DEPARTMENT OF INFORMATICS

TECHNISCHE UNIVERSITÄT MÜNCHEN

Master's Thesis in Informatics

Deep semi supervised carotid vessel wall segmentation and artery localization with shape priors

Marta Saiz Vivó

DEPARTMENT OF INFORMATICS

TECHNISCHE UNIVERSITÄT MÜNCHEN

Master's Thesis in Informatics

Deep semi supervised carotid vessel wall segmentation and artery localization with shape priors

Tiefe halbüberwachte Segmentierung der Halsschlagaderwand und Arterienlokalisation mit Formpriorität

Author:	Marta Saiz Vivó
Supervisor:	Prof. Dr. Nassir Navab
Advisor:	Shahrooz Faghih Roohi
Submission Date:	15-10-2021

I confirm that this master's thesis in informatics is my own work and I have documented all sources and material used.

Munich,   Marta Saiz Vivó

Acknowledgments

Abstract

Carotid atherosclerosis is a risk factor for transient ischemic attacks (TIAs) and stroke, one of the main causes of global mortality. The disease is characterized by the deposit of plaque in the interior of the carotid artery wall . As the plaque grows it causes luminal narrowing and may rupture and trigger ischemic events. Quantitative analysis of plaque components and vessel wall (VW) is useful to identify high-risk plaques and monitor disease progression. For this, inner and outer-wall contour must be delineated from image acquisitions. The Three-dimensional - Motion-Sensitized Driven Equilibrium(MSDE)-prepared Rapid Gradient Echo (3D-MERGE) Magnetic Resonance Imaging (MRI) developed sequence has shown promising performance in atherosclerotic plaque visualization. However, its integration in the clinical practice is limited due to the large number of slices that must be manually delineated which is a time-consuming process.

In this thesis we approach the Carotid Artery Vessel Wall Segmentation Challenge 2021 where the aim is the automatically segment the common carotid artery (CCA), external carotid artery (ECA) and internal carotid artery (ICA) VW with high accuracy and robustness from 3D-MERGE MRI images. The main challenges were, the heterogeneity and small size of the VW morphology and the sparsity of annotations provided. To overcome these challenges, we implemented an automated VW segmentation workflow with deep learning (DL) algorithms based on semi-supervised learning (SSL) with shape constraints. For automatic carotid artery localization, a state-of-the-art object detection algorithm was custom trained with labelled slices (in total 2078). Then, a 3D vessel centreline reconstruction step increased the accuracy of detection for region of interest (ROI) extraction. For VW segmentation a dual-task semi-supervised segmentation network with signed distance map (SDM) regression was trained with labelled (2078) and unlabelled (8665) carotid patch slices where a task-level consistency regularization guided the mutual learning of both networks. This allowed to impose geometrical constraints and take advantage of unlabelled data. Additional shape constraints were applied at a post-processing stage with a persistent homology (PH)-based topological loss function for VW topology correction. The results suggest that in challenging segmentation problems such as vessel wall-Magnetic Resonance Imaging (VW-MRI) segmentation, the model can benefit from enforced shape constraints both during training and at test-time adaptation. Furthermore, the PH-based post-processing module successfully increased the percent of topologically meaningful VW predictions by $\sim 7\%$. The final aim of this work is to advance towards the implementation of DL-based algorithms in the clinical workflow for carotid atherosclerosis lesion assessment.

keywords: segmentation, detection, semi-supervised learning, shape analysis, vessel wall, 3D-MERGE

Kurzfassung

Die Atherosklerose der Karotis ist ein Risikofaktor für transitorische ischämische Attacken und Schlaganfälle, eine der Hauptursachen für die weltweite Sterblichkeit. Die Krankheit ist durch die Ablagerung von Plaque im Inneren der Halsschlagader gekennzeichnet. Die quantitative Analyse von Plaque-Komponenten und der Gefäßwand ist nützlich, um Hochrisiko-Plaques zu identifizieren und den Krankheitsverlauf zu überwachen. Dafür müssen Innen- und Außenwandkontur aus Bildaufnahmen abgegrenzt werden. 3D Black Blood Motion Sensitized Driven Equilibrium-prepped Rapid Gradient Echo (3D-MERGE) Magnetresonanztomographie (MRT) Sequenzen ermöglichen die Visualisierung von atherosklerotischen Plaques. Allerdings ist die Umsetzung in der klinischen Praxis aufgrund der manuellen Segmentierung begrenzt.

In dieser Arbeit nähern wir uns der Carotid Artery Vessel Wall Segmentation Challenge 2021, bei der das Ziel die automatische Segmentierung der gemeinsamen, externen und inneren Halsschlagader mit hoher Genauigkeit und Robustheit aus 3D MERGE . ist MRT-Bilder. Die Hauptherausforderungen waren die Heterogenität und die geringe Größe der Gefäßwandmorphologie und die spärliche Annotationen. Um diese Herausforderungen zu meistern, haben wir einen automatisierten Workflow zur Segmentierung der Gefäßwände implementiert, der auf Deep Semi-Supervised Learning mit Formbeschränkungen basiert. Für die automatische Lokalisation der Halsschlagader wurde ein Objekterkennungsalgorithmus trainiert. Anschließend erhöhte ein 3D-Schritt zur Rekonstruktion der Gefäßmittellinie die Erkennungsgenauigkeit für die Extraktion der interessierenden Region. Für die Segmentierung der Gefäßwände wurde ein semi-überwachtes Segmentierungsnetzwerk mit zwei Aufgaben und einer vorzeichenbehafteten Distanzkartenregression mit markierten (2078) und nicht markierten (8665) Karotis-Patch-Slices trainiert, wobei die Konsistenzregulation auf Aufgabenebene das gegenseitige Lernen beider Netzwerke leitete. Dies ermöglichte es, geometrische Beschränkungen aufzuerlegen und unbeschriftete Daten zu nutzen. Zusätzliche Formbeschränkungen wurden basierend auf der Persistenzhomologie angewendet, um die Gefäßwandtopologie zu korrigieren. Die Ergebnisse zeigen, dass das Modell von erzwungenen Formeinschränkungen bei anspruchsvollen Segmentierungsproblemen wie der MRI-Gefäßwandsegmentierung profitieren kann. Das Persistenzhomologiemodul erhöhte den Anteil der korrekten Gefäßwandtopologien um ∼7% steigern. Das endgültige Ziel dieser Arbeit ist es, die Implementierung von DL-basierten Algorithmen in den klinischen Arbeitsablauf zur Beurteilung von Karotis-Atherosklerose-Läsionen voranzutreiben.

Schlüsselwörter: Segmentierung, Erkennung, Semi-überwachtes Lernen, Formanalyse, Gefäßwand, 3D-MERGE

Contents

List of Figures

List of Tables

1 Introduction

1.1 Motivation and Problem Statement

Cardiovascular disease (CVD) is the leading cause of global mortality. It is estimated that 32% of global deaths in 2019 were caused by CVDs and of these deaths, 85% were due to stroke [1]. Atherosclerosis is the dominant cause of ischemic strokes in CVD and approximately 18-25% of all strokes are caused by atherosclerosis in the carotid arteries [2] . Carotid atherosclerosis is characterized by the deposit of plaque in the carotid artery inner wall which leads to luminal narrowing or stenosis. Approximately 5-10% of the population over the age of 65 suffers subclinical carotid stenosis with a degree greater than 50% [2]. As the disease progresses the atherosclerotic plaques may rupture which leads to clinical events such as TIA or stroke [3].

The monitorization of the disease through atherosclerotic plaque assessment may prevent acute ischemic events with the appropriate care response. Traditional plaque evaluation methods focused solely on the measurement of luminal stenosis. Recently, further plaque features such as plaque burden and intraplaque haemorrhage (IPH) have been proposed to characterize the plaque morphology and composition as well as their associated risks [3]. Based on these features the appropriate medical treatment may be selected between non-invasive interventions such as lifestyle changes and pharmacological treatment and more invasive procedures such as carotid endarterectomy (CEA) [4] and carotid stent placement [5]. Furthermore, surgical planning for CEA procedures may also require the characterization of the atheroma plaque [3].

VW-MRI has proven its capability for identifying high-risk atherosclerotic lesions [6]. However, traditional 2D-MRI acquisitions require a long scan time and have low slice resolution. A 3D-MERGE sequence [7] has been developed capable of fast acquisition times, submillimetre isotropic resolution, and large coverage of the carotid artery. This sequence has shown promising results in carotid plaque characterization [8] however, it has a limited application due to the large number of images where the VW must be manually segmented. Manual segmentation is a time consuming and labour-intensive process, therefore, there is a need to develop computational methods capable of automatic VW segmentation. DL networks have recently revolutionized the field of automatic image segmentation [9].

In this thesis, we approach the Carotid Artery Vessel Wall Segmentation Challenge (CAVWSC) 2021 [10]. In this challenge, the task is to automatically segment the bilateral carotid VW from 3D-MERGE [7] MRI images. The aim is to obtain high accuracy and robust segmentation for common carotid artery (CCA), internal carotid artery (ICA) and external carotid artery (ECA) branches in patients with various carotid conditions. For this, the inner

and outer-wall contours are provided but only for a small number of slices. Furthermore, the carotid VW in the axial view is a small and heterogeneous structure, especially in the bifurcation region, which normally requires ROI extraction.

The goal of this work is to propose an automated framework with DL algorithms based on SSL approaches with shape constraints that can provide accurate and robust carotid VW segmentation from 3D-MERGE MRI images. The proposed method may be a first step towards the integration of DL algorithms in the carotid plaque assessment workflow as preprocessing tools for radiologists to monitor and identify high-risk atherosclerotic lesions from VW-MRI.

1.2 Background

1.2.1 Carotid vessel anatomy

The carotid arteries are blood vessels in the neck responsible for the blood supply to the brain, neck, and face. There are two carotid arteries found in both sides of the neck, known as the left and right carotid arteries. As observed in Figure 1.1, each carotid artery is composed of a CCA that branches into the ICA and ECA. The right CCA originates from the brachiocephalic trunk whilst the left CCA branches directly from the aorta. Both CCAs split into the ICA and ECA at the level of the thyroid cartilage (C4), this region is known as the carotid bifurcation (BF) region. Directly superior to the BF region the ICA experiences a slight dilation, known as the carotid bulb or sinus. The ECA supplies blood to the face and neck, it has a vertical course which lies more superficial than the ICA. As it ascends it gives rise to six branches: the superior thyroid artery, the lingual artery, the facial artery, the ascending pharyngeal artery, the occipital artery, and the posterior auricular artery. The ECA ends in the parotid gland where it divides into the temporal artery and the maxillary artery. The ICA supplies the anterior part of the brain and the eye. As it rises into the cranial cavity it does not give off any branches. Its trajectory is first lateral with respect to ECA and then passes medially and deeper from it [11].

The carotid VW is composed of three layers: the *tunica intima*, the *tunica media*, and the *tunica adventitia*. The intima or lumen wall is the inner wall directly in contact with the blood flow. It is composed of an endothelial monolayer that has a significant role in the formation of the atherosclerotic plaque. The media or middle layer is composed of smooth muscle cells surrounded by a matrix of collagen, elastin, and proteoglycans. The adventitia or outer wall is the external layer mostly composed of collagen which provides support [13].

The reported diameters of CCA, ECA and ICA in the literature vary with different acquisition techniques and measurement methods [14, 15]. In a recent study by Cobiella et al. (2021) [16], where external diameters were measured ex-vivo, the average diameter of CCA reported was 0.97 ± 0.14 cm (n=103 cases), the average for ICA was 0.74 ± 0.16 cm (n=139 cases) and the average for ECA was 0.49 ± 0.08 cm (n=206 cases).
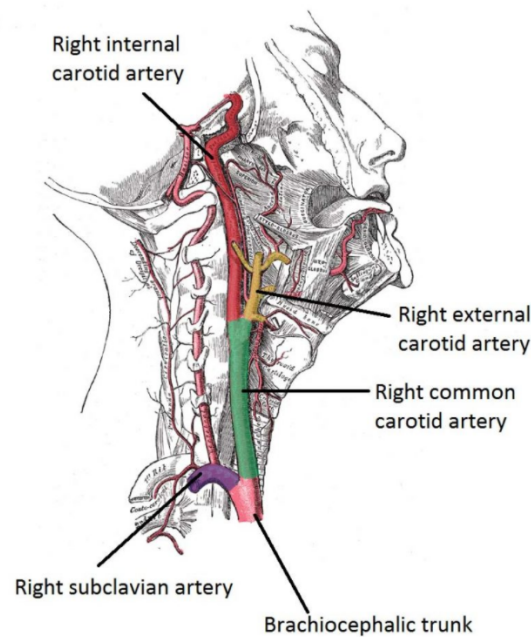
Figure 1.1: Anatomy of the carotid artery [12].

1.2.2 Carotid atherosclerosis

Atherosclerosis is a chronic inflammatory disease caused by the accumulation of low-density lipoprotein (LDL) cholesterol and calcium on the inner arterial wall, this is otherwise known as plaque. As the disease progresses the plaque can continue to grow causing luminal narrowing and blocking the blood flow of the vessel [16]. Figure 1.2 illustrates the progression of atherosclerosis. The first visible sign of atherosclerosis is known as the 'fatty streak' which consists of cholesterol and lipid-containing macrophage cells. The response-to-injury hypothesis states that this initial event is caused by lesions in the endothelium which produce an inflammatory response [17]. Endothelial dysfunction has been observed to occur more likely in arterial regions subjected to a lower shear stress such as curves and bifurcations. Fatty streaks are not symptomatic and in many cases regress back to normal conditions, however, they can also progress into the formation of fibrous plaques [18]. The progressive accumulation of lipids and the infiltration of smooth muscle cells in the intima leads to the formation of a plaque which narrows the vessel lumen. Atherosclerotic plaques can be defined by the risk of rupture they have. Stable plaques cause vascular remodelling where the lumen is only partially occluded and are asymptomatic. They are characterized by a thick fibrous cap (FC) which prevents plaque rupture and a small necrotic core. Furthermore, they may suffer calcifications which has been observed to stabilize the plaque. On the other hand, unstable plaques have thinner fibrous caps and larger necrotic cores that are more likely to rupture. Plaque rupture results in the discharge of prothrombotic factors from the necrotic core into the vessel lumen which leads to thrombus formation and clinical events including myocardial infarction, unstable angina, stroke and sudden cardiac death [17].
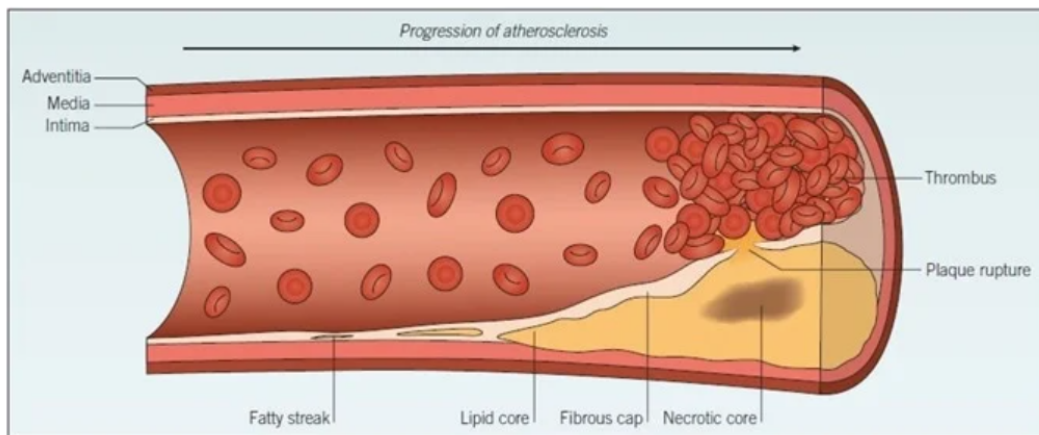
Figure 1.2: Atherosclerosis progression [19].

In the carotid artery, the atherosclerotic plaques generally develop at the bifurcation region leading into the ICA and ECA. As illustrated in Figure 1.3, the ICA is mostly affected involving the regions of the carotid bulb and posterior wall. Furthermore, carotid plaques also extend to the distal portions of the CCA. Besides the known risk factors, the vessel geometry and fluid dynamics play an important role in the development of atherosclerosis. Hemodynamic stress conditions such as low shear and disturbed flow have been reported to cause endothelial dysfunction which leads to plaque formation [3].

The risk factors for the development of carotid atherosclerosis are similar to other vascular diseases. Some of the risk factors include obesity, cigarette smoking, age, diabetes, hypertension and advanced glycation end products (AGE) [3].

Medical treatment of carotid atherosclerosis is dependent on the degree of stenosis and on the symptomatology of the disease. The treatment of asymptomatic patients includes lifestyle intervention and pharmacological agents. Several lifestyle changes could slow the progression of carotid atherosclerosis such as not smoking, ingesting foods with low saturated fats, daily physical exercise and limiting alcohol consumption. Asymptomatic patients with a degree of stenosis lower than 50% should receive intensive medical treatment through pharmacological agents. These include antihypertensive agents, agents that lower the AGE products and increase soluble form of receptor for advanced glycation end products (sRAGE) and lipid lowering agents. In symptomatic patients or asymptomatic with a degree of stenosis above 60% other treatments should be considered such CEA [4]. CEA consists of surgically removing the carotid plaque and repairing the VW. Carotid artery stent placement [5] is a procedure less invasive than CEA and consists of inserting a small mesh cylinder (stent) at the site of the narrowed artery with the inflation of a balloon attached at the end of a catheter. Once it is placed, the balloon is removed, and the stent is left in place which keeps the artery open [5]. In asymptomatic patients, non-invasive medical management of the disease has shown comparable success rates with respect to CEA in reducing the risk of stroke with higher cost-effectiveness [3].
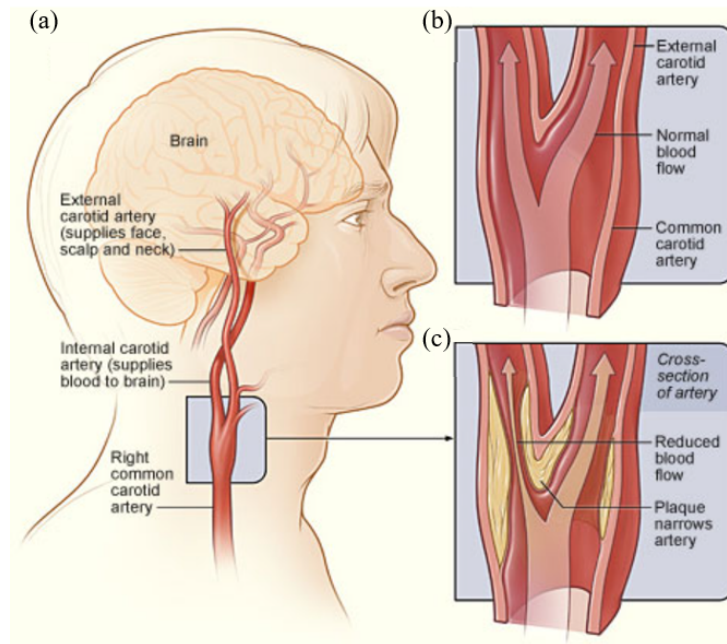
Figure 1.3: (a) Location of right carotid artery. (b) Normal carotid artery. (c) Artery with plaque build-up [20].

