: the 2D-WnetSeg trained with pixel-wise annotations and the combination of the classifier network with K-means (no segmentation network). Overall, the Vessel-CAPTCHA has a performance comparable to the best fully supervised methods (Livne et al., 2019), it avoids any post-processing steps and it provides an important speed-up for training data annotation.

5. Discussion and conclusions

Context and proposed solution Deep convolutional networks have achieved state-of-the-art performance in many medical image segmentation tasks. However, their success has not been as wide for 3D brain vessel segmentation. This can be explained by two factors. First, deep learning techniques are less performing when the object of interest occupies a small portion of the image, as it is the case for brain vessels (Livne et al., 2019). Second, manual pixel-wise annotation of vessels is highly time consuming and complex (Moccia et al., 2018). In this work, we introduced the Vessel-CAPTCHA, an efficient learning framework for vessel annotation and segmentation. The framework formulates the Vessel-CAPTCHA annotation scheme, which allows users to annotate a dataset through simple clicks on patches containing vessels, similarly to the commonly used image-CAPTCHAs of web applications (von Ahn and Dabbish, 2004). As such, our work can be considered a multi-instance learning problem where a bag corresponds to an image patch and the instances are the image pixels to be segmented.

User-provided patch-level tags are used to synthesize pixel-wise pseudo-labels that serve as input to train a 2D patch-based segmentation network. In particular, we use the K-means algorithm to synthesize the pixel-wise pseudo-labels along with the proposed 2D-WnetSeg network, concatenating two 2D-Unets, as backbone architecture. The use of a 2D patch-based segmentation network instead of more complex end-to-end 3D or hybrid architectures, is motivated by the need to increase the object-of-interest to image size ratio, as a way to mitigate the reduced performance of deep learning-based methods when the object of interest does not occupy an important portion of the input image. Furthermore, this simplifies the learning process: at a larger scale, the complexity and uniqueness of each brain vessel tree makes it difficult to

Table 7
Performance summary considering segmentation accuracy, model complexity (Params, GFLOPs), and computational (training and prediction) and user intervention time in minutes. In classical models (NL), user intervention time is measured during inference. In learning-based models, it refers to the time used during training set annotation. For accuracy measures, the bold font denotes best value, with underlined values not significantly different from it ($\alpha = 0.05$).

	Method	Accuracy				Complexity (↓)		Time (↓)		
		DSC (↑)	HD (↓)	95HD (↓)	μD (↓)	Params $\times 10^3$	GFLOPs	Train	Predict	User
NL	Frangi-NP	54.16 ± 8.81	81.04 ± 18.48	14.78 ± 13.83	2.47 ± 2.22			0	25	0
	Sato-NP	55.75 ± 7.15	78.60 ± 16.37	11.53 ± 12.01	2.17 ± 1.07				25	0
	TV-NP	68.41 ± 5.01	60.23 ± 10.08	10.97 ± 11.72	2.10 ± 1.00				35	0
	Frangi-PP	68.44 ± 3.15	20.60 ± 10.91	9.01 ± 10.38	2.36 ± 2.01				25	25
	Sato-PP	69.01 ± 3.67	21.53 ± 9.11	8.86 ± 10.09	2.10 ± 1.01				25	25
	TV-PP	70.74 ± 3.38	**20.11 ± 8.45**	8.31 ± 8.23	2.07 ± 1.02				35	25
FS	Vessel 2D-Unet	77.66 ± 4.32	74.78 ± 16.73	12.60 ± 18.16	0.60 ± 0.11	31.38	15.6	90	< 1	327
	DeepVesselNet	76.13 ± 5.51	75.32 ± 12.94	4.32 ± 1.16	1.65 ± 0.26	0.05	NA	960	< 1	327
	2D-WnetSeg	76.63 ± 4.26	80.69 ± 23.20	13.15 ± 19.67	2.13 ± 2.37	16.34	25.90	90	< 1	327
LS	3D-Unet	68.50 ± 3.37	76.12 ± 8.47	15.72 ± 2.23	2.56 ± 1.44	16.21	1669.53	60	< 1	327
	Pseudo-labeling	54.90 ± 5.86	68.50 ± 9.58	24.19 ± 5.25	4.48 ± 1.67	31.38	15.6	910	< 1	0
	PnetCl + K-means	64.96 ± 4.76	65.82 ± 7.99	16.66 ± 3.85	2.62 ± 0.65	0.62	0.993	60	~1	75.5
	Vessel-CAPTCHA (ours)	**79.32 ± 3.02**	51.70 ± 5.92	**4.06 ± 1.50**	**0.50 ± 0.09**	16.34	25.90	90	<1	75.5

NL, No labels; FS, Fully supervised; LS, Limited supervision; NP, No post-processing; PP, Post-processing; NA, Not available.

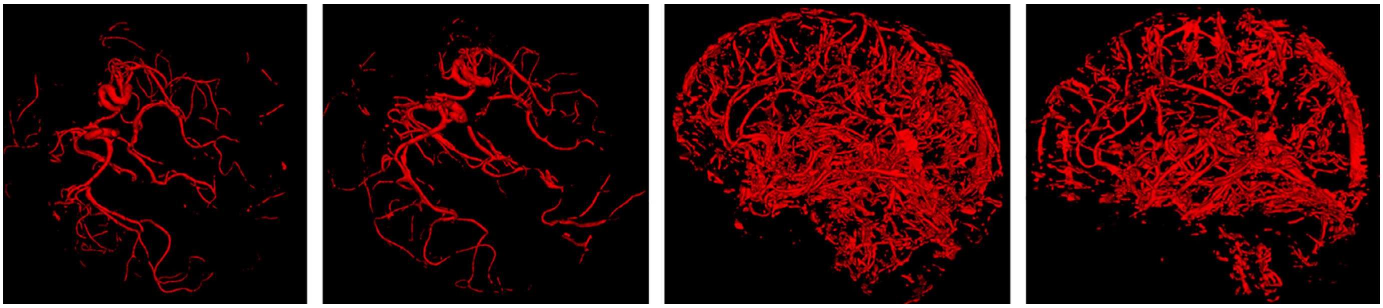

Fig. 18. 3D renderings of obtained segmentations in two TOF images (left) and two SWI (right).

learn common underlying patterns (Moriconi et al., 2019), whereas, at a local scale, the characteristic patterns of vessels are similar between each other, allowing the network to learn them. Reducing the input size is a common strategy in learning-based vessel segmentation, beyond brain vessel tree segmentation (Kitrungrotsakul et al., 2019; Koziński et al., 2020). The lower results obtained by 3D networks validate our choice of a 2D patch-based segmentation network.

To further ease the annotation process, our framework includes a classification network that can label training data without further user effort. This network is trained using the same user-provided patch tags and it allows to classify image patches from unseen images that can be used to enlarge the original training set without the need for further user annotations.

Framework evaluation We evaluated the proposed framework in terms of its accuracy and required annotation time, using a synthetic dataset and two image modalities, TOF and SWI. Our framework achieved performances comparable to those of current state-of-the-art deep learning approaches for brain vessel segmentation (Livne et al., 2019; Tetteh et al., 2020), while reducing the annotation burden by 77% on average. A visual inspection of the extracted trees showed a good continuity of the extracted vessel trees across image slices (Fig. 18). When compared to other approaches subject of limited supervision, our simple yet effective framework demonstrated its superiority. Our promising results, with competitive accuracies and a significant reduction of the user-required effort, should enable the wider use of deep learning techniques for vessel segmentation.

Our results show that the classifier network not only allows to enlarge the training dataset, but it can act as a second opinion to assess the segmentations. This concept could be further extended to guide a user in the manual correction of a segmentation mask. In this work, we used the classification network as an expert. However, the disagreements between the segmentation and classification network (i.e. 2D-WnetSeg segments a vessel in a patch classified as non-vessel or vice versa) could be used as a measure of uncertainty. Since WnetSeg and PnetCl architectures are significantly different, they extract low-level and high-level features differently. As such, they are complementary to each other: if both agree on a prediction over a patch, the prediction can be considered as one of high confidence, whereas when there is a disagreement the patch can be suggested to the rater for revision.

Limitations and perspectives Although our work focuses on the brain vessel tree, we consider that the proposed framework is general enough that it can be easily extended to other vascular structures (Aughwane et al., 2019), other tubular structures with complex networks to annotate (Zuluaga et al., 2014a), or different image modalities. However, for some modalities the K-means algorithm used to obtain pixel-wise pseudo-labels can be limited. As an example, the coronary vessel tree imaged with computed tomography angiography is likely to present calcified or lipid plaques that appear as hyper and hypo-intense objects, respectively (Zuluaga et al., 2011). In the current setup, they would be segmented as a vessel (calcified plaques) or the background (lipid plaques). A natural extension of this work would be to develop novel self-supervised methods, beyond those studied in this work, which can cope with the characteristics of different vessel/tubular trees and image modalities.

Our main effort in this work has been directed towards a simplified annotation process and the development of mechanisms that can mitigate the negative effects of 'simpler' annotations to achieve performances comparable to the state-of-the-art. Nevertheless, we consider that there are different ways that could be explored to achieve a higher performance and 3D vessel continuity. For instance, similarly to what has been proposed by (Koziński et al., 2020; Phellan et al., 2017), the annotations could be performed in different image planes. The use of multiple planes could contribute to improve the 3D consistency of the extracted vessel tree and its continuity. Currently, these are done in the axial plane. In addition, the Vessel-CAPTCHA allows for flexible annotations as, for some users, it is simpler to label vessels by following their trajectory. Now, all this information is discarded (see Fig. 2(e) and (g)), when in some cases it may have relevant content. The challenge here would be to identify when the patch annotations contain relevant information beyond the mere identification of the patch. Finally, one last limitation of the current framework is related to the selection of the patch grid scheme. While it is convenient to present non-overlapping patches to the user, in some cases, this may degrade the framework's performance. This is particularly true when the grid partition results in the split of vessels, in particular the smaller ones, across two or more patches causing them to lose their characteristic shape. The use of overlapping patches is a straightforward extension of this work that could reduce the number of misclassified vessels.