1.2.3 Carotid vessel wall imaging techniques

The identification of carotid atherosclerosis has been conventionally performed by measuring the degree of luminal stenosis through imaging techniques such as sonography, catheter-based angiography and, more recently, Computed Tomography Angiography (CTA) and MRA [21]. However, histopathologic studies demonstrated that there are considerable differences between plaques with the same degree of stenosis which could increase the risk of stroke [2]. In addition, it was observed that mild degrees of stenosis may still result in cerebrovascular events [22]. This led to a more focused research on the features of the plaque associated with ischemic events [23]. Advanced imaging techniques such as Multidetector Computed Tomography (MDCT), Ultrasound (US) and high-field MRI have enabled the identification of plaque features associated with high-risk plaques. Rather than simply measuring the percentage of luminal stenosis, advanced imaging methods enable the identification of "vulnerable plaques" which are prone to rupture and cause acute clinical events [2]. Recent research works have focused on investigating the optimal approach for high-risk plaque imaging [24]. Carotid vessel wall imaging (VWI) techniques enable the visualization of the location and extent of the atherosclerotic plaque and could be useful for assisting surgical planning before CEA or carotid stent placement. Furthermore, carotid VWI could also aid in the decision making of borderline cases. For example, if a large lipid-rich necrotic core with IPH is detected inside the carotid plaque of a patient with repeat strokes it could lead to considering the CEA procedure. In subclinical cases VWI methods could also be useful in monitoring the progression of the atherosclerotic plaque [2].

Previous research works [2, 23] have introduced new concepts for the evaluation of atherosclerosis severity. First, that the degree of stenosis is a weak marker for atherosclerosis as it does not fully characterize the volume and extension of the plaque. Second, a set of plaque features identifiable with imaging techniques are closely linked to an increased risk of stroke regardless of the degree of stenosis. These features can be accurately predicted thanks to the advancement of imaging techniques that target the VW instead of only the vessel lumen [2]. The lumen morphology of the carotid plaques can be smooth or present irregularities. Surface irregularity is considered as a risk factor for embolism and TIA or stroke [25]. Another important plaque feature is IPH. IPH is a commonly employed feature to identify vulnerable plaques and strong associations have been observed between the presence of IPH in the plaque and cerebrovascular accidents [26]. Furthermore, it can be used to give a measure of plaque progression. The association of intraplaque neovascularization with IPH and plaque remodelling has also been recently studied [2]. The FC which separates the LRNC of the plaque from the lumen is one of the most important risk predictors to be considered [2]. Vulnerable plaques are characterized by thin FC and a large percent area of LRNC and are highly associated with cardiac events. Lastly, plaque volume or burden has demonstrated to play a role in determining the 'vulnerability' of the plaque. An increased plaque volume is related to higher risk of cardiovascular events and some authors hypothesized that it may be a better indicator than the degree of stenosis. [2].

Carotid VWI techniques can be used to extract plaque features with the aim of predicting the risk of cardiac events and monitoring the disease progression. The state-of-the-art non-invasive imaging modalities used for VWI are US, MRI and Computed Tomography (CT). US imaging can be used to evaluate the plaque based on echogenicity factors. For example, hypoechogenic elements are associated with LRNC and hyperechogenic regions or shadowing may indicate calcifications. However, the visualization of deeper tissues may be limited due to the acoustic shadowing caused by the calcified plaques. CT scans can provide high resolution imaging of the carotid wall which renders it a promising tool for the quantification of plaque volume. Furthermore, the recent advancements in MDCT allows it to differentiate between fatty and calcified plaques which are associated with different risks of rupture. However, it presents several limitations such as the radiation dose delivered to patients, the risk associated when injecting contrast materials and the limited contrast in fat tissue [2].

In high-resolution MRI, the high field strength (1.5T-3T) has improved signal-to-noise ratio and enables the evaluation of plaque features other than the degree of stenosis. Multi-contrast carotid MRI (including T1W, T2W, proton density and TOF sequences) provides comprehensive information of plaque components such as the FC, LRNC, calcifications and IPH [27]. MRI has demonstrated high sensitivity in the detection of IPH which for some authors converts it in the leading imaging technique for the detection of vulnerable plaques [2]. In addition, it does not employ ionizing radiation which is harmful for the patient [8].

Figure 1.4 shows a comparison of high-resolution MRA, multi-contrast MRI and ex-vivo histological slide of a carotid VW with plaque. As can be observed, VW-MRI is able to differentiate between loose matrix (purple outline) and LRNC (yellow outline) which allows the identification of vulnerable plaques. Furthermore, as can be seen in the image, plaque evaluation requires the prior step of VW segmentation. VW segmentation involves defining the contour of the inner wall or lumen (red) and the outer-wall (blue) [28].
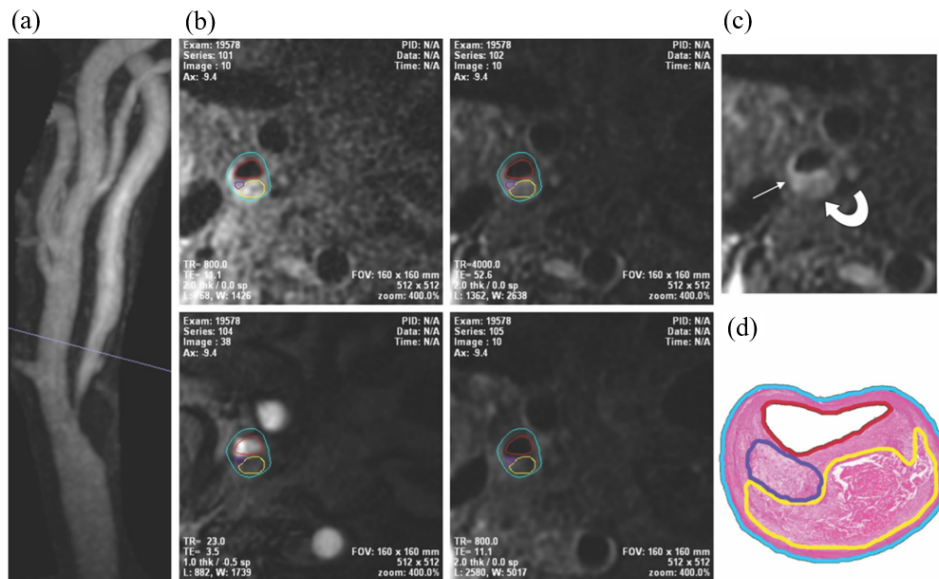
Figure 1.4: Comparison of high-resolution MR angiography, 3T multi-contrast MRI and histological slide of carotid vessel with plaque. (a) MRA. (b) T1W, T2W, TOF and contrast-enhanced T1W. (c) T2W where region of LRNC is dark (curved arrow) and loose matrix is bright (arrow). (d) Histologic slide with LRNC (yellow outline), loose matrix (purple outline), inner-wall (red outline), outer-wall (blue outline) [6, 28].

3D-MERGE sequence

High-resolution black-blood MRI has emerged as a promising technique for atherosclerotic plaque characterization [29, 30]. In this sequence the signal from flowing blood is suppressed, which avoids pulsation or ghosting artifacts common in bright-blood techniques that cause an overestimation of stenosis. Furthermore, the lack of intraluminal signal allows to clearly delineate the carotid VW [8]. For this reason, it is the predominant technique for VW evaluation, and it has been employed for plaque feature analysis such as plaque burden in several clinical trials [8, 31]. On the other hand, black-blood MRI is mostly employed as a 2D sequence where the longitudinal coverage is limited to approximately 15 mm above and below the carotid bifurcation. This coverage may not be sufficient to visualize the full extent

of the carotid plaque atheroma. For example, the plaque length is an important risk factor that may not be completely quantified with the limited carotid coverage provided by traditional 2D-MRI sequences [8]. In addition, plaque burden quantification can be used to monitor the disease progression or the response to treatment however, an accurate measurement requires inner and outer-wall segmentation in images with high resolution in all three directions. 2D-MRI techniques have limited slice resolution as opposed to 3D techniques [7].

A 3D black-blood imaging technique known as 3D-MERGE [7] has recently been implemented to characterize carotid plaque composition and morphology. It consists of a 3D isotropic sequence with Motion Sensitized Driven Equilibrium (MSDE) preparation for maximum blood suppression and a gradient echo sequence with short pulse repetition times for a faster 3D acquisition. 3D-MERGE sequences can acquire images with entire coverage of the cervical bilateral carotids and with submillimetre isotropic resolution ($\sim$0.7 mm^3) with a field strength of 3T. Furthermore, they have faster acquisition scan times ($\sim$2 minutes) compared with other 3D scan techniques therefore ensuring patient compliance and reduced motion artifacts. The high slice resolution, the expanded coverage of the carotid and the short scan times have rendered it a promising sequence for characterization of carotid plaques in clinical applications. Figure 1.5 shows an example of a 3D-MERGE acquisition of the carotid artery.
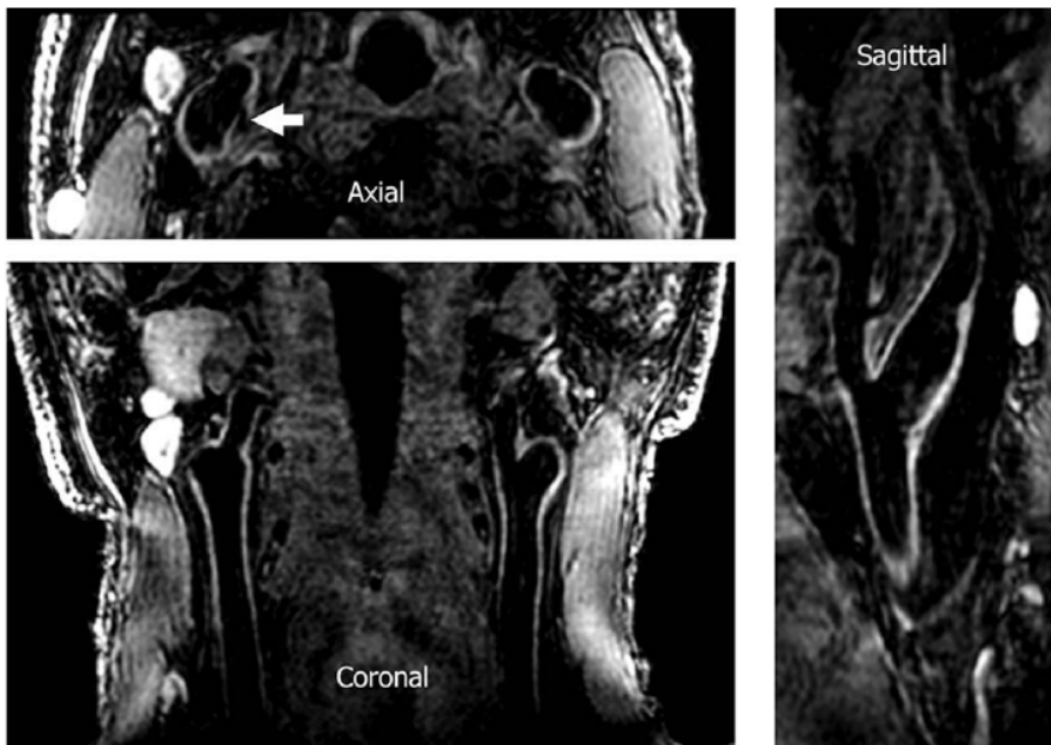
Figure 1.5: Isotropic 3D-MERGE sequence acquisition in three orthogonal planes. Calcification of right carotid plaque is visible (arrow) [7].

1.3 Structure

The rest of this manuscript is organized as follows. In Section II, we provide the state-of-the art of VW DL segmentation methods. The methodology employed in our proposed framework and the experimental data used are described in Section III, followed by the results and discussion in Section IV. In Section V we draw the conclusions and in Section VI we present some of the limitations and future work.

2 State of the art

During the years, semi-automated and automated computational methods have been proposed for VW image segmentation. In the field of VW-MRI segmentation Yuan et al.(1999) [32] applied active contour models for vessel boundary detection whilst Ladaks et al.(2001) [33] applied discrete dynamic contours for outer-wall and lumen segmentation in a slice-by-slice mode. Underhill et al. (2006) [34] applied active shape models in black-blood MRI sequences. However, the above methods required manual interaction for contour initialization. To increase the accuracy of segmentation, Unal et al.(2008) [35] proposed the introduction of shape priors to an edge detection method in intravascular ultrasound (IVUS) images, however, parameter tuning was not trivial. Liu et al.(2012) [36] extended the segmentation task to 3D MRI images with a B-spline snake evolution model that propagated outer and inner wall segmentations between slices, the main limitation was that snake models may fail in low quality images or severely diseased areas. 3D deformable models have also been applied for vessel segmentation. For example, Van´t Klooster et al.(2012) [37] combined 3D TOF-MRA data with 2D vessel images and Arias-Lorza et al.(2015) [38] employed multiple sequences of MRA (proton density-weighted (PDW) black-blood sequences, phase contrast, PDW-echo planar) for 3D graph-cut optimization. More recently, Luo et al.(2019) [39] proposed to modify the traditional level set method with a double adaptative threshold for a semi-automatic segmentation of TOF-MRA carotid artery images. In general, the above methods required extensive human input in the form of either contour initialization, parameter tuning or multiple image registration.

With the advent of deep learning, automatic computational methods for image segmentation have experienced a major break-through. Multi-layer networks that apply convolutional operations, otherwise known as Convolutional Neural Networks (CNNs), have shown promising performance in image classification tasks [40]. In addition, Long et al.(2015) [41] proposed a CNN network with Fully Convolutional Networks (FCNs), capable of efficient pixel-wise segmentation. Based on FCNs, Ronneberger et al.(2015) [42] proposed a U-shaped network with skip connections between downsampling and up-sampling paths, otherwise known as U-Net, achieving promising results in biomedical applications.

Several variants of FCNs and U-Net have been recently proposed for carotid VW segmentation [9]. For example, Shi et al.(2019) [43] applied a 2D U-Net that combined information of lumen and outer-wall with a multi-class loss function in MRI image segmentation. Wu et al.(2019) [44] proposed a 2.5D U-Net with a triple dice loss function that leveraged spatial information between adjacent slices achieving promising results. However, both methods required user interaction for artery localization and ROI extraction. More recently, Chen et al.(2019) [45] proposed a CNN with polar conversion for the task of CCA and ICA segmentation. Furthermore, they automatized ROI extraction with the integration of an object

detection algorithm coupled with tracklet refinement to the segmentation framework.

Carotid VW segmentation methods have also been extended to 3D applications. For example, Zhu et al.(2021) [46] recently proposed a 3D U-Net with patch and global level training. On the other hand, 3D methods require fully labelled annotations in every image slice which may be challenging in medical image applications.

Recent research works have focused on the introduction of prior knowledge to generate anatomically plausible medical image segmentations. Embedding prior knowledge such as shape in deep learning networks has demonstrated increased accuracy and robustness of segmentation [47].

In this field, Xue et al.(2020) [48] proposed the integration of level set functions that encode implicit shape representations such as SDMs to the task of image segmentation. By converting the segmentation task into SDM regression, they showed superior segmentation performance and better smoothness and continuity in shape in the task of multi-organ segmentation. Navarro et al.(2019) [49] proposed the integration of SDM prediction as an auxiliary task for multi-task learning segmentation. Different from level set functions, Clough et al.(2019) [50] proposed the integration of topological priors to the loss function through the concept of PH. The topological loss function could be applied both during training and at a post-processing stage without the requirement of ground-truth annotations and promising results were achieved in the task of left ventricle myocardium segmentation.

Medical image annotations are often expensive to acquire, especially in 3D volumetric images. To reduce labelling costs, SSL methods aim to learn from a limited amount of labelled data and a large set of unlabelled data directly. Recently, SSL methods have successfully incorporated unlabelled data by enforcing the consistency of unlabelled data predictions through a regularization term in the loss function [51, 52]. Furthermore, the recent integration of SDM regression to SSL frameworks has achieved state of the art results in medical image segmentation [53–55]. In this setting, Luo et al.(2020) [53] proposed a dual-task consistency framework with level set functions whilst Li et al.(2020) [54] enforced consistency of SDM unlabelled predictions through an adversarial loss. More recently, Zhang et al.(2021) [55] proposed a dual-task mutual learning framework that combined two individual networks, trained for binary mask segmentation and SDM regression, through a cross-task consistency loss. The aim was to leverage information from unlabelled data whilst imposing geometrical shape constraints and promising results were achieved for left atrium MRI segmentation.

In this thesis, we propose an automated carotid vessel wall segmentation workflow for 3D-MERGE images. First, a state-of-the-art object detection algorithm [56] was employed for artery localization and automatic ROI extraction. Inspired by the works of Zhang et al.(2021) [55] and Clough et al.(2019) [50] we propose a VW segmentation framework that imposes shape constraints at two levels, both during training and at a postprocessing stage, through SDM regression and topological priors. To our knowledge, shape constrained convolutional networks and SSL methods have not been previously applied to the task of carotid VW-MRI segmentation.

3 Materials and Methods

3.1 Materials

3.1.1 MR Images

The dataset employed in this thesis was provided by the CAVWSC 2021 [10]. The data acquisition was performed by the CARE II [31] multicentre study, where patients with recent ischemic stroke or transient ischaemic attacks were enrolled as subjects. Each subject underwent bilateral carotid arteries MRI acquisition with standardised protocols. Details of the dataset are shown in Table 3.1. The dataset consists of fast 3D-MERGE [7] sequences. All scans were acquired on a Philips Achieva 3T scanner with an 8-channel phase array coil. The field-of-view (FOV) was of 250 x 160 mm^2. The 3D-MERGE sequence was acquired in the coronal plane and the scan parameters were optimized to image the entire longitudinal extent of the carotid artery covered by the coil. The scans were acquired with a submillimetre and isotropic resolution of 0.7 x 0.7 x 0.7 mm^3 and were zero-resampled to 0.35 x 0.35 x 0.35 mm^3. Figure 3.1 illustrates a sample 3D acquisition of the carotid artery. As can be observed, the whole neck was covered with the 3D-MERGE image scan, including the CCA, ICA and ECA.

Each scan consisted of a volumetric image resliced in the axial plane with 720 slices per volume and average size of 100 x 720 (height x width) pixels. The axial slices were interpolated to 720 x 720 pixel matrix size with zero padding. As can be observed in Figure 3.1 the quality of the image slices varied along the slice depth due to the coronal acquisition scan with only the several hundreds of slices in the middle (approximately 200-500) with sufficient image quality for carotid VW evaluation.

Two sets of data were provided the challenge, both from the same CARE II study and with the same imaging parameters. The training and testing set consisted of 25 patients.