Declaration of Competing Interest

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

CRediT authorship contribution statement

Vien Ngoc Dang: Methodology, Investigation, Data curation, Validation, Software, Visualization, Writing – original draft, Writing – review & editing. **Francesco Galati:** Methodology, Investigation, Validation, Software, Visualization, Writing – review & editing. **Rosa Cortese:** Conceptualization, Data curation, Writing – review & editing. **Giuseppe Di Giacomo:** Data curation, Investigation. **Viola Marconetto:** Data curation, Investigation. **Prateek Mathur:** Data curation, Investigation. **Karim Lekadir:** Resources, Writing – review & editing. **Marco Lorenzi:** Conceptualization, Writing – original draft, Writing – review & editing. **Ferran Prados:** Conceptualization, Resources, Writing – review & editing. **Maria A. Zuluaga:** Conceptualization, Methodology, Resources, Data curation, Visualization, Writing – original draft, Writing – review & editing, Supervision.

Acknowledgments

ML and MAZ are supported by the French government, through the 3IA Côte d'Azur Investments in the Future project managed by the National Research Agency (ANR) (ANR-19-P3IA-0002). ML is partially funded by the ANR JCJC project Fed-BioMed (19-CE45-0006-01). KL received funding from the Spanish Ministry of Science, Innovation and Universities under grant agreement RTI2018-099898-B-I00 (HeartBrainCom). FP is funded in part by the National Institute for Health Research University College London Hospitals Biomedical Research Centre (NIHR BRC UCLH/UCL High Impact), and by a Non-Clinical Postdoctoral Guarantors of Brain fellowship.

References

Ahn, J., Kwak, S., 2018. Learning pixel-level semantic affinity with image-level supervision for weakly supervised semantic segmentation. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 4981–4990. doi:10.1109/cvpr.2018.00523.

Aughwane, R., Schaaf, C., Hutchinson, J., Virasami, A., Zuluaga, M., Sebire, N., Arthurs, O., Vercauteren, T., Ourselin, S., Melbourne, A., David, A., 2019. Micro-CT and histological investigation of the spatial pattern of feto-placental vascular density. Placenta 88, 36–43. doi:10.1016/j.placenta.2019.09.014.

Bae, W., Noh, J., Kim, G., 2020. Rethinking class activation mapping for weakly supervised object localization. In: Computer Vision. In: LNCS, 12360, pp. 618–634. doi:10.1007/978-3-030-58555-6_37.

Bai, W., Suzuki, H., Qin, C., Tarroni, G., Oktay, O., Matthews, P.M., Rueckert, D., 2018. Recurrent neural networks for aortic image sequence segmentation with sparse annotations. In: Medical Image Computing and Computer Assisted Intervention. In: LNCS, 11073, pp. 586–594. doi:10.1007/978-3-030-00937-3_67.

Benmansour, F., Cohen, L.D., 2009. Fast object segmentation by growing minimal paths from a single point on 2D or 3D images. J. Math. Imaging Vis. 33 (2), 209–221.

Bernier, M., Cunnane, S.C., Whittingstall, K., 2018. The morphology of the human cerebrovascular system. Hum. Brain Mapp. 39 (12), 4962–4975. doi:10.1002/hbm.24337.

Bruggemann, J., Lander, G.C., Su, A.I., 2018. Exploring applications of crowdsourcing to cryo-EM. J. Struct. Biol. 203 (1), 37–45.

Campanella, G., Hanna, M.G., Geneslaw, L., Miraflor, A., Silva, V.W.K., Busam, K.J., Brogi, E., Reuter, V.E., Klimstra, D.S., Fuchs, T.J., 2019. Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. Nat. Med. 25 (8), 1301–1309. doi:10.1038/s41591-019-0508-1.

Can, Y.B., Chaitanya, K., Mustafa, B., Koch, L.M., Konukoglu, E., Baumgartner, C.F., 2018. Learning to segment medical images with scribble-supervision alone. In: Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. In: LNCS, 11045, pp. 236–244. doi:10.1007/978-3-030-00889-5_27.

Cetin, S., Unal, G., 2015. A higher-order tensor vessel tractography for segmentation of vascular structures. IEEE Trans. Med. Imaging 34 (10), 2172–2185. doi:10.1109/tmi.2015.2425535.

Chen, H., Qi, X., Yu, L., Heng, P.-A., 2016. Dcan: deep contour-aware networks for accurate gland segmentation. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition, pp. 2487–2496. doi:10.1109/CVPR.2016.273.

Cheplygina, V., de Bruijne, M., Pluim, J.P., 2019. Not-so-supervised: a survey of semi-supervised, multi-instance, and transfer learning in medical image analysis. Med. Image Anal. 54, 280–296. doi:10.1016/j.media.2019.03.009.

Cheplygina, V., Perez-Rovira, A., Kuo, W., Tiddens, H.A.W.M., de Bruijne, M., 2016. Early experiences with crowdsourcing airway annotations in chest CT. In: Deep Learning and Data Labeling for Medical Applications, pp. 209–218. doi:10.1007/978-3-319-46976-8_22.

Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O., 2016. 3D U-net: learning dense volumetric segmentation from sparse annotation. In: Medical Image Computing and Computer-Assisted Intervention. In: LNCS, 9901, pp. 424–432. doi:10.1007/978-3-319-46723-8_49.

Dai, J., He, K., Sun, J., 2015. Boxsup: exploiting bounding boxes to supervise convolutional networks for semantic segmentation. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition, pp. 1635–1643. doi:10.1109/ICCV.2015.191.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., Fei-Fei, L., 2009. Imagenet: a large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition, pp. 248–255. doi:10.1109/CVPR.2009.5206848.

Dias, M., Monteiro, J., Estima, J., Silva, J., Martins, B., 2019. Semantic segmentation of high-resolution aerial imagery with W-Net models. In: Progress in Artificial Intelligence. EPIA 2019. In: LNCS, 11805, pp. 486–498. doi:10.1007/978-3-030-30244-3_40.

Dietterich, T.G., Lathrop, R.H., Lozano-Pérez, T., 1997. Solving the multiple instance problem with axis-parallel rectangles. Artif. Intell. 89 (1–2), 31–71.

Elson, J., Douceur, J.R., Howell, J., Saul, J., 2007. Asirra: a CAPTCHA that exploits interest-aligned manual image categorization. In: ACM Conference on Computer and Communications Security, 7, pp. 366–374. doi:10.1145/1315245.1315291.

Feng, X., Yang, J., Laine, A. F., Angelini, E. D., 2017. Discriminative localization in CNNs for weakly-supervised segmentation of pulmonary nodules. CoRR abs/1707.01086

Frangi, A.F., Niessen, W.J., Vincken, K.L., Viergever, M.A., 1998. Multiscale vessel enhancement filtering. In: Medical Image Computing and Computer-Assisted Intervention. In: LNCS, 1496, pp. 130–137. doi:10.1007/bfb0056195.

Full, P.M., Isensee, F., Jger, P.F., Maier-Hein, K., 2021. Studying robustness of semantic segmentation under domain shift in cardiac MRI. In: Statistical Atlases and Computational Models of the Heart. M&Ms and EMIDEC Challenges, pp. 238–249. doi:10.1007/978-3-030-68107-4_24.

Gao, M., Huang, J., Huang, X., Zhang, S., Metaxas, D.N., 2012. Simplified labeling process for medical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention, pp. 387–394. doi:10.1007/978-3-642-33418-4_48.

Hassouna, M.S., Farag, A., Hushek, S., Moriarty, T., 2006. Cerebrovascular segmentation from TOF using stochastic models. Med. Image Anal. 10 (1), 2–18. doi:10.1016/j.media.2004.11.009.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778. doi:10.1109/CVPR.2016.90.

Hong, S., Yeo, D., Kwak, S., Lee, H., Han, B., 2017. Weakly supervised semantic segmentation using web-crawled videos. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition, pp. 7322–7330. doi:10.1109/cvpr.2017.239.

Hou, L., Samaras, D., Kurc, T.M., Gao, Y., Davis, J.E., Saltz, J.H., 2016. Patch-based convolutional neural network for whole slide tissue image classification. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition, pp. 2424–2433. doi:10.1109/cvpr.2016.266.

Ilse, M., Tomczak, J., Welling, M., 2018. Attention-based deep multiple instance learning. In: International conference on machine learning. PMLR, pp. 2127–2136.

Izadyyazdanabadi, M., Belykh, E., Cavallo, C., Zhao, X., Gandhi, S., Moreira, L.B., Eschbacher, J., Nakaji, P., Preul, M.C., Yang, Y., 2018. Weakly-supervised learning-based feature localization for confocal laser endomicroscopy glioma images. In: Medical Image Computing and Computer Assisted Intervention. In: LNCS, 11071, pp. 300–308.

Jia, Z., Huang, X., Chang, E.I.-C., Xu, Y., 2017. Constrained deep weak supervision for histopathology image segmentation. IEEE Trans. Med. Imaging 36 (11), 2376–2388. doi:10.1109/tmi.2017.2724070.

Kamnitsas, K., Ledig, C., Newcombe, V.F., Simpson, J.P., Kane, A.D., Menon, D.K., Rueckert, D., Glocker, B., 2017. Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. Med. Image Anal. 36, 61–78. doi:10.1016/j.media.2016.10.004.

Kandil, H., Soliman, A., Taher, F., Mahmoud, A., Elmaghraby, A., El-Baz, A., 2018. Using 3-D CNNs and local blood flow information to segment cerebral vasculature. In: 2018 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT), pp. 701–705. doi:10.1109/isspit.2018.8642676.

Ke, R., Bugeau, A., Papadakis, N., Schuetz, P., Schnlieb, C.-B., 2020. Learning to segment microscopy images with lazy labels. In: Computer Vision – ECCV 2020 Workshops, pp. 411–428. doi:10.1007/978-3-030-66415-2_27.