Table 3.1: CARE-II dataset parameters

Dataset name	Acquisition resolution	Interpolated resolution (mm^3)	FOV* (mm^2)	Slices per scan	Image size	Interpolated image size
CARE-II	0.7 x 0.7 x 0.7	0.35 x 0.35 x 0.35	250 × 160	720	100 x 720	720 x 720

*FOV: field-of-view
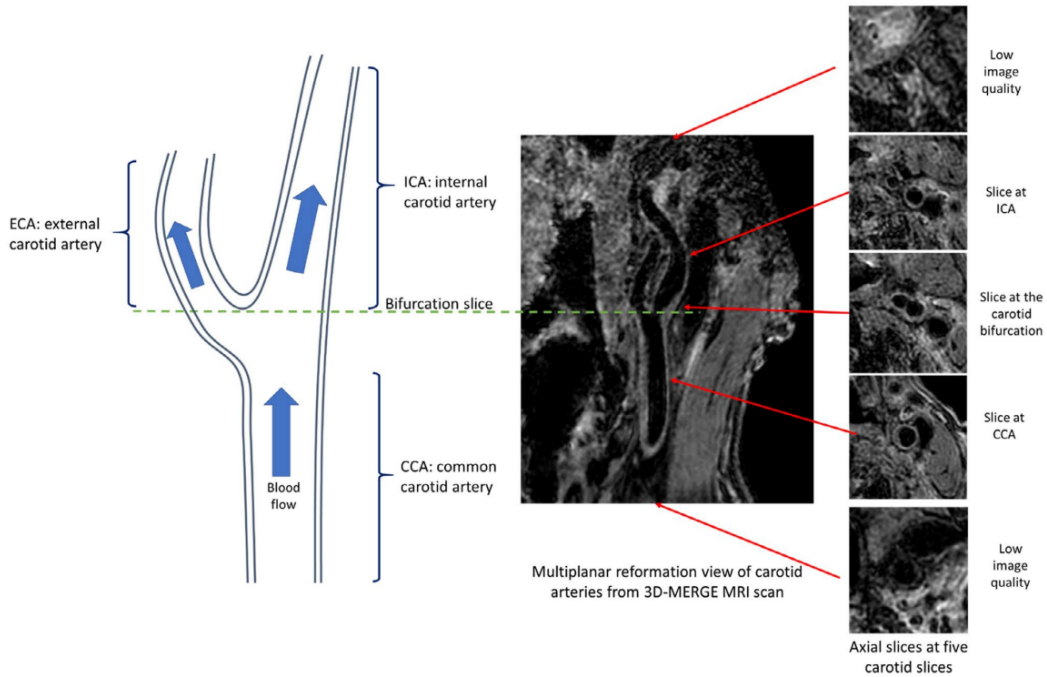
Figure 3.1: Illustration of 3D-MERGE images from CARE II study in different carotid regions [57].

3.1.2 Human labelling

The lumen and outer carotid VW were manually traced by trained reviewers on axial slices with the aid of a custom designed carotid VW annotation Computer-Aided System for Cardiovascular Disease Evaluation (CASCADE) software package [58].

In the training set, the annotations included the CCA, ECA and ICA for left (L) and right (R) arteries and were given in separate slices. The challenge also provided the location of the bifurcation slice to differentiate between the CCA and ICA/ECA. Figure 3.2 illustrates the lumen (red) and outer-wall (blue) contours for a sample slice in the original provided image (Figure 3.2(a)) and with a x9 scale (Figure 3.2(b)). Contours were provided as coordinates considering the top-left corner of the image as the origin.

For each scan or patient, only the slices with enough image quality for VW diagnosis were annotated, in total only 20% of the slices [10]. Consequently, the 100-200 slices of the outer-most edges were unlabelled due to the low image quality provided by the coronal scan. In addition, in several cases, the annotated slices were sparsely labelled which meant that not all carotid types were annotated for one particular slice. In total, the number of annotated slices in the dataset was 2336 and the number of annotated carotid structures extracted was 2670. It must be remarked that each patient had a different number of annotations for each carotid branch. The number of unlabelled carotid arteries with experimental value was computed as the number of carotid structures, within first and last labelled slice location, with no carotid label. In total, 8665 unlabelled carotid structures were identified.
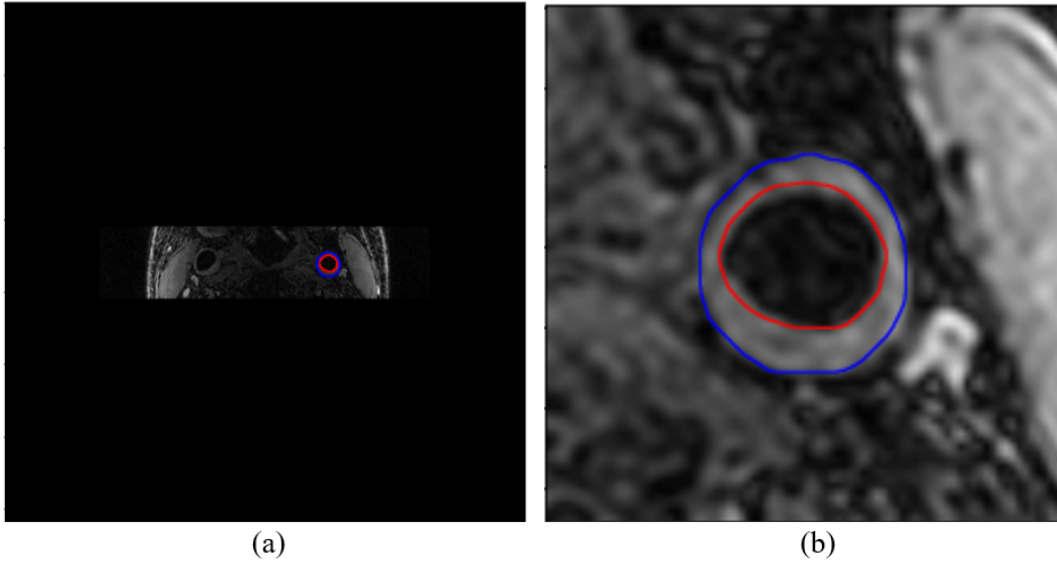
Figure 3.2: (a) Sample image slice (720x720) with lumen (red) and outer-wall (blue) annotations for ICA-L. (b) Lumen and outer-wall annotations with x9 scale.

3.2 Methods

3.2.1 Theoretical concepts

Deep Learning for object detection

Object detection is an important task related to the field of computer vision. It involves the detection and localization of object instances belonging to a certain class, in digital images and videos [58]. Object detection can provide valuable information related to many applications such as human behaviour analysis [59], autonomous driving [60] and face recognition [61]. More recently, object detection algorithms have also been proposed for medical applications, for example, for medical image processing [62–64].

The goal of object detection is to develop computational tools that can determine the position of an object in a given image (object localization) and the category it belongs to (object classification) automatically. The object localization task involves drawing a bounding box around the object of interest and locate it within an image. However, this may become challenging as the number of object occurrences in a sample can range from 0 to many and this information is not known beforehand, therefore the length of the output variable is not constant [58].

Object recognition requires the extraction of meaningful and semantic representations of the image, otherwise known as feature vectors. Traditional approaches involve the use of feature descriptors such as histograms of oriented gradients, however these must be manually designed and tailored for background and illumination conditions. With the advent of deep learning and CNN, multilayer convolutional operations have been applied to object detection

methods for adaptative and robust feature extraction. This has led to a rapid expansion of object detection algorithms, known as CNN-based object detection methods [58].

CNN-based object detectors can be categorized in two main types: one-stage and two-stage detectors. Two-stage detectors first require a region proposal network (typically a region CNN [65]) and based on the features extracted from the proposed regions, object classification and bounding box regression is performed with a separate architecture. Two-stage methods normally achieve the highest detection accuracy however their inference time is significantly slower. The most representative models are faster Region based Convolutional Neural Network (R-CNN) [66] and mask R-CNN [67]. On the other hand, one-stage object detection methods predict bounding boxes over the entire image, without the requirement of a region proposal network, which significantly reduces inference time. One of the most popular one-stage object detectors is the You only look once (YOLO) [68] algorithm.

YOLO [68] is a non-region-based object detection method that utilizes the entire feature map to predict the object category and bounding box. Therefore, a single unified architecture can perform global regression and classification . YOLO pipeline is very simple; it can be trained on full images by dividing the input into S x S grids and each grid is responsible of detecting the object whose centre it contains. In this manner, the algorithm can 'only look once' at the image to output the bounding box predictions. Using the feature information from the whole image reduces the error rate compared with local detection. Figure 3.3 shows an illustration of this framework. The main advantage of YOLO algorithm is its high inference speed due to performing object detection in one stage, however, this also comes at the cost of underperforming in the detection of small and irregular objects. The latest developed algorithm, YOLO version 4 (YOLOv4) [56] incorporates additional methods to boost the detection accuracy limitations from YOLO whilst retaining the high inference speed.
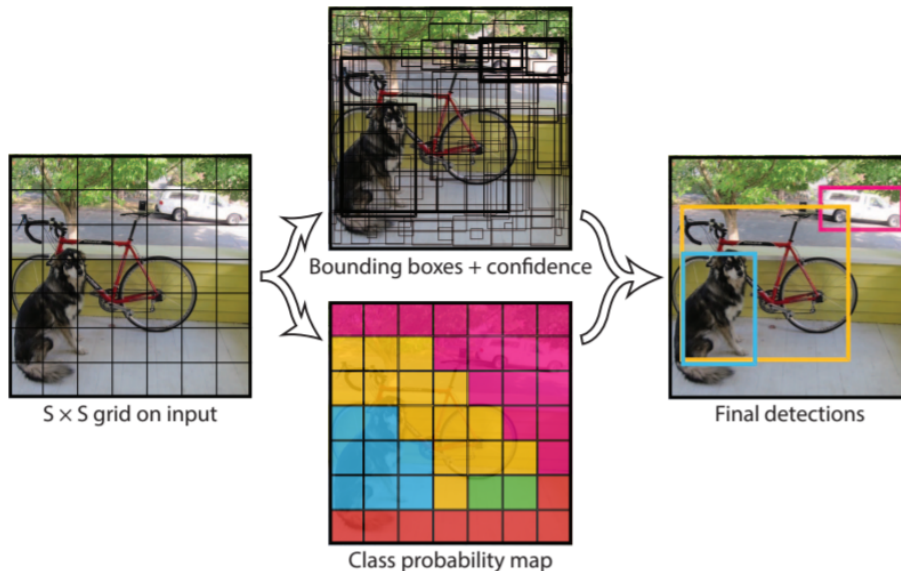
Figure 3.3: YOLO one-stage object detection framework [68].

In this thesis, we employ a one-stage CNN-based object detection algorithm known as YOLOv4 [56] for artery lumen localization. YOLOv4 is characterized by real-time inference speed with high accuracy of detection and has shown promising results in several computer vision applications such as vehicle detection [69] or aircraft segmentation [70]. Recently, it has been employed for medical image processing applications such as lesion detection [63, 64].

Loss function

The optimization of CNN-based object detection algorithms is based on the minimization of a loss function that computes the error between the predicted and ground-truth bounding box. In the following section, $B = (x, y, w, h)$ is the predicted bounding box and $B^{gt} = (x^{gt}, y^{gt}, w^{gt}, h^{gt})$ is the ground-truth bounding box.

YOLOv4 employs a loss function based on the Intersection over Union (IoU), known as the complete IoU (cIoU) [71]. The cIoU loss function is defined in Equation 3.1

$$L_{CIoU} = 1 - IoU + \frac{\rho(\mathbf{b}, \mathbf{b}^{gt})}{c^2} + \alpha v \tag{3.1}$$

where $\mathbf{b}$ and $\mathbf{b}^{gt}$ refers to the central points of B and B^{gt}. ρ is the Euclidean distance and c is the diagonal length of the smallest box encompassing B and B^{gt}. The IoU is a widely used accuracy metric for object detection and is defined in Equation 3.2

$$IoU = \frac{|B \cap B^{gt}|}{|B \cup B^{gt}|} \tag{3.2}$$

v measures the consistency aspect ratio as defined in Equation 3.3

$$v = \frac{4}{\pi^2} (\arctan \frac{w^{gt}}{h^{gt}}) - (\arctan \frac{w}{h})^2 \tag{3.3}$$

α is a trade-off parameter that gives higher priority to non-overlapping bounding boxes during regression. Equation 3.5 defines the parameter α.

$$\alpha = \frac{v}{(1 - IoU) + v} \tag{3.4}$$

As observed, different from the standard IoU loss, the cIoU loss considers three geometric properties: the overlap area, the central point distance and the aspect ratio of prediction and ground-truth bounding box which leads to faster convergence and performance than IoU loss.

Deep learning for semantic image segmentation

Semantic image segmentation is a pixel-wise classification task that assigns a category to each pixel in an image. For example, in binary image segmentation the pixels are classified in two categories: pixels belonging to the target object (1) and pixels belonging to the background (0) [72].

In recent years, DL methods have revolutionized traditional computer vision problems such as automated image segmentation [72]. In this field, CNN networks have experienced major success due to their ability of learning hierarchical image feature representations during training. For end-to-end semantic segmentation FCNs [41] were first proposed by replacing fully connected layers with convolutional layers in classification models and with the introduction of an up-sampling path to generate dense predictions. FCNs introduced encoder-decoder architectures for image segmentation however they often failed to consider global spatial information. To solve this, an improved FCN variant was proposed known as U-Net [42]. U-Nets are CNNs with symmetric encoder-decoder architectures and skip connections that propagate features from the down-sampling to the up-sampling path. In the encoder, convolutional operations are applied to extract contextual information whilst the decoder is an expanding path that localizes the detected patterns such as contour and boundaries in the final resolution image. U-Net networks have achieved state-of-the-art results in medical image segmentation tasks [42, 72].

Loss function

The optimization of DL-based algorithms such as CNNs for image segmentation is normally based on the minimization of a supervised loss function that computes the error between the output of the algorithm (prediction) and the 'true' or expected value (ground-truth).

The soft Dice loss function was first proposed by Milletari et al. [73]. It is based on the dice coefficient (defined in chapter 4 - Equation 4.2) which measures the similarity between the predicted (p) and the corresponding ground-truth (y) probability map for each pixel i. The dice loss function is defined in Equation 3.5

$$L_{dice}(y, p) = 1 - \frac{\sum_i^N y_i p_i}{\sum_i^N y_i + \sum_i^N p_i + \sigma} \tag{3.5}$$

where σ is a smoothing factor equal to 0.00001.

Activation function

Activation functions are employed in neural networks to transform the weighted sum of the inputs at each node of a layer into an output signal. Most activation functions are non-linear to introduce non-linearities to the network design [74].

Activation functions can be applied in hidden network layers (layers that receives input from previous layer and pass an output to the next layer) and output layers (layers that output the final prediction) to limit the range of values of the output probability map. The choice of activation function normally depends on whether it is applied on a hidden or output layer, and in the output layer it will depend on the type of the prediction that the model requires [74].

The Rectified Linear Unit (ReLU) function is an activation function commonly used in hidden layers. Equation 3.6 shows the mathematical definition.

$$f(x) = max(0, x) \tag{3.6}$$

Figure 3.4(a) shows a graphical representation of the ReLU function. As observed when the input is negative the returned value is 0, otherwise the input value is returned. This function is widely employed as it is less susceptible to vanishing gradients [74].
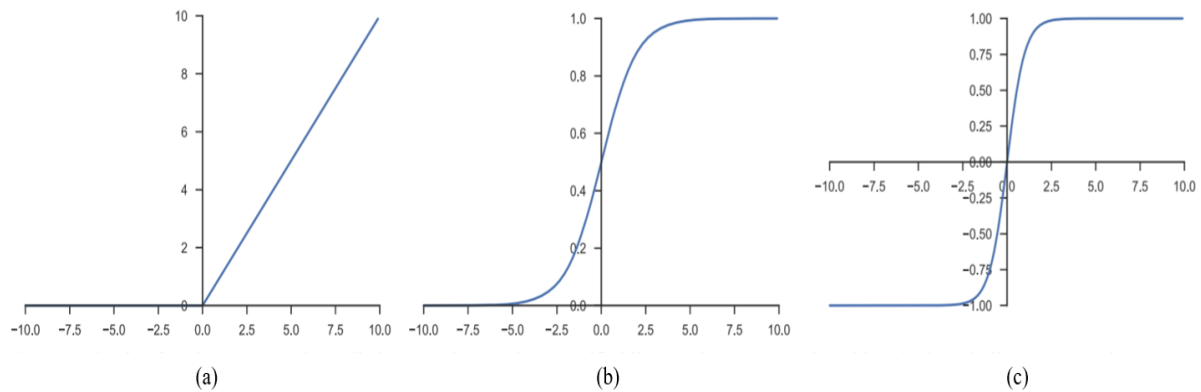
Figure 3.4: Graphical representation of (a) ReLU function, (b) sigmoid and (c) tanh [75].

Output activation functions are activation functions applied to the last layer of the network to limit the range of output probability maps into a set of finite values.

The sigmoid or logistic activation function is the standard output activation function for binary classification and segmentation with neural networks. The sigmoid function is defined in Equation 3.7.

$$f(x) = \sigma(x) = \frac{1}{1 + e^{-x}} \tag{3.7}$$

As can be observed in Figure 3.4(b) , the sigmoid activation function returns values in the range of [0,1], therefore it is commonly employed to output a prediction probability map where each pixel contains the probability of belonging to a binary class. The predicted probability map can be then converted to a binary mask through the application of a threshold (i.e., 0.5).

The hyperbolic tangent (tanh) is a logistic function commonly employed as an output activation function in neural networks. Different from the sigmoid function it is symmetric around the origin. The tanh function is defined in Equation 3.8

$$f(x) = \tanh(x) = 2\sigma(2x) - 1 \tag{3.8}$$

where σ corresponds to the sigmoid function.

Tanh returns outputs values in the range [-1,1], as shown in Figure 3.4(c). Due to this property, it has recently been proposed as an output activation function in the prediction of normalized SDMs (also in the range [-1,1]) [48].

Limitations of the U-Net

The U-Net has become the benchmark neural network for medical image segmentation applications, however there are some limitations associated with the architecture and loss functions that are utilized [47].

The Dice and cross-entropy loss are commonly used in combination with U-Nets for medical image segmentation, however, they ignore high-level features such as shape or topology of the object of interest. Furthermore, these losses penalize all mistakes equally without geometrical or anatomical considerations which could lead to uneven target boundaries and outliers in the output segmentation. The architecture of the vanilla U-Net also does not integrate any anatomical or contextual constraints which further results in a lack of spatial awareness in the final predictions. For this reason, recent research works have focused on the integration of knowledge priors such as shape into the U-Net or its 3D variant, for medical image segmentation [53–55]. Embedding prior knowledge could also lead to increased robustness and accuracy of the network as well as generating anatomically plausible segmentations [47].

Deep neural networks such as the U-Net normally require large amounts of fully annotated data for successful training. However, in medical image segmentation applications large and fully labelled datasets are normally expensive and difficult to acquire. When given the complexity of the model, the amount of training data is not large enough over-fitting may occur where the model does not generalize well for unseen data [76].

Several regularization techniques have been proposed to improve the generalization ability of neural networks and avoid over-fitting. Dropout consists of randomly disconnecting layer outputs with a probability p during training. This can be seen as a form of noise injection which increases the robustness of the model. Batch normalization standardizes the layer

inputs to a Gaussian distribution (standard deviation = 1 and mean = 0) which also introduces noise to the network training. Both techniques make the model more robust and reduce its variance. Data augmentation is also a useful strategy to reduce the variance of the model. By applying transformations to the training data such as geometrical or noise perturbations we can generate new training samples for our dataset [76].

Recently, other strategies have been proposed to regularize the training procedure and improve the generalization ability of the model. In the following section we describe three concepts which form the basis of the proposed VW segmentation framework: Semi-supervised Learning [77], Deep Mutual Learning [78] and shape prior regularization [48, 50, 55].

Semi-supervised learning

Semi-supervised learning (SSL) approaches aim to increase the model performance by leveraging information from a small amount of labelled data and a large amount of unlabelled data [77].

In SSL we are provided with a dataset D consisting of a subset of labelled samples D_l and a subset of unlabelled samples D_u were normally the portion of labelled samples is significantly smaller compared with the unlabelled samples. The aim of SSL is to leverage unlabelled samples D_u to increase the performance of a model $f(\theta)$ with trainable parameters θ. For example, D_u might provide additional information about the structure of data distribution $p(x)$ which could lead to a better estimation of the decision boundary between classes, this can be further observed in Figure 3.5 [77].

One of the first SSL strategies consisted of self-training methods [79] where a model, first trained with labelled samples, was used to generate pseudo-labels for unlabelled data. The pseudo labels were then added to the training set for the next iteration. In the deep SSL setting, Bai et al. (2017) [80] proposed an iterative framework for cardiac image segmentation were in each iteration pseudo labels were generated and were refined by a Conditional Random Field (CRF), then the refined pseudo-labels were used to update the network. However, pseudo-labelling methods rely on the high confidence of pseudo-labels and often annotations generated by the model may be noisy and have a detrimental effect on subsequent iterations. Another approach is to utilize unlabelled data through adversarial training. For example, Zhang et al. (2017) [81] proposed a deep adversarial model that encouraged the predictions of unlabelled images to be similar to the labelled ones.

Consistency regularization

Consistency regularization is a category of SSL that introduces an unsupervised loss term to the loss function [77]. It is based on the smoothness assumption (i.e., data points close in the image space should also be close on the label space) or the manifold assumption (i.e., two data points in a local low-dimensional manifold have similar class labels) [79]. Consequently, the output of a predictive model should be robust to realistic perturbations of input data. In this setting, unlabelled data can be leveraged to produce a decision boundary which captures the data manifold [77]. Figure 3.5 illustrates the described SSL assumptions.
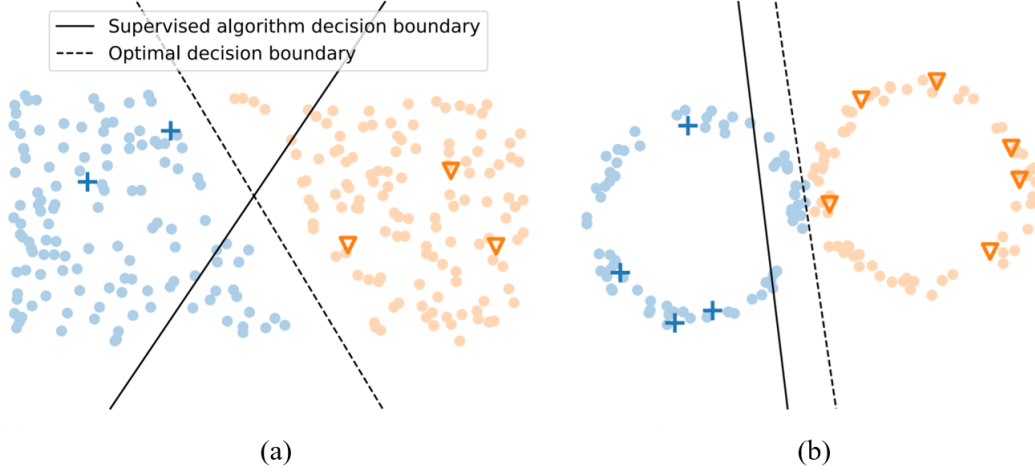
Figure 3.5: Illustration of semi-supervised learning assumptions. (a) Smoothness or low-density assumptions. (b) Manifold assumptions [51].

Consistency regularization methods aim to optimize a loss function L consisting of a supervised loss L_{sup} and an unsupervised loss or consistency regularization term L_{cons} that enforces consistency between two models' outputs. This is formulated in Equation 3.9

$$L = L_{sup} + \lambda L_{cons} \tag{3.9}$$

where L_{sup} is applied to D_l and L_{cons} is applied to $D = D_l \cup D_u$. λ is a weight coefficient that controls the trade-off between the supervised and unsupervised term.

The consistency regularization term L_{cons} can be formulated as follows. Given an unlabelled data point $x \in D_u$ and its perturbed version $\hat{x}$, the objective is to minimize the distance between the two model outputs $d(f_\theta(x), f_\theta(\hat{x}))$. One of the commonly employed distance metrics for consistency regularization is the mean square error (MSE) and the KL divergence [82]. Equation 3.10 illustrates the MSE distance metric as a consistency regularization term for model outputs $Y = f_\theta(x)$ and $\hat{Y} = f_\theta(\hat{x})$

$$L_{cons} = d_{MSE}(Y, \hat{Y}) = \frac{1}{N} \sum_{i \in N} (Y_i, \hat{Y}_i)^2 \tag{3.10}$$

where i corresponds to each pixel value in a $N \times N$ probability map.

It should be noted that consistency can also be applied between two perturbed inputs and between outputs of different networks, normally in a teacher-student structure [82].

Several methods have applied consistency regularization with different techniques [77]. For example, Li et al.(2020) [52] proposed a teacher-student model where consistency was enforced between the transformed output of a student model with a random augmentation technique (i.e., dropout, Gaussian noise, random data transformation) and the output of a teacher model from a transformed input. Based on the consistency of both outputs the weights of the teacher model were updated with an exponential moving average (EMA). Recently, Ouali et al.(2020) [83] leveraged unlabelled samples by enforcing consistency between the predictions of a main and auxiliary decoders with different perturbed inputs.

On the other hand, the above methods only consider the consistency at a data level and ignore the consistency between different tasks. Furthermore, they require several forward passes to compute the consistency loss which results in high computational costs. More recently, Zamir et al. (2020) [84] proposed to enforce consistency across different tasks based on the invariance of inference paths. This novel strategy showed promising results in that different from traditional multi-task learning approaches, consistency is explicitly enforced between predictions of different tasks during training. However, this method has only been applied in a fully supervised paradigm. Luo et al. (2020) [53] applied cross-task consistency in a semi-supervised setting achieving promising results.

Deep mutual learning

Model distillation [85] is a widely used technique to transfer knowledge from a teacher to a student model. The aim is to transfer from a powerful and large network to a smaller network that will have equal representation capacity but with significantly smaller computational or memory requirement. In a related but different concept, deep mutual learning (DML) [78] proposes that the knowledge is transferred between a pool of small student networks. Specifically, each student network is trained with a supervised loss and a mimicry loss that aligns the posterior probability of each student. In this manner, each student network can learn in a peer-teaching scenario which outperforms the conventional independent learning. Furthermore, this technique is applicable to different sized models.

Equation 3.11 and Equation 3.12 show the loss functions of two networks (Θ_1 and Θ_2) for the DML approach

$$L_{\Theta_1} = L_{C_1} + D_{KL}(\boldsymbol{p_2} \parallel \boldsymbol{p_1}) \tag{3.11}$$

$$L_{\Theta_2} = L_{C_2} + D_{KL}(\boldsymbol{p_1} \parallel \boldsymbol{p_2}) \tag{3.12}$$

where L_{C_1} and L_{C_1} are the supervised loss and $\boldsymbol{p_1}$ and $\boldsymbol{p_2}$ are the posterior probability of networks Θ_1 and Θ_2.

As observed through the supervised and KL divergence loss the networks learn to solve the task and to mimic the performance of the other network [78]. Furthermore, as each network starts from a different initialization, their learnt representation and their probability estimates are different. This variability provides additional information in DML. Overall, training collaboratively two networks through DML is effective to improve the generalization ability of both networks [78].

Dual-task mutual learning for semi-supervised segmentation

In this thesis, inspired by the work of Zhang et al.(2021) [55] we implement a VW segmentation framework based on the concepts of DML and SSL with cross-task consistency regularization. First, it follows the DML approach by simultaneously training two individual networks for collaborative learning and peer knowledge transfer. In the proposed framework each network is trained for a different task guided by a supervised and an unsupervised loss term. One network is trained for binary mask segmentation and the other for SDM regression. As each network learns reliable representations of the final segmentation target starting from different task-level conditions, the estimated probabilities are different which in the DML approach is exploited as additional knowledge. However, different from Equation 3.11 and Equation 3.12, the unsupervised loss term is not a mimicry loss. Instead, DTML is implemented in a semi-supervised setting, where an unsupervised cross-task consistency loss is integrated in the loss function to leverage information from unlabelled data by encouraging consistent predictions between networks with different tasks. The consistency loss guides the mutual learning of both networks and enforces shape constraints through signed distance map predictions.

Shape-aware regularization

The shape constraints enforced in this work during training and post-processing step will be detailed below.

Signed distance map

Level set functions have been traditionally employed in computer vision tasks to capture geometric active contours and distance information [86, 87]. As observed in Figure 3.6 , by viewing the surface of an object as a signed distance function (SDF), the implicit contour of a shape can be extracted as the intersection of the surface with a plane.

The integration of SDFs in CNN architectures has rejuvenated level set methods in their application to DL-based image segmentation [89]. In medical image applications, several works have reported a superior segmentation performance when incorporating signed distance map (SDM) regression as an auxiliary task for binary segmentation [90]. SDM regression offers several advantages. Instead of directly extracting the boundary or surface of a target shape, we can extract the distance of each pixel to the boundary. Therefore, the incorporation of SDM prediction to the training strategy can bias the network to learn the per pixel information of the boundary location and capture implicit geometrical properties [91].

(a) (b)
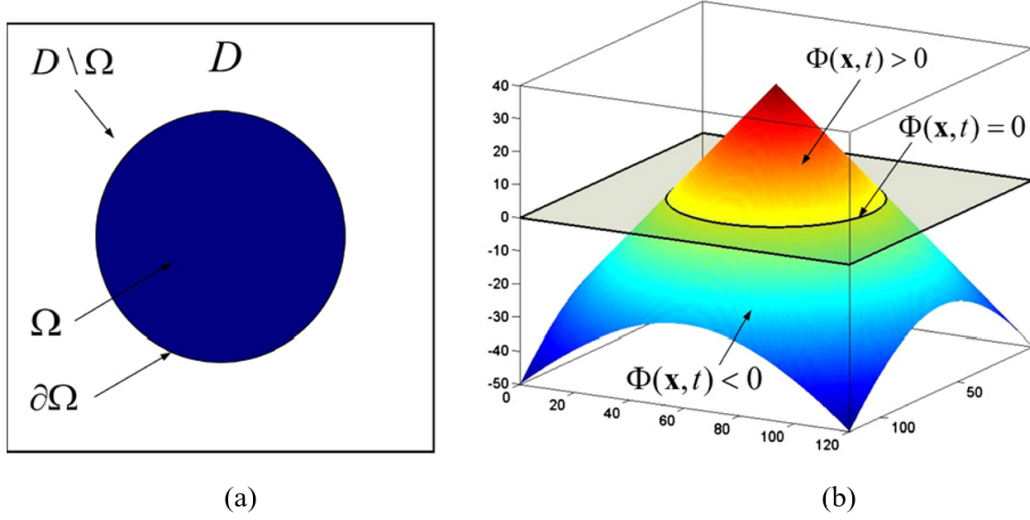
Figure 3.6: Level set description of a 2D shape. (a) design domain. (b) level set model [88].

The level set or SDM can be directly transformed from a binary segmentation mask with the function defined in Equation 3.13

$$
G_{SDF} = \begin{cases} -\inf_{y \in \partial G} \|x - y\|_2, & x \in G_{in} \\ 0, & x \in \partial G \\ -\inf_{y \in \partial G} \|x - y\|_2, & x \in G_{out} \end{cases} \tag{3.13}
$$

where $\|x - y\|_2$ represents the Euclidean distance between pixels x and y, and G_{in}, ∂G and G_{out} represent the inside, boundary and outside of a target segmentation object.

As illustrated in Figure 3.6 and Equation 3.13, SDMs take negative values inside the target object, positive values outside and 0 values on the boundary (zero-level set). The intensity value of each pixel is defined as the distance to the closest target boundary. In this work, the signed distance value is normalized between [-1,1] [55].

To transform the SDM back to the binary segmentation mask a smooth approximation of an inverse transform is employed, similar to [53]. This transformation is defined in Equation 3.14.

$$
G_{mask} = \frac{1}{1 + e^{-k \cdot z}} = \sigma(k \cdot z), z \in G_{SDF} \tag{3.14}
$$

where z is the per pixel value of the SDM and k is a transform factor, following the original implementation it is selected as $k = 1500$ [55].

As observed, Equation 3.14 is equal to a sigmoid function (section 3.2.1 - Equation 3.7) multiplied by a factor *k*, therefore it can be seen as an activation function applied to the networks output [53].

Equation 3.15 proves the differentiability of the inverse function.

$$\frac{\partial T^{-}1}{dz} = \left(\frac{1}{1 + e^{-k \cdot z}} \right)'$$
$$= k \cdot \frac{1}{1 + e^{-k \cdot z}} \cdot \left(1 - \frac{1}{1 + e^{-k \cdot z}} \right)$$

(3.15)

In this work, SDMs are generated by adding an auxiliary regression head with a hyperbolic tangent (section 3.2.1 - Equation 3.8) to a classic encoder-decoder architecture. The aim was to design a geometry-aware approach that could refine output VW segmentation masks by leveraging shape priors from the transformed distance maps.

Persistent homology

Persistent homology (PH) is a mathematical tool employed to measure the persistence of topological features across different scales. It is an integral part of topological data analysis (TDA), an emerging field where the topological attributes of the data are used to extract meaningful information from high-dimensional and noisy datasets [92]. PH has been applied to a variety of information extraction tasks, ranging from biological systems [93] to computer vision [94]. Recently, PH has been applied in medical image segmentation tasks achieving promising results in noisy images [50].

The general idea of PH is to extract the topology of simplicial complexes which can be used to represent high dimensional data points. In the present problem, PH is applied to 2D data represented by cubical complexes. A cubical complex is a set of points, lines, squares and n-dimensional counterparts used in the computation of homology in the topological space [50].

If we consider a 2D image, represented by an $N_x \times N_y$ array and defined by a set of pixel intensities X_{ij}, the output probability of a CNN parametrised with weights w at the pixel (i, j) is $S_{ij} = (f(X; w)) \in [0, 1]$. In cubical complexes, an elementary cube covering the region of pixel (i, j) can be represented as $Q_{ij} = [i, i + 1] \times [j, j + 1]$. Considering $B(p)$ as the super-level set of S_{ij} in other words, the set of pixels with S_{ij} above a threshold p, Equation 3.16 can be formulated as [50]

$$B(p) = \bigcup_{i,j} Q_{ij} : S_{ij} \geq p$$

(3.16)

By lowering the filtration value p more pixels are considered in the subset $B(p)$, this is defined in Equation 3.17.

$$\varnothing \subseteq B(1) \subseteq B(p_1) \subseteq B(p_2)... \subseteq B(0) \subseteq [0, N_x] \times [0, N_y] \qquad (3.17)$$

As more pixels join the pixel set $B(p)$, the cubical complex Q grows, and topological features may be created or destroyed. PH measures the number of topological components that appear at different $B(p)$ for each value p. This can be represented through *Betti* numbers of different dimensions. A *Betti* number of dimensions 0 (β_0) correspond to the number of connected components present on the image, a *Betti* number of dimension 1 (β_1) corresponds to the number of holes [92].

In order to analyse topological features present at different filtration values or thresholds p, persistence barcode diagrams can be generated to represent the birth (p threshold after which the topological feature first appears) and death (p threshold after it disappears) of different topological features. The birth and death values can be denoted as b_{kl} and d_{kl}, respectively, for dimension k and l_{th} largest bar. The length of the bar therefore represents the difference between b_{kl} and d_{kl}, which is known as the persistence of the topological feature. Only the most persistent features (longest bars) will be represented in this diagram as they are considered more meaningful. An example of barcode diagram and its corresponding image can be observed in Figure 3.7. In Figure 3.7(b), the x-axis represents different filtration values. Each bar corresponds to the persistence of a topological dimension, the red bar represents topological features of dimension 0 (β_0) and the green bars represent topological features of dimension 1 (β_1). As can be observed in the probability map (Figure 3.7(a)), 3 connected components (objects) are present with one of them having a hole, therefore 4 bars are represented in the persistence barcode ($\beta_0 = 3$ and $\beta_1 = 1$) [50].

The b_{kl} and d_{kl} values for each topological component correspond to a particular pixel intensity in image S. Therefore, if the intensity of that pixel is changed the b_{kl} and d_{kl} values would also change. This means that b_{kl} and d_{kl} are differentiable with respect to the image array S. Since $S = f(X; w)$ is also differentiable, the gradient of b_{kl} and d_{kl} can be calculated with respect to the weight of the network. This implies that the weights of the network can be adjusted to adhere to a specific barcode diagram or expected topology in the image [50].
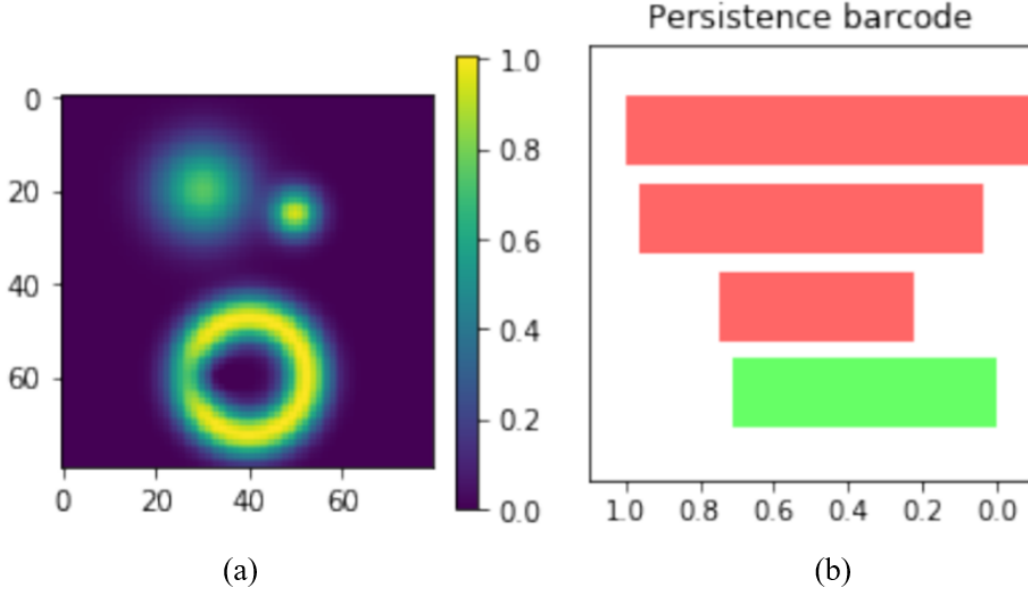
Figure 3.7: (a) Sample segmentation probability map of size 80 x 80. (b) PH barcode diagram of the probability map. The x-axis of the barcode represents different filtration values. Red bars correspond to 0-dimensional persistent features (β_0). Green bars correspond to 1-dimensional persistent features (β_1) [50]

Persistent homology-based loss function

Due to the differentiable properties of persistence barcodes, the topology of the predicted segmentation can be compared to the expected topology or topological prior and through gradient descent the prediction could be brought closer to the correct topology.