Kitrungrotsakul, T., Han, X.-H., Iwamoto, Y., Lin, L., Foruzan, A.H., Xiong, W., Chen, Y.-W., 2019. Vesselnet: a deep convolutional neural network with multi pathways for robust hepatic vessel segmentation. Comput. Med. Imaging Graph. 75, 74–83. doi:10.1016/j.compmedimag.2019.05.002.

Klepaczko, A., Szczypiński, P., Deistung, A., Reichenbach, J.R., Materka, A., 2016. Simulation of MR angiography imaging for validation of cerebral arteries segmentation algorithms. Comput. Methods Prog. Biomed. 137, 293–309. doi:10.1016/j.cmpb.2016.09.020.

Koziński, M., Mosinska, A., Salzmann, M., Fua, P., 2020. Tracing in 2D to reduce the annotation effort for 3D deep delineation of linear structures. Med. Image Anal. 60, 101590. doi:10.1016/j.media.2019.101590.

Krähenbühl, P., Koltun, V., 2011. Efficient inference in fully connected CRFs with gaussian edge potentials. In: Shawe-Taylor, J., Zemel, R.S., Bartlett, P.L., Pereira, F.C.N., Weinberger, K.Q. (Eds.), Advances in Neural Information Processing Systems 24, pp. 109–117.

Kraus, O.Z., Ba, J.L., Frey, B.J., 2016. Classifying and segmenting microscopy images with deep multiple instance learning. Bioinformatics 32 (12), i52–i59.

LaMontagne, P. J., Benzinger, T. L., Morris, J. C., Keefe, S., Hornbeck, R., Xiong, C., Grant, E., Hassenstab, J., Moulder, K., Vlassenko, A. G., Raichle, M. E., Cruchaga, C., Marcus, D., 2019. OASIS-3: longitudinal neuroimaging, clinical, and cognitive dataset for normal aging and Alzheimer's disease. medRxiv. 10.1101/2019.12.13.19014902

Law, M.W.K., Chung, A.C.S., 2008. Three dimensional curvilinear structure detection using optimally oriented flux. In: Proceedings of the European Conference on Computer Vision. In: LNCS, 5305, pp. 368–382. doi:10.1007/978-3-540-88693-8_27.

Leibig, C., Allken, V., Ayhan, M.S., Berens, P., Wahl, S., 2017. Leveraging uncertainty information from deep neural networks for disease detection. Sci. Rep. 7 (1). doi:10.1038/s41598-017-17876-z.

Lerousseau, M., Vakalopoulou, M., Classe, M., Adam, J., Battistella, E., Carré, A., Estienne, T., Henry, T., Deutsch, E., Paragios, N., 2020. Weakly supervised multiple instance learning histopathological tumor segmentation. In: Medical Image Computing and Computer Assisted Intervention. In: LNCS, 12265, pp. 470–479. doi:10.1007/978-3-030-59722-1_45.

Lesage, D., Angelini, E.D., Bloch, I., Funka-Lea, G., 2009. A review of 3D vessel lumen segmentation techniques: models, features and extraction schemes. Med. Image Anal. 13 (6), 819–845. doi:10.1016/j.media.2009.07.011.

Li, K., Vakharia, V.N., Sparks, R., Rodionov, R., Vos, S.B., McEvoy, A.W., Miserocchi, A., Wang, M., Ourselin, S., Duncan, J.S., 2019. Stereoelectroencephalography electrode placement: detection of blood vessel conflicts. Epilepsia 60 (9), 1942–1948.

Li, N., Wang, W.-T., Sati, P., Pham, D.L., Butman, J.A., 2014. Quantitative assessment of susceptibility-weighted imaging processing methods. J. Magn. Reson. Imaging 40 (6), 1463–1473.

Li, X., Yang, F., Cheng, H., Liu, W., Shen, D., 2018. Contour knowledge transfer for salient object detection. In: Proceedings of the European Conference on Computer Vision, pp. 370–385. doi:10.1007/978-3-030-01267-0_22.

Liang, Q., Nan, Y., Coppola, G., Zou, K., Sun, W., Zhang, D., Wang, Y., Yu, G., 2019. Weakly supervised biomedical image segmentation by reiterative learning. IEEE J. Biomed. Health Inform. 23 (3), 1205–1214. doi:10.1109/JBHI.2018.2850040.

Lin, D., Dai, J., Jia, J., He, K., Sun, J., 2016. ScribbleSup: scribble-supervised convolutional networks for semantic segmentation. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition, pp. 3159–3167. doi:10.1109/CVPR.2016.344.

Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghafoorian, M., Van Der Laak, J.A., Van Ginneken, B., Sánchez, C.I., 2017. A survey on deep learning in medical image analysis. Med. Image Anal. 42, 60–88. doi:10.1016/j.media.2017.07.005.

Liu, G., Wu, J., Zhou, Z.-H., 2012. Key instance detection in multi-instance learning. In: Asian Conference on Machine Learning. PMLR, pp. 253–268.

Livne, M., Rieger, J., Aydin, O.U., Taha, A.A., Akay, E.M., Kossen, T., Sobesky, J., Kelleher, J.D., Hildebrand, K., Frey, D., Madai, V.I., 2019. A U-net deep learning framework for high performance vessel segmentation in patients with cerebrovascular disease. Front. Neurosci. 13. doi:10.3389/fnins.2019.00097.

Lu, M.Y., Williamson, D.F.K., Chen, T.Y., Chen, R.J., Barbieri, M., Mahmood, F., 2021. Data-efficient and weakly supervised computational pathology on whole-slide images. Nat. Biomed. Eng. 5 (6), 555–570. doi:10.1038/s41551-020-00682-w.

Lundervold, A.S., Lundervold, A., 2019. An overview of deep learning in medical imaging focusing on MRI. Z. Med. Phys. 29 (2), 102–127. doi:10.1016/j.zemedi.2018.11.002.

Luo, A., Li, X., Yang, F., Jiao, Z., Cheng, H., 2020. Webly-supervised learning for salient object detection. Pattern Recognit. 103, 107308. doi:10.1016/j.patcog.2020.107308.

Maron, O., Lozano-Pérez, T., 1997. A framework for multiple-instance learning. In: Jordan, M.I., Kearns, M.J., Solla, S.A. (Eds.), Advances in Neural Information Processing Systems, 10, pp. 570–576.

Matuszewski, D.J., Sintorn, I.-M., 2018. Minimal annotation training for segmentation of microscopy images. In: 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), pp. 387–390.

Milletari, F., Navab, N., Ahmadi, S.-A., 2016. V-net: fully convolutional neural networks for volumetric medical image segmentation. In: 2016 Fourth International Conference on 3D Vision (3DV), pp. 565–571. doi:10.1109/3dv.2016.79.

Moccia, S., Momi, E.D., Hadji, S.E., Mattos, L.S., 2018. Blood vessel segmentation algorithms — review of methods, datasets and evaluation metrics. Comput. Methods Prog. Biomed. 158, 71–91. doi:10.1016/j.cmpb.2018.02.001.

Moriconi, S., Zuluaga, M.A., Jager, H.R., Nachev, P., Ourselin, S., Cardoso, M.J., 2019. Inference of cerebrovascular topology with geodesic minimum spanning trees. IEEE Trans. Med. Imaging 38 (1), 225–239. doi:10.1109/tmi.2018.2860239.

Morrison, M.A., Payabvash, S., Chen, Y., Avadiappan, S., Shah, M., Zou, X., Hess, C.P., Lupo, J.M., 2018. A user-guided tool for semi-automated cerebral microbleed detection and volume segmentation: evaluating vascular injury and data labelling for machine learning. NeuroImage 20, 498–505. doi:10.1016/j.nicl.2018.08.002.

Ni, J., Wu, J., Wang, H., Tong, J., Chen, Z., Wong, K.K., Abbott, D., 2020. Global channel attention networks for intracranial vessel segmentation. Comput. Biol. Med. 118, 103639. doi:10.1016/j.compbiomed.2020.103639.

Niu, Z., Zhong, G., Yu, H., 2021. A review on the attention mechanism of deep learning. Neurocomputing 452, 48–62. doi:10.1016/j.neucom.2021.03.091.

Ørting, S.N., Doyle, A., Hilten, A.V., Hirth, M., Inel, O., Madan, C.R., Mavridis, P., Spiers, H., Cheplygina, V., 2020. A survey of crowdsourcing in medical image analysis. Hum. Comput. 7, 1–26. doi:10.15346/hc.v7i1.1.

Ouyang, X., Xue, Z., Zhan, Y., Zhou, X.S., Wang, Q., Zhou, Y., Wang, Q., Cheng, J.-Z., 2019. Weakly supervised segmentation framework with uncertainty: a study on pneumothorax segmentation in chest x-ray. In: Medical Image Computing and Computer Assisted Intervention, pp. 613–621. doi:10.1007/978-3-030-32226-7_68.

Pepe, A., Schussnig, R., Li, J., Gsaxner, C., Chen, X., Fries, T.-P., Egger, J., 2020. IRIS: interactive real-time feedback image segmentation with deep learning. In: Proc. SPIE Medical Imaging 2020: Biomedical Applications in Molecular, Structural, and Functional Imaging, 11317, p. 113170R. doi:10.1117/12.2551354.

Phellan, R., Peixinho, A., Falcão, A., Forkert, N.D., 2017. Vascular segmentation in TOF MRA images of the brain using a deep convolutional neural network. In: Intravascular Imaging and Computer Assisted Stenting, and Large-Scale Annotation of Biomedical Data and Expert Label Synthesis. In: LNCS, 10552, pp. 39–46. doi:10.1007/978-3-319-67534-3_5.

Qi, H., Collins, S., Noble, A., 2017. Weakly supervised learning of placental ultrasound images with residual networks. In: Annual Conference on Medical Image Understanding and Analysis, pp. 98–108. doi:10.1007/978-3-319-60964-5_9.