In this work, we follow the PH loss function proposed by Clough et al.(2019) [50]. Considering the *Betti* numbers of the desired topology as β_k^* the PH loss function can be defined as in Equation 3.18 and Equation 3.19.

$$L_k(\beta_k^*) = \sum_{l=1}^{\beta_k}(1 - |b_{k,l} - d_{k,l}|^2) + \sum_{l=\beta_k^*+1}^{\infty} |b_{k,l} - d_{k,l}|^2 \tag{3.18}$$

$$L_{topo} = \sum_k L_K(\beta_k^*) \tag{3.19}$$

It must be remarked that β_k^* can be considered as a topological prior which is determined by the experimenter and does not require ground-truth annotations. For example, in the present problem of VW image segmentation the expected topology is a ring shape, therefore the loss function will be minimised when β_k^* corresponds to $\beta_0 = 1$ (one connected component) and $\beta_1 = 1$ (one hole).

3.2.2 Network architecture

YOLOv4

YOLOv4 [56] is a state-of-the-art one-stage object detector. As can be observed in Figure 3.8, its architecture can be divided in 3 basic building blocks: the backbone, the head and the neck. First the input images are processed through the backbone to extract high-level feature representations. Then, the neck module down samples the extracted feature for deep feature selection. In this step, the output features from the backbone network are mixed and combined and convolutional operations perform feature aggregation. Finally, the head of the detector generates dense anchor-based predictions, similar to YOLO version 3 algorithm [95].
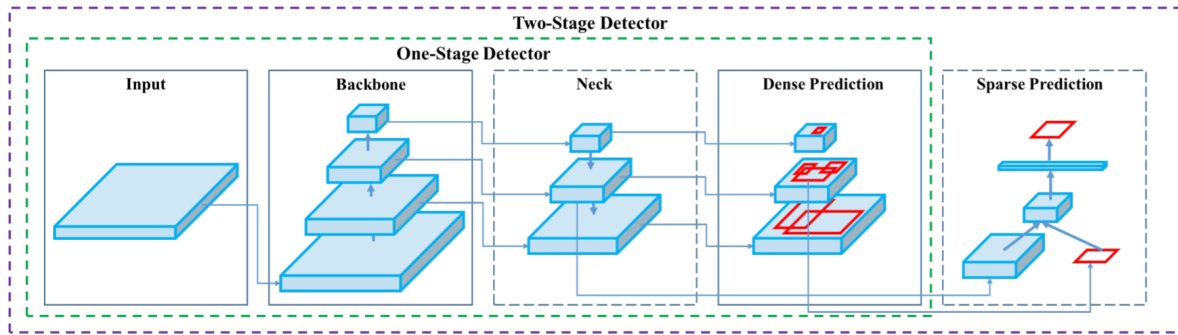
Figure 3.8: Object detector architecture [56].

The backbone consists of a CSPDarknet53 [56] network pretrained on Image-Net dataset and implemented with a Mish activation function [96]. CSPDarknet53 is a modified version of Darknet-53 [95] that includes cross-stage partial connections [97] and consists of 29 convolutional layers with 3×3 kernel size and receptive field size of 725×725. As observed in Figure 3.9, CSPDarknet53 is composed of a partial dense block and a partial transition layer that splits the gradient flow and merges it with a cross-stage hierarchy method to increase gradient richness whilst reducing computation time [97].

At the neck stage, YOLOv4 uses a modified version PANet [98] for feature aggregation that concatenates instead of adds the features from the previous and current layer. Furthermore, the object detection algorithm integrates a SPP module [99] to extract more significant context features and increase the receptive field size [56].

The role of the head detector is to output the coordinates of the predicted bounding box, the confidence score of the object and the classification label. The output predictions are based on anchor boxes which are dimension priors selected by the researcher. YOLOv4 divides the image into $n \times n$ grid cells, every grid infers the bounding box coordinates along with a confidence score [56].
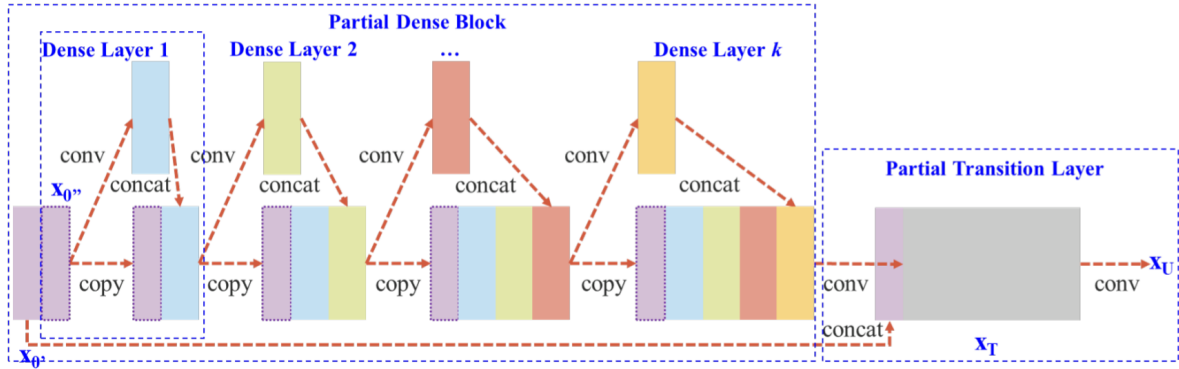
Figure 3.9: DenseNet with cross-stage partial connections [97].

YOLOv4 outputs 4 anchor-based bounding box coordinates predictions, (t_x, t_y, t_w, t_h) where (t_x, t_y) corresponds to the centre coordinates with respect to a grid cell (c_x, c_y) and (t_w, t_h) corresponds to the width and height of the bounding box. The offset coordinates can be decoded to the final bounding box coordinates (b_x, b_y, b_w, b_h) as can be observed in Figure 3.10, where (p_x, p_h) correspond to the anchor box width and height and σ to a sigmoid function [56, 95]
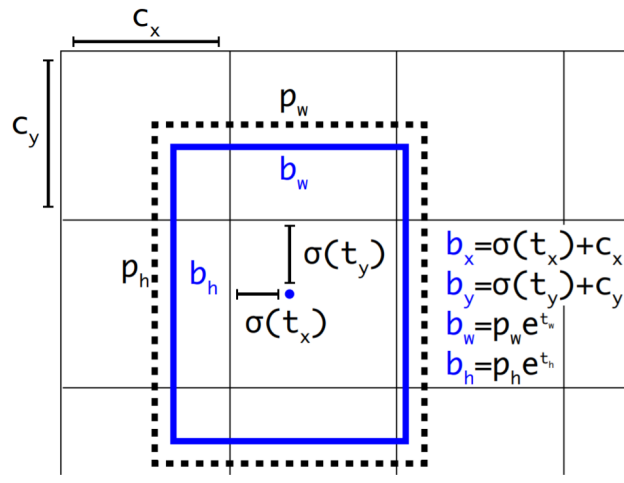
Figure 3.10: Bounding box (blue rectangle) with dimensions priors (black rectangle) [95].

As a final output, YOLOv4 generates an output tensor of size: $n \times n \times (C + 1) + 5 \times k$ where k is the number of anchor boxes and C the number of classes. The constant equal to 5 corresponds to 4 bounding box offsets coordinates and the confidence prediction of the object. $n \times n$ is the size of the grid the image is divided into. In the current implementation $C = 1$ as we only detect one object class (carotid artery) [56].

Modified U-Net

In the following section, the backbone network architecture employed for VW segmentation will be described. It consists of a modified U-Net that integrates an additional output layer for SDM regression and has 16 base filters. The modified U-Net architecture can be considered as a 2D version of the V-Net [73] without residual connections and with two output branches as proposed by [53]. The network architecture is illustrated in Figure 3.11. Similar to a traditional U-Net, it is characterized by encoder-decoder architecture with skip connections and 2D convolutions. The composition of the encoder and decoder pathway are detailed below.
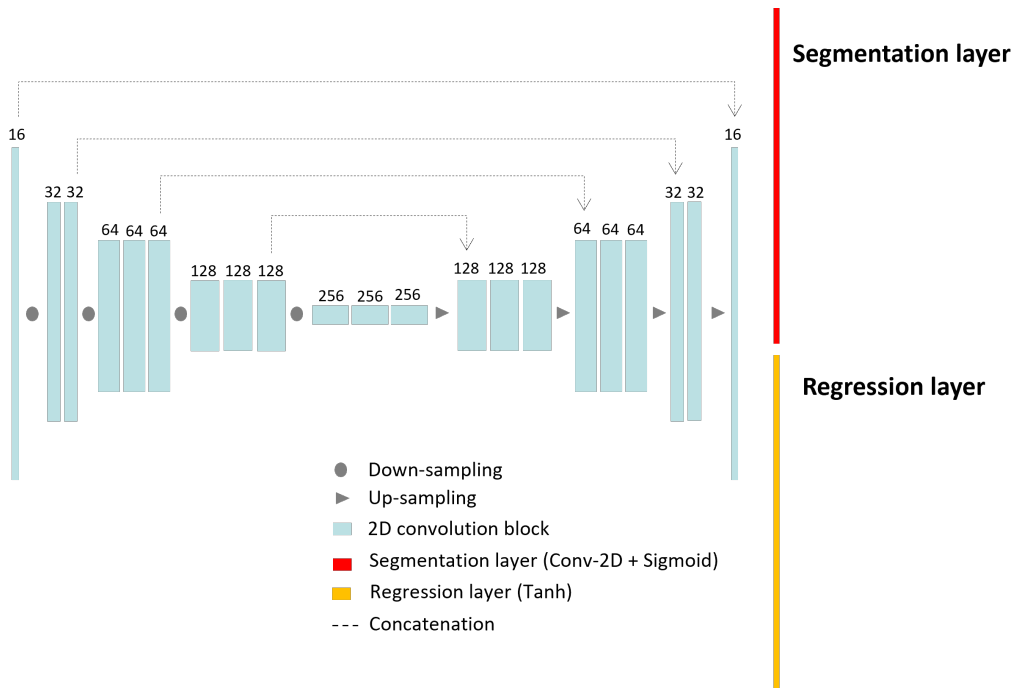
Figure 3.11: Backbone architecture of VW segmentation model. Modified U-Net inspired by [73] and [53].

.

The encoder or down-sampling path extracts contextual information from an input image. As observed in Figure 15, the network has a 5-layer depth, each layer of the encoder is composed of a convolutional block followed by a down-sampling block. The convolutional block consists of a 2D convolutional layer with kernel size 3×3, $stride = 1$ and $padding = 1$, followed by a batch normalization layer and a ReLU activation function (defined in section 3.2.1- Equation 3.6). The number of convolutional blocks varies depending on the depth of the layer. The down sampling block consists of a convolutional layer with $stride = 2$ that halves the spatial resolution of input feature maps. The number of base filters equals 16 and it multiplies by 2 as the depth of the layer increases. In this manner, the encoder extracts increasingly abstracts representations of the image.

The decoder or up-sampling path reconstructs the output image at full resolution from the feature representations extracted by the encoder. As observed in Figure 3.11, each layer consists of a convolutional and an up-sampling block. The convolutional block from the decoder follows the same structure as in the encoder. The up-sampling block consists of an up-sampling operation with bilinear interpolation that doubles the spatial resolution of feature maps. Between the up-sampling and convolutional block a concatenation layer is applied that integrates the feature maps from the previous up-sampling layer with the output feature maps from the corresponding level at the encoder. This is otherwise known as skip connections. Following the work of [53], the output of the decoder consists of 2 branches. The segmentation layer consists of a convolutional operation followed by a sigmoid activation function (section 3.2.1 - Equation 3.7) that outputs the pixel-wise probability maps in the range [0,1] for VW binary segmentation. The regression layer consists of a tanh (section 3.2.1 - Equation 3.8) activation function that outputs the SDM prediction maps in the range [-1,1]. Dropout is applied in the last layer of the encoder and before the final layer in the decoder with a probability of $p = 0.5$.

3.2.3 Proposed framework

In this section we introduce the proposed framework for automated VW segmentation. The overall framework can be observed in Figure 3.12.
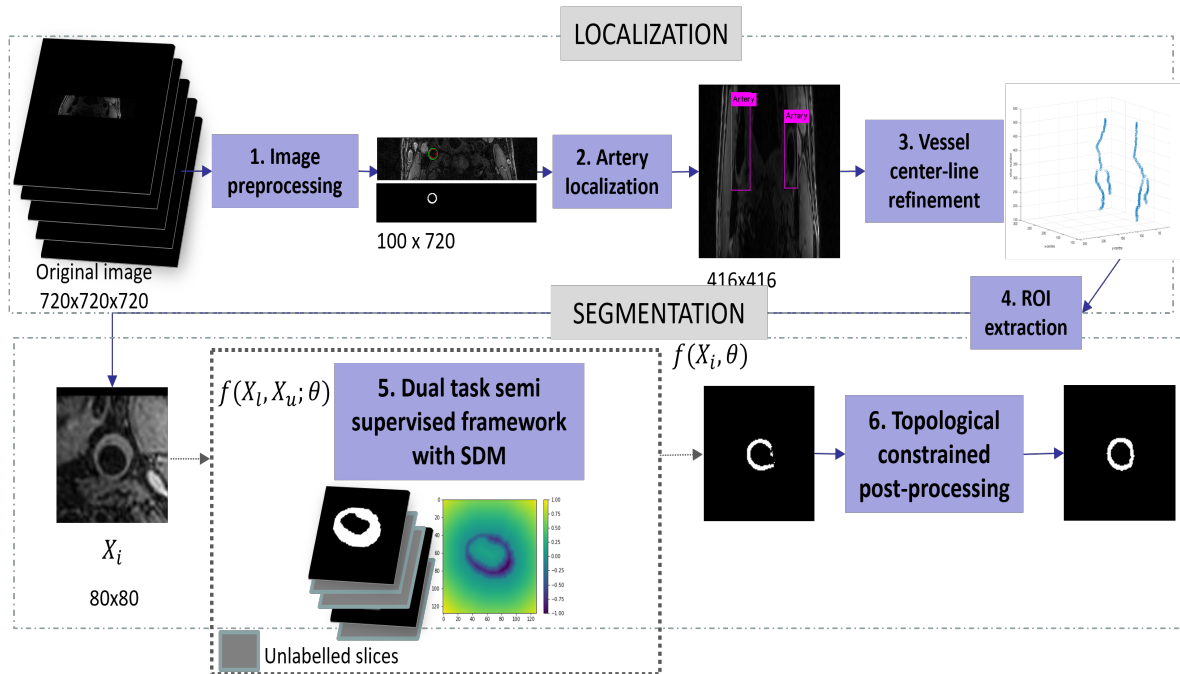
Figure 3.12: Proposed framework for VW localization and segmentation.

The proposed framework consists of localization and a segmentation module. The purpose of the localization module is the automatic detection of the artery lumen centre. For this, after an initial preprocessing stage a state-of-the-art object detection algorithm was custom trained for carotid artery detection. The outputs were bounding boxes encompassing the artery from which the artery centre could be automatically extracted. To refine the artery centre predictions, a vessel centreline reconstruction step was applied. In this step, the predicted artery localization was tracked through consecutive slices for the 3D reconstruction of the vessel centreline. This provided additional spatial information that could aid in the identification and correction of miss-detections. From the refined artery centre predictions, the ROI was extracted. The localization module was applied to both labelled and unlabelled slices and the ROI was extracted as image patches of size 80×80 around the predicted artery centre. The ROI patches were then resized to size 128×128 and input to the segmentation module.

For VW segmentation, we implemented a DTML framework [55] that consists of two networks, the first network was tasked with the prediction of a binary mask segmentation and the second with the prediction a level set representation or SDM. Through SDM regression geometrical shape constraints were enforced by the network. Both tasks were joined by a consistency loss applied in a SSL setting which guided the mutual learning of both networks. The aim was to explore the effect of shape constraint for the task of VW segmentation and to leverage information from the large number of unlabelled slices provided.

The initial predictions obtained from this framework were then evaluated for their topological correctness. Segmentations that did not follow the expected topology of VW were input to a post-processing module that applied a PH-based topological loss function [50], encouraging the network to output predictions with the desired topology. Therefore, in this framework shape constraints were applied both during training of the network through signed distance map and during the post-processing stage through topological priors.

Image preprocessing

The images provided by the challenge were processed for subsequent experiments. First, the slices with labelled annotations were identified and extracted from the dataset. Then, the original zero-padded images of size 720×720 were thresholded with Otsu threshold to identify the padded regions which were cropped out of the image. The final image size of the axial slices was 100×400. Figure 3.13 illustrates this process.