Quellec, G., Lamard, M., Abràmoff, M.D., Decencière, E., Lay, B., Erginay, A., Cochener, B., Cazuguel, G., 2012. A multiple-instance learning framework for diabetic retinopathy screening. Med. Image Anal. 16 (6), 1228–1240. doi:10.1016/j.media.2012.06.003.

Radbruch, A., Mucke, J., Schweser, F., Deistung, A., Ringleb, P.A., Ziener, C.H., Roethke, M., Schlemmer, H.-P., Heiland, S., Reichenbach, J.R., Bendszus, M., Rohde, S., 2013. Comparison of susceptibility weighted imaging and TOF-angiography for the detection of thrombi in acute stroke. PLoS One 8 (5), e63459. doi:10.1371/journal.pone.0063459.

Rajchl, M., Lee, M.C.H., Oktay, O., Kamnitsas, K., Passerat-Palmbach, J., Bai, W., Damodaram, M., Rutherford, M.A., Hajnal, J.V., Kainz, B., Rueckert, D., 2017. DeepCut: object segmentation from bounding box annotations using convolutional neural networks. IEEE Trans. Med. Imaging 36 (2), 674–683. doi:10.1109/tmi.2016.2621185.

Raza, H., Ravanbakhsh, M., Klein, T., Nabi, M., 2019. Weakly supervised one shot segmentation. In: 2019 IEEE/CVF International Conference on Computer Vision Workshop, pp. 1401–1406. doi:10.1109/ICCVW.2019.00176.

Rempfler, M., Schneider, M., Ielacqua, G.D., Xiao, X., Stock, S.R., Klohs, J., Székely, G., Andres, B., Menze, B.H., 2015. Reconstructing cerebrovascular networks under local physiological constraints by integer programming. Med. Image Anal. 25 (1), 86–94. doi:10.1016/j.media.2015.03.008.

Robben, D., Tretken, E., Sunaert, S., Thijs, V., Wilms, G., Fua, P., Maes, F., Suetens, P., 2016. Simultaneous segmentation and anatomical labeling of the cerebral vasculature. Med. Image Anal. 32, 201–215. doi:10.1016/j.media.2016.03.006.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention. In: LNCS, 9351, pp. 234–241. doi:10.1007/978-3-319-24574-4_28.

Sato, Y., Nakajima, S., Atsumi, H., Koller, T., Gerig, G., Yoshida, S., Kikinis, R., 1997. 3D multi-scale line filter for segmentation and visualization of curvilinear structures in medical images. In: First Joint Conference Computer Vision, Virtual Reality and Robotics in Medicine and Medical Robotics and Computer-Assisted Surgery. In: LNCS, 1205, pp. 213–222. doi:10.1007/bfb0029240.

Schlegl, T., Waldstein, S.M., Vogl, W.-D., Schmidt-Erfurth, U., Langs, G., 2015. Predicting semantic descriptions from medical images with convolutional neural networks. In: Information Processing in Medical Imaging. In: LNCS, 9123, pp. 437–448. doi:10.1007/978-3-319-19992-4_34.

Schneider, M., Reichold, J., Weber, B., Székely, G., Hirsch, S., 2012. Tissue metabolism driven arterial tree generation. Med. Image Anal. 16 (7), 1397–1414. doi:10.1016/j.media.2012.04.009.

Setio, A.A.A., Ciompi, F., Litjens, G., Gerke, P., Jacobs, C., van Riel, S.J., Wille, M.M.W., Naqibullah, M., Sanchez, C.I., van Ginneken, B., 2016. Pulmonary nodule detection in CT images: false positive reduction using multi-view convolutional networks. IEEE Trans. Med. Imaging 35 (5), 1160–1169. doi:10.1109/tmi.2016.2536809.

Shelhamer, E., Long, J., Darrell, T., 2017. Fully convolutional networks for semantic segmentation. IEEE Trans. Pattern Anal. Mach. Intell. 39 (4), 640–651. doi:10.1109/tpami.2016.2572683.

Shen, Y., Wu, N., Phang, J., Park, J., Liu, K., Tyagi, S., Heacock, L., Kim, S.G., Moy, L., Cho, K., et al., 2021. An interpretable classifier for high-resolution breast cancer screening images utilizing weakly supervised localization. Med. Image Anal. 68, 101908.

Shin, S.Y., Lee, S., Yun, I.D., Kim, S.M., Lee, K.M., 2019. Joint weakly and semi-supervised deep learning for localization and classification of masses in breast ultrasound images. IEEE Trans. Med. Imaging 38, 762–774.

Simonyan, K., Zisserman, A., 2015. Very deep convolutional networks for large-scale image recognition. In: 3rd International Conference on Learning Representations, Conference Track Proceedings. http://arxiv.org/abs/1409.1556

Taher, F., Soliman, A., Kandil, H., Mahmoud, A., Shalaby, A., Gimel'farb, G., El-Baz, A., 2020. Accurate segmentation of cerebrovasculature from TOF-MRA images using appearance descriptors. IEEE Access 8, 96139–96149. doi:10.1109/access.2020.2982869.

Tajbakhsh, N., Jeyaseelan, L., Li, Q., Chiang, J.N., Wu, Z., Ding, X., 2020. Embracing imperfect datasets: a review of deep learning solutions for medical image segmentation. Med. Image Anal. 63, 101693.