The coordinates of lumen and outer-wall contours provided were extracted with the CASCADE [58] software and the coordinates were transformed to the cropped image. Then, a fill operation was applied between the contours to obtain the VW binary segmentation mask. This is illustrated in Figure 3.14.
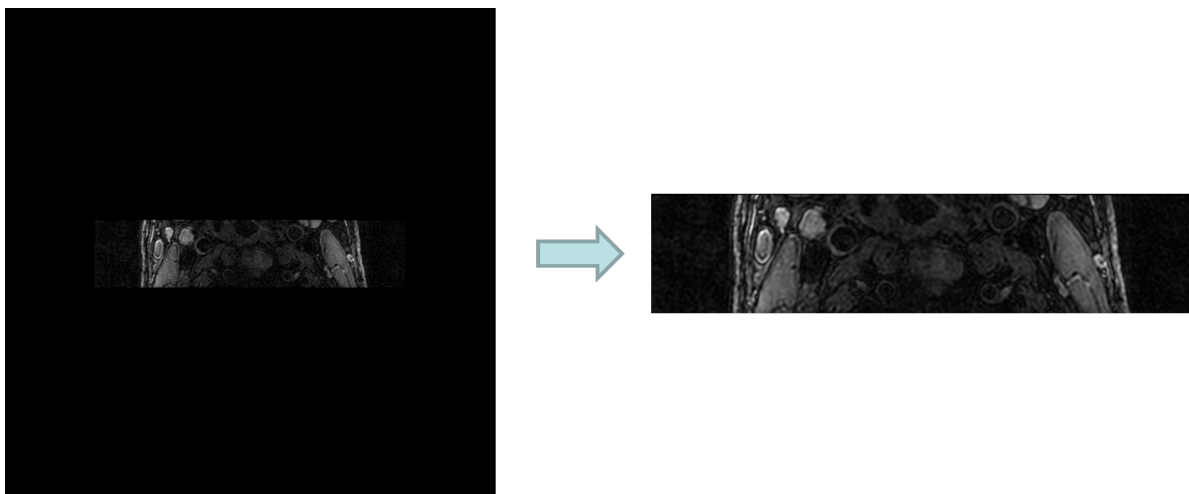
Figure 3.13: Crop operation for zero-padding removal. Left: original image array of size 720x720. Right: cropped image array of size 100x400.
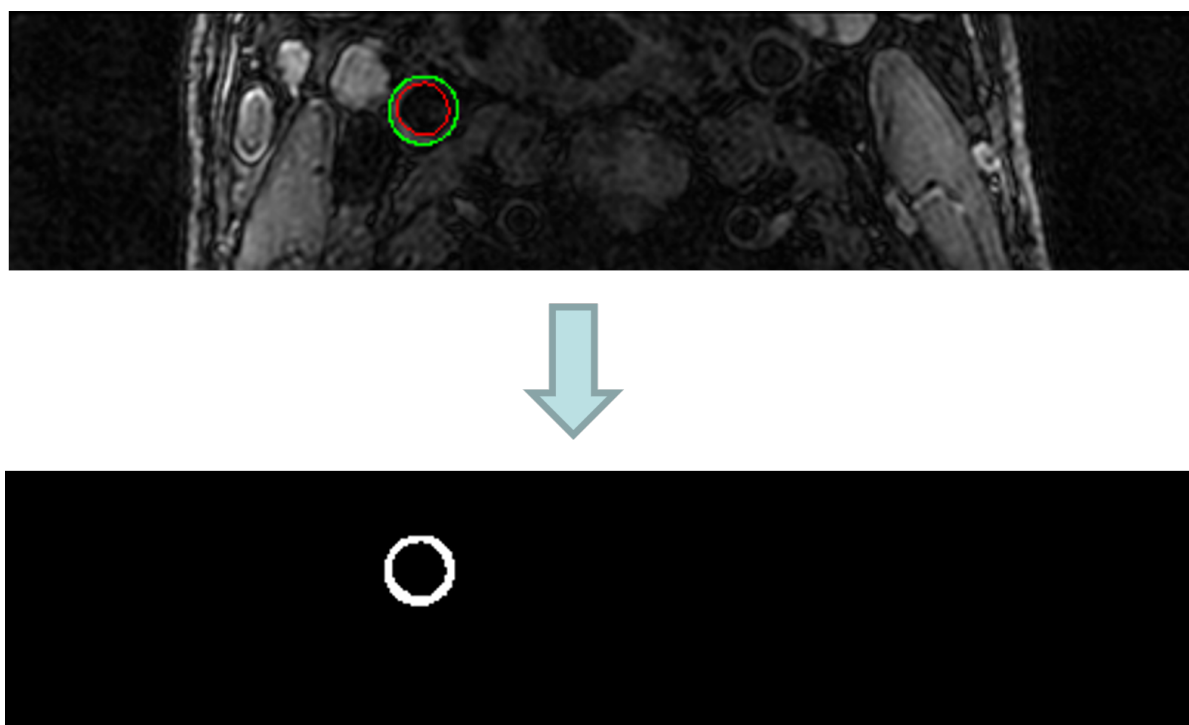
Figure 3.14: VW binary mask extraction. Above: lumen (red) and outer-wall (green) plotted contours. Below: ground-truth VW binary mask.

The intensity of the images was adjusted to a constant range and the contrast was maximized with a linear normalization operation as shown in Equation 3.20, forcing the lowest pixel intensity value to be 0 and the highest to be 255

$$u_i = \left(\frac{v_i - v_{in}}{v_{max} - v_{min}} \right) \tag{3.20}$$

where v_i and u_i correspond to the intensity of the image at pixel i for the cropped and normalized image, respectively. v_{min} and v_{max} correspond to the minimum and maximum intensity of the cropped image.

Artery centre localization

The aim of our proposed artery localization strategy is to automatically and accurately detect artery centres for ROI extraction prior to the segmentation task. Unlike other organs such as the brain or the liver, the VW is a very small and thin structure, normally comprising less than 0.1% of the image [45] (on axial slices), therefore, ROI extraction is normally required to 'zoom in' the target region and reduce class imbalance between foreground pixels (those corresponding to the VW) and background.

For artery centre localization we employ a 2-step approach: target region identification and vessel centreline refinement. First, the target object region is identified with a state-of-the-art object detection algorithm, known as YOLOv4 [56] . From the predicted artery localization the outer-wall centre is extracted. Then, a vessel centreline refinement method is adopted to track the artery centres across consecutive slices spatially and correct inaccurate artery centre predictions.

Target region identification

For target region identification we employed a CNN-based object detection algorithm known as YOLOv4 [56]. The aim was to predict minimum encompassing rectangles (bounding boxes) of the carotid arteries in all slices. YOLOv4 was selected due to its state-of-the-art results in detection accuracy and inference speed which render the process time efficient.

1. Ground-truth generation

The object detection algorithm requires an input training pair of (D_i, B_i) where D is the input image and B denotes the ground-truth bounding box coordinates for each sample i.

The first step was to extract the ground-truth bounding box from the binary segmentation masks. For this, a contour approximation algorithm was applied to detect the outer-wall contour from the binary segmentation masks. Figure 3.15 illustrates the bounding box extraction from the ground-truth binary mask.
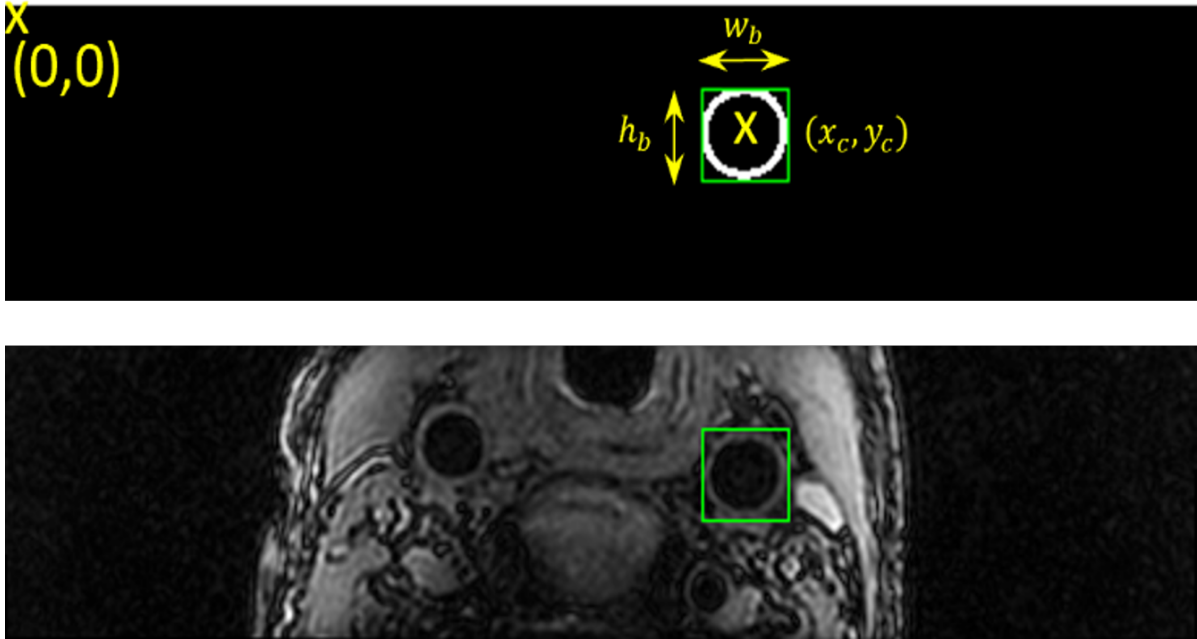
Figure 3.15: Above: ground-truth bounding box extraction from VW binary mask. Below: corresponding image with ground-truth bounding box (green rectangle).

Once the ground-truth bounding box coordinates were extracted they were reformatted to fit the input requirements of YOLOv4. Equation 3.21 shows the attributes of the YOLOv4 label file

$$\left(c, \frac{x_c}{w_i}, \frac{y_c}{h_i}, \frac{w_b}{w_i}, \frac{h_b}{h_i} \right) \tag{3.21}$$

where c corresponds to the class identification number (in this work $c = 0$ as we only consider one object class), x_c and y_c define the bounding box centre coordinates with respect to the top-left corner of the image and (w_b, h_b) correspond to the size of bounding box b. The coordinates were normalised with respect to the size (w_i, h_i) of image i.

2. Image preprocessing

The input resolution of the images is an important parameter as it determines the number of pixels that the model will perform detection on. For example, increasing the input resolution may lead to better accuracy results however at the expense of longer training and inference times. On the other hand, larger pixel resolutions can help the object detector to detect smaller objects. In addition, YOLOv4 algorithms expect squared input sizes that are multiples of 32. Due to the small size of the artery and YOLOv4 prerequisites, we select as input resolution 416×416 which is also a standard value used in YOLOv4 training.

The images were resized applying the factor in Equation 3.22 with bicubic interpolation.

$$\left(\frac{416}{w}, \frac{416}{h}\right) \tag{3.22}$$

The same transformation was then applied to the generated ground-truth bounding box as illustrated in Figure 3.16.
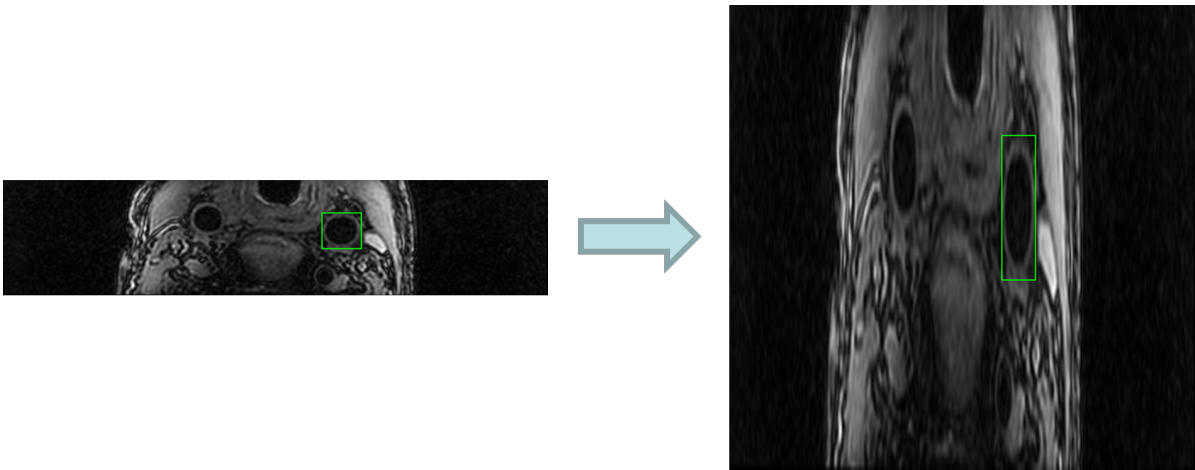
Figure 3.16: Right: original cropped image of size 100x400. Left: resized YOLOv4 input image to size 416x416. Ground-truth bounding box in green.

3. Data augmentation

Due to the limited number of annotated images, especially, for certain artery types such as ICA/ECA, several data augmentation strategies were employed to increase the robustness of the algorithm and increase the size of the training set.

First, for each labelled slice the ground-truth bounding box was propagated 6 consecutive slices in each side with the exception of neighbouring slices already containing annotations.

Figure 3.17 illustrates an example of slice propagation, as can be observed, in neighbouring slices the size and location of the artery does not suffer noticeable changes and remains bounded within the propagated bounding box.

On the other hand, the bounding box propagation requires at least one labelled slice of the target carotid branch for each patient, and some carotid branches such as ECA-L(left) and ECA-R(right) are only annotated on 6 patients. To overcome this and increase the robustness of the algorithm, we treat YOLOv4 training from a semi-supervised approach by generating bounding-box pseudo-labels. Initially, the object detection algorithm is pretrained with the labelled and augmented slices, then unlabelled slices are passed as input during test-time and pseudo labels are generated for ECA and ICA structures. This way, the bounding box training set for these carotid branches can be increased.
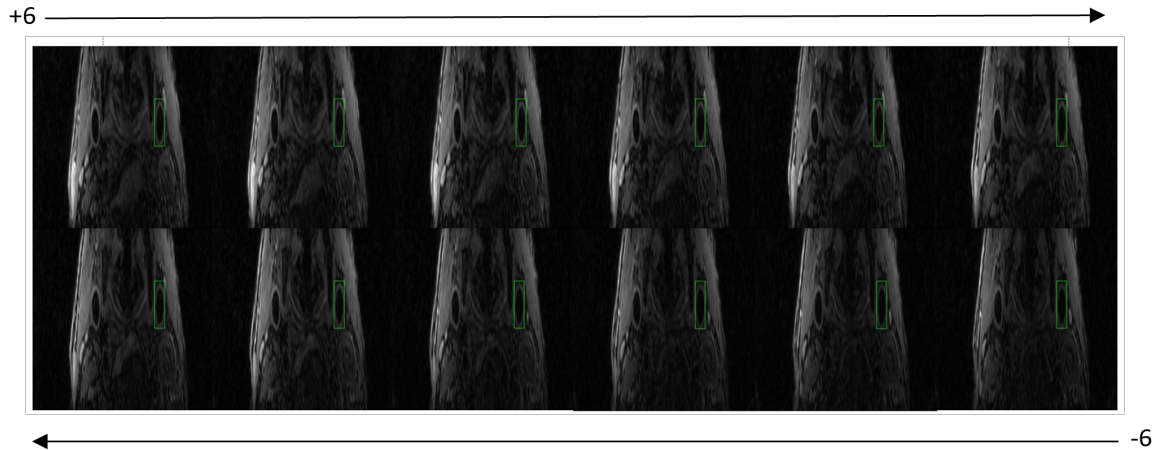
Figure 3.17: Label propagation strategy to 6 consecutive slices in both directions (± 6).

Table 3.2 summarises the number of annotated slices for each structure after each 'data augmentation' strategy is applied. Finally, YOLOv4 is trained with 21260 bounding box annotations.

Table 3.2: Data augmentation for object detection.

N° of annotations	Initial	With propagation	Without propagation
ICA-L	249	5001	5274
ICA-R	224	4652	4925
ECA-L	24	408	621
ECA-R	39	399	620
CCA-L	1096	4186	4186
CCA-R	1038	5634	5634

ICA: internal carotid artery, ECA: external carotid artery, CCA: common carotid artery. L: left, R: right.

Once the ground-truth was generated, the next step was to integrate labelled annotations of different carotid types into the same or corresponding image slice. As observed in Figure 3.18, due to the sparsely labelled slices provided by the challenge, the number of ground-truth bounding boxes employed to train the object detection algorithm ranged from 1 to 4.

Finally, data augmentation transformations were applied off-line and on-line, as introduced in the YOLOv4 algorithm. Online data augmentation consisted of mosaic data augmentation and photometric distortion with saturation = 1.5 and exposure = 1.5. Offline data augmentation consisted of random transformations of rotation $[-15°, 15°]$ and shear $[-15°, 15°]$. Figure 3.19 illustrates an example of off-line data augmentation results for 3 slices.
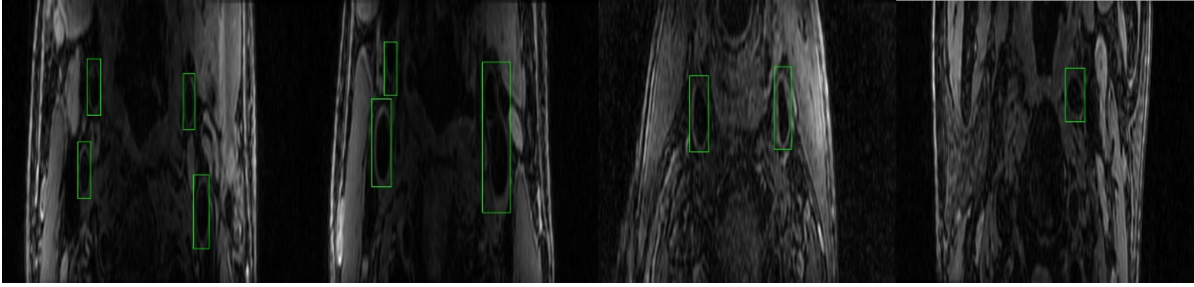
Figure 3.18: Input slices with different number of ground-truth bounding boxes.
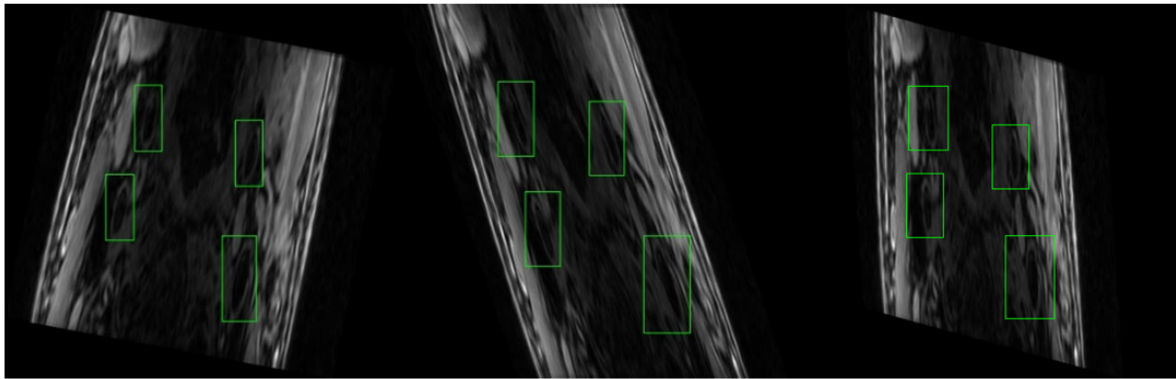
(a)   (b)   (c)

Figure 3.19: Offline data augmentation transformations on 3 sample slices. (a) Rotation. (b) Shear. (c) Rotation & Shear.

Vessel centreline refinement

The predictions of the trained object detection algorithm were refined by a vessel centreline tracking method. Inspired by [45], vessel centreline reconstruction was performed by mapping the axial localization (x and y coordinates) of the detected artery centres against the slice number for each patient. Different from [45] the vessel centreline was reconstructed in 3D which provided additional insight of the miss-detected artery localizations. Figure 3.20 illustrates the proposed vessel centreline reconstruction method for one patient. Following the assumption that artery centres have continuity between slices, miss-detections could be identified as artery centre localizations that spatially deviate from the reconstructed artery centre line. Once the deviated artery centres were identified they were corrected by propagating the localization of the nearest prediction within the vessel centreline. In cases of no-detection the same refinement process was performed. In this manner, the initial predictions of the localization algorithm were refined with respect to the vessel centreline.
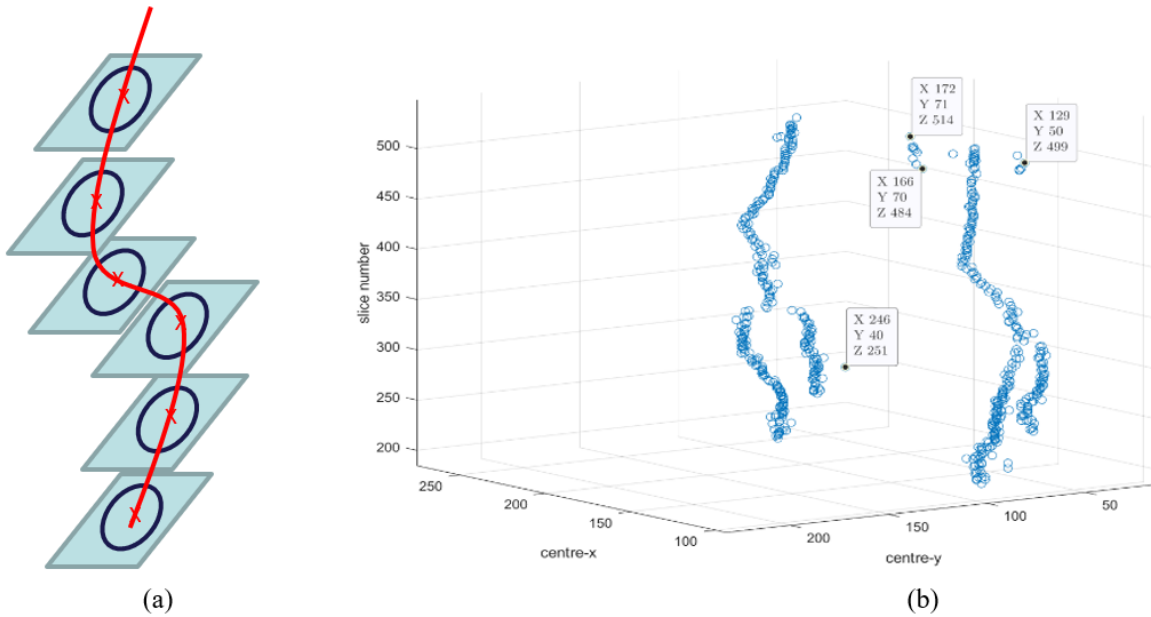
Figure 3.20: (a) Illustration of vessel tracking. (b) 3D vessel centreline reconstruction graph for left and right carotid artery. Labelled points correspond to identified miss-detections.

Vessel wall segmentation

In this step, the aim was to obtain accurate binary segmentation prediction masks of the carotid vessel wall (VW) for ECA, ICA and CCA bilateral structures. For this, 2 dual-task networks were employed, inspired by [55].

Image preprocessing

The image preprocessing stage for VW segmentation framework consisted primarily of patch extraction around the region of interest (ROI).

ROI extraction can be understood as 'zooming' into the target region after its localization to increase the VW size with respect to the background and reduce foreground and background pixel class imbalance. This facilitates the thin VW segmentation in posterior binary segmentation methods.

Once the artery vessel centre prediction was refined, to ensure inter-slice continuity the ROI of VWl segmentation maps were extracted. For this, all VW binary segmentation maps were cropped around the predicted artery centre with a square crop of 80×80 pixels. The ROI extraction process is illustrated in Figure 3.21. As can be observed, zero-padding was applied in regions were the cropped patch exceeded the image dimension. The cropped images were then resized to a final input size of 128×128.
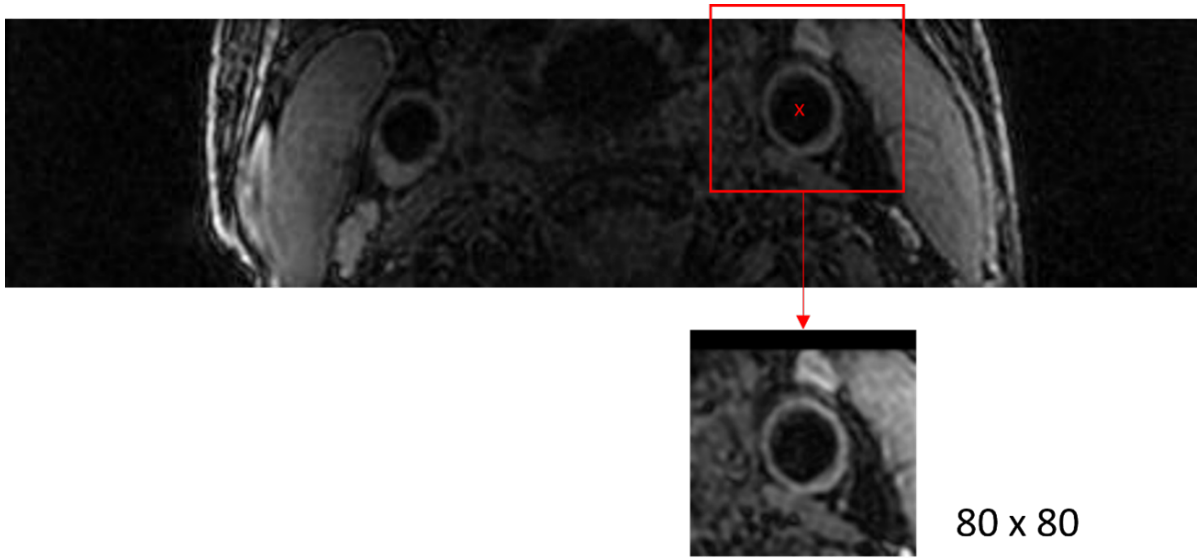
Figure 3.21: ROI extraction.

Dual-task mutual learning framework

In the following section we introduce the proposed 2D VW segmentation framework inspired by the dual-task mutual learning (DTML) approach [55]. Figure 3.21 illustrates the DTML VW segmentation framework which consists of the integration of two individual networks. Both networks share the same backbone architecture, detailed in section 3.2.2. The upper network (M_s) aims at predicting a binary mask segmentation whilst the lower one (M_d) aims at regressing the SDM. The dual-task training allowed to learn the segmentation target from different perspectives and with the integration of SDM regression, geometrical constraints could be enforced to extract more reliable information. Furthermore, as the segmentation target can be represented by both a pixel-wise probability map and a high-level contour or SDM both predictions should be consistent. A cross-task consistency loss [84] was therefore applied between the predicted binary mask from the segmentation network and the SDM prediction transformed to a binary representation. Different from traditional consistency training schemes, consistency was enforced between different tasks leading to a task-level regularization that stabilized the learning procedure. Furthermore, the cross-task consistency loss guided the transfer of knowledge between both networks, in a concept similar to DML [78] where, different from the traditional one-way transfer from teacher to student, an ensemble of student networks can collaboratively and implicitly learn useful knowledge during the training process. The cross-task consistency loss required no supervision which allowed to utilize unlabelled cases for the extraction of additional rich information to improve the model's generalization ability.
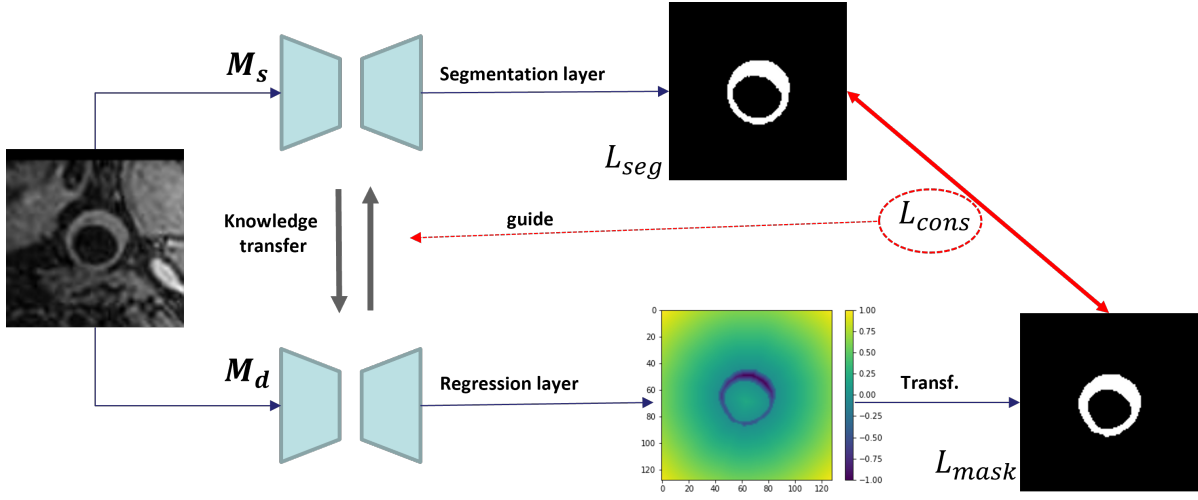
Figure 3.22: Dual-task mutual learning (DTML) framework [55] for VW segmentation.

Semi-supervised dual-task mutual learning

The implemented dual-task networks employ the same backbone architecture (section 3.2.2), however, for M_s only the segmentation head is activated whilst for M_d only the regression head. Given an input image $X \in \mathbb{R}^{HxW}$ the M_s network outputs a pixel-wise segmentation map ($\hat{Y}_{seg}$) and the M_d network outputs a SDM ($\hat{Y}_{dis}$) [55].

The output maps are defined in Equation 3.23

$$\hat{Y}_{seg} = f_{seg}(X; \theta_{seg}), \hat{Y}_{dis} = f_{seg}(X; \theta_{dis}) \tag{3.23}$$

where $\hat{Y}_{seg} \in [0,1]^{HxW}$ and $\hat{Y}_{dis} \in \mathbb{R}^{HxW}$, θ_{seg} and θ_{dis} are the parameters of the segmentation networks M_s and M_d, respectively.

The dual task networks were implemented in a semi-supervised setting through consistency regularization (section 3.2.1 - Equation 3.9). For semi-supervised segmentation of 2D medical images, we can consider a training dataset D consisting of M labelled cases and N unlabelled cases where $M \ll N$. The labelled data set can be denoted as $D_l = \{X_i, Y_i\}_i^M$ and the unlabelled dataset as $D_u = \{X_i\}_i^N$ where $X_i \in \mathbb{R}^{HxW}$ and $Y \in \{0,1\}^{HxW}$ represent the input and ground-truth, respectively.

Both networks were trained simultaneously. Each network backpropagates a loss function composed of a supervised term (L_{seg} for M_s and L_{mask} for M_s) applied to D_l and an unsupervised term (L_{seg}) applied to $D_l \cup D_u$.

Equation 3.24 and Equation 3.25 defines the loss function for M_s and M_d, respectively.

$$L_{seg} + \lambda_{cons} * L_{cons} \tag{3.24}$$

$$L_{mask} + \lambda_{cons} * L_{cons} \tag{3.25}$$

Both M_s and M_d employ the Dice loss function (section 3.2.1 - Equation 3.5) as the supervised term. In M_s the supervised loss L_{seg} is applied between the predicted and ground-truth binary segmentation mask. In M_d the supervised loss L_{mask} is applied between the output SDM transformed to a binary mask representation and ground-truth binary mask. The segmentation task learnt pixel-wise information of the target segmentation whilst the regression task learnt geometrical information of the target segmentation shape.

As observed in Equation 3.24 and Equation 3.25, both networks are joined with a cross-task consistency loss L_{cons}. Equation 3.26 defines the implementation of L_{cons}. The consistency was enforced through a MSE loss function applied between the M_s segmentation prediction and the transformed M_d regression prediction

$$L_{cons} = \lambda_{cons} \| f_{seg}(X; \theta_{seg}) - f_{mask}^{-1}(f_{dis}(X; \theta_{dis})) \|_2^2 \tag{3.26}$$

where λ_{cons} represents the trade-off weight between the supervised and consistency loss and f_{mask}^{-1} is the inverse transformation applied to SDM predictions. $\| \|_2$ denotes the L2 norm. $f_{seg}(X; \theta_{seg})$ and $f_{dis}(X; \theta_{dis})$ correspond to the models M_s and M_d, respectively.

The weight λ_{cons} represents a Gaussian ramp-up function, defined in Equation 3.24 [55]

$$\lambda_{cons}(t) = 0.0001 * e^{-5*(1 - \frac{T}{T_{max}})} \tag{3.27}$$

where t is the number of iterations.

In summary, the goal of the segmentation framework is to collaboratively minimize two loss functions, defined in Equation 3.28 and Equation 3.29 , that correspond to the pixel-wise segmentation and SDM regression tasks, respectively.

$$\min_{\theta_{seg}} \sum_{i \in D_l} L_{seg}(f_{seg}(X_i; \theta_{seg}), Y_i) + \sum_{i \in D} L_{cons}(f_{mask}^{-1}(f_{dis}(X_i; \theta_{dis})), f_{seg}(X_i; \theta_{seg})) \tag{3.28}$$

$$\min_{\theta_{dis}} \sum_{i \in D_l} L_{mask}(f_{mask}^{-1}(X_i; \theta_{dis}), Y_i) + \sum_{i \in D} L_{cons}(f_{mask}^{-1}(f_{dis}(X_i; \theta_{dis})), f_{seg}(X_i; \theta_{seg})) \tag{3.29}$$

Persistent homology post-processing module

Inspired by the works of Clough et al.(2019) [50] we implemented the following post-processing framework based on the concept of Persistent homology (PH) (section 3.2.1) to further correct the topology of the predicted segmentations.

In the post-processing framework, the trained CNN network f with weights w was used to infer prediction segmentations at test-time. Then, for each sample X_n in the testing set the persistence barcode, was computed. Figure 3.23 shows an illustration of the persistence barcode and the corresponding predicted segmentation for 3 slices. The Figure 3.23 above shows a prediction with a correct or expected topology, this corresponds to $\beta_1^* = 1$ (one hole) and $\beta_0^* = 1$ (one connected component). The persistence barcode associated was used as the 'true' barcode in the loss function. Below, examples of incorrect predictions are shown where the VW segmentation corresponds to more than one connected component (Figure 3.23 below-left) or a broken down or C-shaped prediction (Figure 3.23 below-right). As observed, the persistence barcode and *Betti* numbers deviate from the expected values. In this manner, predictions with incorrect topology were detected and were input to the PH post-processing module with the aim of correcting their topology.
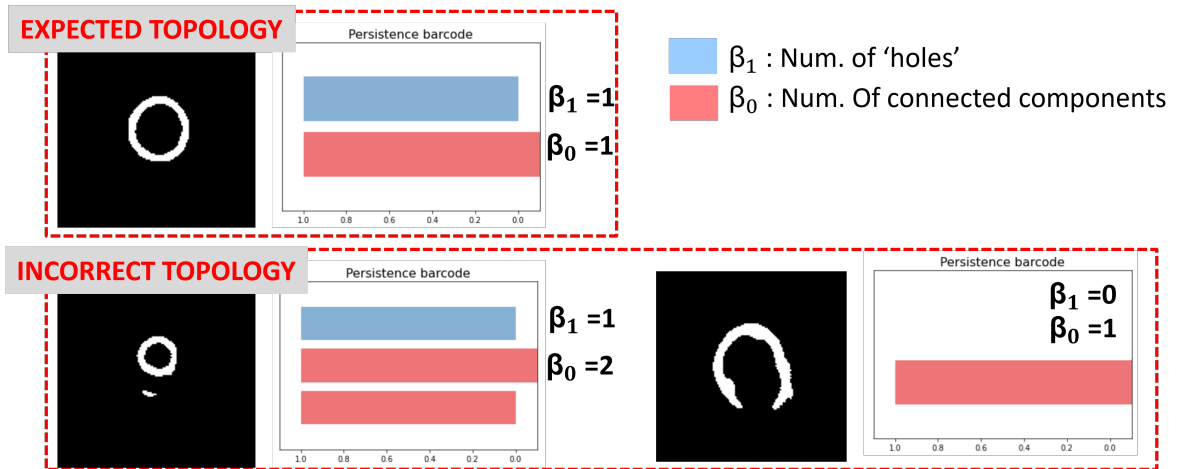
Figure 3.23: Illustration of persistence barcode extraction for different topologies in 3 sample predictions. Above: Binary mask and barcode for prediction with correct VW topology. Below: binary mask and barcode for 2 incorrect VW topology predictions.

For each incorrect testing prediction X_n, the topological loss function defined in section 3.2.1 - Equation 3.19 as L_{topo}, which penalized the differences between the predicted persistence barcode and the 'true' persistence barcode, was minimised. In the PH-based post-processing module, L_{topo} guided the test-time adaptation of the weights from the pretrained network $f(X_n, w)$ to correct the inferred segmentation topology. This resulted in an updated set of weights (w_n) learnt for each individual testing slice. However, since many segmentations could potentially minimise L_{topo}, a similarity constraint in the form of a MSE function between the original and the updated networks prediction is applied to limit the modifications to a minimal set necessary to correct the predicted topology. Equation 3.30 illustrates the PH-based loss function employed

$$
\begin{aligned}
L(X_n; w, w_n) = \frac{1}{V} |f(X_n, w) - f(X_n, w_n)|^2 \\
+ \lambda L_{topo}(X_n, w_n)
\end{aligned}
\tag{3.30}
$$

where V is the number of pixels and λ the weight coefficient between the mean square loss constraint and the topological loss. Equation 3.30 was minimised for the expected topology illustrated in Figure 3.23 above with a weight coefficient of $\lambda = 1000$.

3.2.4 Implementation

The CARE II training dataset provided by the CAVWSC 2021 [10] challenge was employed to train the proposed framework. The training set consisted of 25 cases with 2670 carotid VW annotations. Data partitioning was performed in both localization and segmentation modules where 20% of the training set was randomly selected and used as the validation set. In total, 20 cases (2078 annotations) were used as the training set and 5 cases (592 annotations) as the validation set. For the training of the segmentation module, unlabelled slices were also utilized. From the unlabelled and sparsely labelled slices in the training set a total of 8665 unannotated carotid patches were extracted. Unlabelled slices were considered as slices with no carotid annotation between the first and last labelled slice for each case (outer-most unlabelled slices were not considered due to their lack of image quality). A separate set of 25 cases (in total 4198 annotated patches), provided by the CAVWSC 2021 challenge was used to test the localization and segmentation module. The proposed framework was trained in Pytorch [100] using an NVIDIA ®Tesla T4 [101] GPU.

For the custom-training of the object localization algorithm, extensive data augmentation techniques were employed (section 3.2.3 - Table 3.2) that augmented the labelled training dataset to a total of 21260 bounding box annotations. The YOLOv4 object detection network was trained for 6000 iterations with a batch size of 64 and batch subdivision size of 24. The algorithm was optimized with Stochastic Gradient Descent (SGD) with *momentum* = 0.949 and *decay* = 0.005. The input resolution was 416x416 and multi-scale training was applied to improve the generalization ability of the network. The network was initialized with pretrained weights from the original implementation. Following the original implementation [56], the Complete IoU loss function (cIoU) (section 3.2.1 - Equation 3.1) with learning rate 0.001 was employed and 9 anchor boxes were used as dimension priors. 0em

In the VW segmentation module, the DTML framework was trained with the ROI carotid patches extracted for labelled (2670) and unlabelled (8665) slices. The framework was trained with a SGD optimizer (*momentum* = 0.9 and weight *decay* = 0.0001) for 18000 iterations. The base learning rate was 0.01 and decayed by 0.1 every 7500 iterations. The batch size is 4 and consisted of 2 labelled and 2 unlabelled images. The network was trained with carotid patches (after ROI extraction) resized to 128×128. To avoid over-fitting, data augmentation techniques were applied on-the-fly during training and consisted of rotation ($\pm 90°$) and flip (horizontal and vertical) transformations. At inference time, the pixel-wise prediction generated by the M_s network was employed as the segmentation result.

At the post-processing stage, the topological loss function based on PH [50] was applied only to prediction segmentations with detected incorrect topology (in total 215 slices). For the test-time adaptation an Adam optimizer was used with a learning rate of 10^{-5} for 100 iterations. Each slice required at worst 5 minutes for the topological correction. All experiments performed in the post-processing module were implemented using Pytorch [100]. For the computation of persistence barcodes the Topolayer package introduced in [102] was employed.