Tetteh, G., Efremov, V., Forkert, N.D., Schneider, M., Kirschke, J., Weber, B., Zimmer, C., Piraud, M., Menze, B.H., 2020. Deepvesselnet: vessel segmentation, centerline prediction, and bifurcation detection in 3-D angiographic volumes. Front. Neurosci. 14.

von Ahn, L., Dabbish, L., 2004. Labeling images with a computer game. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 319–326. doi:10.1145/985692.985733.

Vrugt, J.A., Robinson, B.A., 2007. Treatment of uncertainty using ensemble methods: comparison of sequential data assimilation and Bayesian model averaging. Water Resour. Res. 43 (1). doi:10.1029/2005wr004838.

Wang, G., Li, W., Zuluaga, M.A., Pratt, R., Patel, P.A., Aertsen, M., Doel, T., David, A.L., Deprest, J., Ourselin, S., Vercauteren, T., 2018. Interactive medical image segmentation using deep learning with image-specific fine tuning. IEEE Trans. Med. Imaging 37 (7), 1562–1573. doi:10.1109/tmi.2018.2791721.

Wang, G., Zuluaga, M.A., Li, W., Pratt, R., Patel, P.A., Aertsen, M., Doel, T., David, A.L., Deprest, J., Ourselin, S., Vercauteren, T., 2019. DeepIGeoS: a deep interactive geodesic framework for medical image segmentation. IEEE Trans. Pattern Anal. Mach. Intell. 41 (7), 1559–1572. doi:10.1109/tpami.2018.2840695.

Wang, G., Zuluaga, M.A., Pratt, R., Aertsen, M., David, A.L., Deprest, J., Vercauteren, T., Ourselin, S., 2015. Slic-Seg: Slice-by-slice segmentation propagation of the placenta in fetal MRI using one-plane scribbles and online learning. In: Medical Image Computing and Computer-Assisted Intervention. In: LNCS, 9351, pp. 29–37. doi:10.1007/978-3-319-24574-4_4.

World Health Organization, 2020. Global health estimates. https://www.who.int/data/global-health-estimates.

Xu, G., Song, Z., Sun, Z., Ku, C., Yang, Z., Liu, C., Wang, S., Ma, J., Xu, W., 2019. CAMEL: a weakly supervised learning framework for histopathology image segmentation. In: 2019 IEEE/CVF International Conference on Computer Vision, pp. 10681–10690. doi:10.1109/ICCV.2019.01078.

Xu, Y., Zhu, J.-Y., Chang, E.I.-C., Lai, M., Tu, Z., 2014. Weakly supervised histopathology cancer image segmentation and classification. Med. Image Anal. 18 (3), 591–604. doi:10.1016/j.media.2014.01.010.

Zhao, R., Liao, W., Zou, B., Chen, Z., Li, S., 2019. Weakly-supervised simultaneous evidence identification and segmentation for automated glaucoma diagnosis. In: The Thirty-Third AAAI Conference on Artificial Intelligence, AAAI 2019, 33, pp. 809–816. doi:10.1609/aaai.v33i01.3301809.

Zhao, S., Zhou, M., Tian, Y., Xu, P., Wu, Z., Deng, Q., 2015. Extraction of vessel networks based on multiview projection and phase field model. Neurocomputing 162, 234–244. doi:10.1016/j.neucom.2015.03.048.

Zhou, B., Khosla, A., Lapedriza, À., Oliva, A., Torralba, A., 2016. Learning deep features for discriminative localization. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition, pp. 2921–2929. doi:10.1109/CVPR.2016.319.

Zou, Y., Zhang, Z., Zhang, H., Li, C., Bian, X., Huang, J., Pfister, T., 2021. Pseudoseg: designing pseudo labels for semantic segmentation. In: 9th International Conference on Learning Representations, ICLR 2021.

Zuluaga, M.A., Hush, D., Leyton, E.J.D., Hoyos, M.H., Orkisz, M., 2011. Learning from only positive and unlabeled data to detect lesions in vascular CT images. In: Medical Image Computing and Computer-Assisted Intervention. In: LNCS, 6893, pp. 9–16. doi:10.1007/978-3-642-23626-6_2.

Zuluaga, M.A., Orkisz, M., Dong, P., Pacureanu, A., Gouttenoire, P.-J., Peyrin, F., 2014. Bone canalicular network segmentation in 3D nano-CT images through geodesic voting and image tessellation. Phys. Med. Biol. 59 (9), 2155–2171. doi:10.1088/0031-9155/59/9/2155.

Zuluaga, M.A., Rodionov, R., Nowell, M., Achhala, S., Zombori, G., Cardoso, M.J., Miserocchi, A., McEvoy, A.W., Duncan, J.S., Ourselin, S., 2014. SEEG trajectory planning: combining stability, structure and scale in vessel extraction. In: Medical Image Computing and Computer-Assisted Intervention. In: LNCS, 8674, pp. 651–658. doi:10.1007/978-3-319-10470-6_81.