4 Results and Discussion

The following section details the results of the proposed framework. Section 4.1 shows the results of the initial artery localization module. Section 4.2 illustrates the results of the segmentation module.

Both the artery localization and segmentation modules were evaluated with a separate testing dataset provided by the CAVWSC 2021 challenge [10]. The testing dataset was acquired from the CARE II study [31] and consisted of 25 cases. Similar to the training dataset, not all slices were labelled, and several slices were sparsely labelled. In total, 4189 carotid annotations in 3307 separate slices for bilateral CCA, ECA and ICA carotid types were employed for evaluation.

Evaluation metrics

In the following section the evaluation metrics employed in this work will be defined.

The evaluation of the object detection system was performed with two metrics, the IoU between the predicted and ground-truth bounding boxes (section 3.2.1 - Equation 3.2) and the pixel-wise mean absolute distance (MAD) between the ground-truth vessel centre (or nearest in cases with no ground-truth annotations) and the predicted artery centre.

The distance metric employed for the MAD measurement was the Euclidean distance. The pairwise Euclidean distance between 2 points (i.e. predicted and ground-truth artery centre localization) in a 2-dimensional space is defined in Equation 4.1

$$d(P_1, P_2) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \qquad (4.1)$$

where P_1 and P_2 are 2 points with coordinates (x_1, y_1) and (x_2, y_2), respectively.

The performance of the VW segmentation module was evaluated with ground-truth binary maps representing the VW area. The metrics employed were the Dice coefficient, the Jaccard Index, the Haussdorf Distance (HD) and the topological correctness. In the following section, *A* corresponds to the ground-truth binary mask and *B* to the predicted binary mask.

The Dice similarity coefficient [103] represents the spatial overlap between the binary segmentation prediction and the ground-truth. The definition of Dice coefficient can be seen in Equation 4.2.

$$Dice(A, B) = \frac{2(A \cap B)}{A + B} \tag{4.2}$$

A Dice = 1 indicates a complete area overlap between predicted and ground-truth target segmentation.

The Jaccard index [104] is a similarity coefficient applied between ground truth and predicted segmentation. It is defined in Equation 4.3.

$$Jaccard(A, B) = \frac{|A \cap B|}{|A \cup B|} \tag{4.3}$$

The HD [105] is defined as the maximum surface distance between two objects in both directions. It is used to measure the shape similarity between a ground-truth and prediction binary map. The HD between prediction and ground-truth binary map is defined in Equation 4.4

$$HD(A, B) = max\{\sup_{x \in A} \inf_{y \in B} \|x + y\|, \sup_{y \in B} \inf_{x \in A} \|x + y\|\} \tag{4.4}$$

where x and y correspond to pixels of binary image arrays A and B, respectively.

The topological correctness is defined in Equation 4.5. A correct topology for VW segmentation is defined by a discrete binary segmentation consisting of one connected component with one hole inside [50].

$$\% \text{ Correct topologies} = \frac{\text{N}^{\text{o}} \text{ of predictions with correct topology}}{\text{N}^{\text{o}} \text{ of predictions}} * 100 \tag{4.5}$$

4.1 Artery localization

In the following section, the results of the target region detection and vessel centreline refinement steps will be presented.

4.1.1 Target region identification

For automatic artery centre localization predictions, the custom-trained YOLOv4 object detection algorithm was employed. The object detection algorithm predicted in total 7121 bounding boxes, however, due to the sparse annotations of the testing set provided by the challenge, the quantitative evaluation was performed only on output predictions with a corresponding ground-truth annotation.

To represent the detection accuracy in distinct regions, such as CCA, BF and ECA/ICA, the bifurcation location provided by the CASCADE [7] software was used. The predicted bounding boxes were classified with respect to their locations for different carotid branches.

Table 4.1 shows the artery centre detection results for each carotid type. A threshold of 50% IoU between ground-truth and prediction was set to identify successful or hit-detections. No-detections were identified as arteries with ground-truth labels but no output prediction and miss-detections were defined as arteries with ground-truth labels and an inaccurate prediction ($IoU < 50\%$). Furthermore, the MAD in pixels between the artery centre prediction and the ground-truth or nearest ground-truth annotation was also computed. The detection accuracy was computed as percentage of hit-detections with respect to all ground-truth slices as observed in Equation 4.6.

$$\% \text{ Detection} = \frac{\text{N}^\text{o} \text{ hit-detections}}{\text{N}^\text{o} \text{ ground-truth bounding box}} * 100 \tag{4.6}$$

Table 4.1: YOLOv4 artery localization results.

Carotid type	Ground-truth	No-detection	Miss-detection	Hit-detection	Average IoU (%)	MAD (pixel)	% Detection
ICA	2279	154	50	2075	72.1 ± 9.3	2.5 ± 1.5	91.0
ECA	486	83	34	369	72.4 ± 8.6	2.3 ± 1.3	75.9
CCA	1234	31	8	1195	79.6 ± 7.5	2.2 ± 1.2	96.8
BF	190	19	2	178	80.3 ± 8.2	2.2 ± 1.4	93.7
Total	4189	292	94	3803	76.1 ± 8.4	2.3 ± 1.4	89.4

ICA: internal carotid artery, ECA: external carotid artery, CCA: common carotid artery, BF: bifurcation. MAD: mean absolute distance, IoU: Intersection over Union.

As can be observed in Table 4.1, the object detection system was able to detect artery locations robustly. This can be seen in the high accuracy of detection results in different carotid types, especially in the BF region (93.7%), a structure known to be highly heterogeneous and therefore challenging to detect. On the other hand, the accuracy of detection is significantly lower for ECA (75.9%), this was to be expected due to the smaller number of annotations provided for this carotid type. Furthermore, the ECA has generally a smaller diameter than the ICA and can therefore be more challenging to detect by the object detection algorithm. The mean artery centre deviation was of 2.3 pixels which corresponds to approximately 0.7 mm. Given that the carotid artery lumen diameter is between 4 and 7 mm [106] we considered it as high accuracy results.

For a qualitative evaluation of the detection results, Figure 4.1 illustrates samples of successful artery predictions for 3 slices from 3 distinct patients and for different carotid types. Pink bounding boxes represent the predictions from the object detection algorithm and green bounding boxes represent the ground-truth.
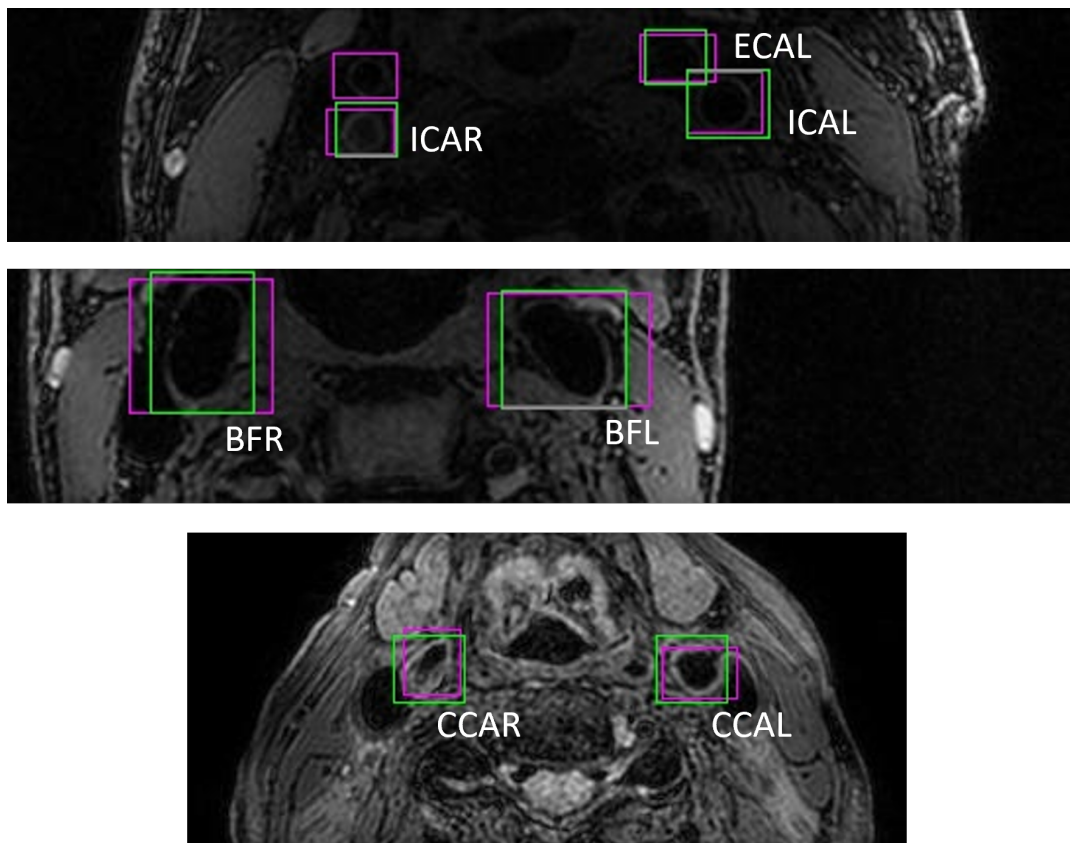
Figure 4.1: Qualitative evaluation. Hit-detections in 3 sample slices. Green bounding box: ground-truth and pink bounding box: prediction

Figure 4.2 illustrates two sample slices of failed detections, signalled by a red arrow. Figure 4.2 above represents an example of miss-detection and Figure 4.2 below of a no-detection. In this case, the miss-detection can be identified by its no overlap with a corresponding ground-truth box when only one bounding box is expected, as it corresponds to the CCA slice location.
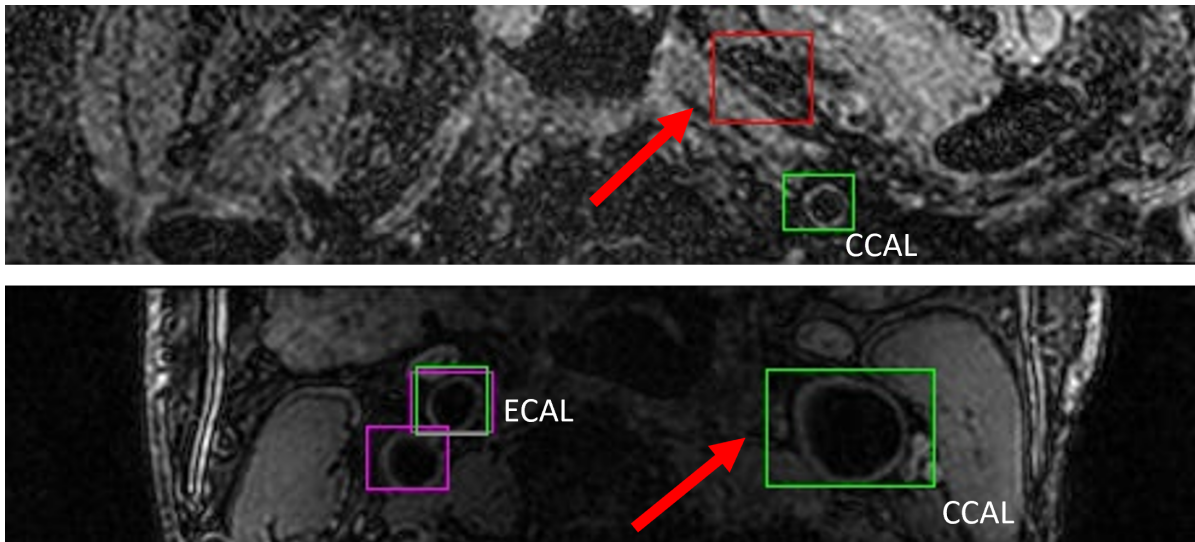
Figure 4.2: Qualitative evaluation. Failed detection in 2 sample slices. Green bounding box: ground-truth and pink bounding box: prediction. Red arrow indicates failed artery detection.

As can be qualitatively observed in Figure 4.1, the localization system was capable of accurately detecting different carotid types with different morphology. Especially, in the BF slice (Figure 4.1 below), which is a challenging region. Furthermore, this was performed in slices with different image quality which further proves the robustness of the algorithm.

On the other hand, the algorithm also failed in some slices as can be observed in Figure 4.2. The red bounding box in Figure 4.2 above illustrates an example of miss-detection where the detected bounding box has a 0% overlap with the corresponding ground-truth annotation. This could be due to the observable lower image quality in the above image. Furthermore, the similar image contrast properties of the miss-detected structure with the target carotid artery could have further confounded the algorithm. Figure 4.2 below represents an example of a no-detection where the algorithm failed to output a prediction for the target carotid type. This could be due to the markedly large size of the non-detected carotid artery compared with other ICA or ECA structures observed in the slice.

4.1.2 Vessel centreline tracking

The vessel centreline was spatially reconstructed from the output predictions and compared with the reconstruction of the available ground-truth annotations for a more comprehensive evaluation of the artery localization results. Furthermore, this provided additional spatial information for the identification and posterior correction of miss-detections.

The qualitative evaluation of the vessel centreline refinement method for one patient is illustrated in Figure 4.3. Figure 4.3 above illustrates the results of the 3D vessel centreline reconstruction graphs before (Figure 4.3(a)) and after (Figure 4.3(b)) the proposed refinement step is applied. x and y-axis correspond to localization of artery centre in each slice (z-axis). Blue circles correspond to the localization of predicted artery s and brown circles to the localization of the available ground-truth artery centres. The labelled points represent the identified miss-detections though vessel-centreline reconstruction. To further illustrate the refinement applied a sample miss-detection (indicated by red arrow) was selected and the miss-detection, prediction and ground-truth bounding box of the corresponding slice are illustrated below. This is shown before (Figure 4.3(c)) and after (Figure 4.3(d)) refinement step is applied.

Table 4.2 shows the number of additional miss-detections (from predictions with no associated ground-truth) detected with vessel centreline reconstruction (N=44). For a quantitative evaluation of the vessel centreline refinement, the MAD in pixels between the artery centre and nearest labelled ground-truth was computed for the identified miss-detections before and after refinement.

Table 4.2: Quantitative evaluation of vessel centreline refinement

N° of Miss-detections	MAD before refinement (pixel)	MAD after refinement (pixel)
44	31.6	2.75

MAD: mean absolute distance

As can be observed in Figure 4.3, by tracking the artery centres in consecutive slices the vessel centreline was reconstructed in 3D where the left and right CCA, ECA and ICA branches can be easily distinguished. The high-degree of overlap between the ground-truth (brown) and predicted (blue) artery centres illustrates the high accuracy results achieved by the object detection algorithm. Furthermore, as observed by the labelled points in Figure 4.3(a),in some slices the object detection algorithm predicted incorrect artery localizations. Due to their spatial deviation with respect to the reconstructed centreline they could be easily identified as miss-detections, even in slices with no ground-truth. As observed in Figure 4.3(b) after the refinement step has been applied,the number of spatially deviated artery centres reduces. This can be further observed in Figure 4.3(c) and Figure 4.3(d) which show the refinement of the sample miss-detection (red bounding box) signalled by the red arrow. In Table 4.2 the effectiveness of this approach can be further observed as the spatial distance between the miss-detected artery centre localization and the nearest ground-truth significantly reduces.
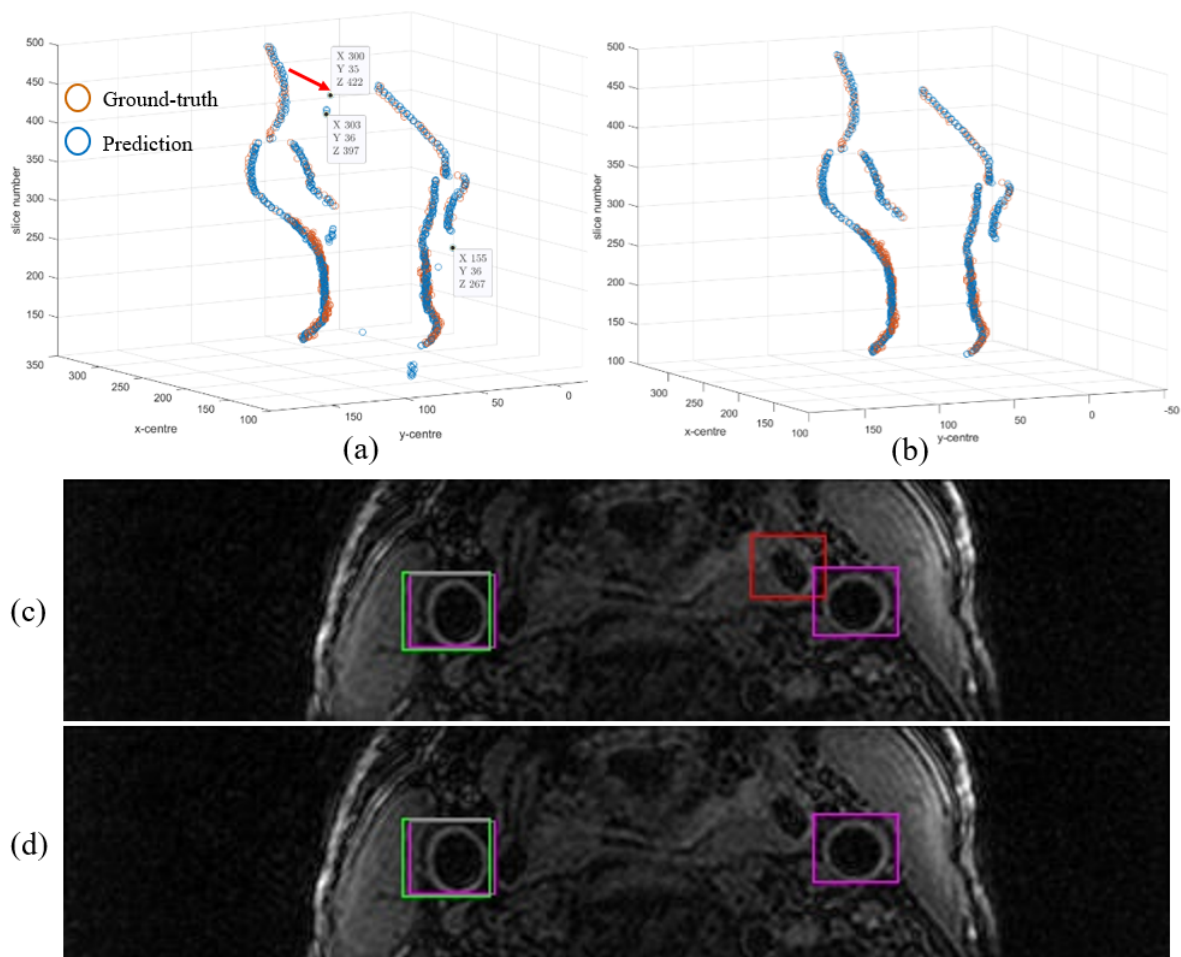
Figure 4.3: Qualitative evaluation of proposed vessel centreline refinement method. (a)-(b): Vessel centreline 3D reconstruction graph (a) before refinement (b) after refinement. Labelled circles correspond to identified miss-detections. (c)-(d): Slice location of sample miss-detection (red arrow) (c) before refinement (d) after refinement. Red bounding box: miss-detection, green bounding box: ground-truth, pink bounding box: prediction

4.2 Vessel wall segmentation

4.2.1 Vessel wall segmentation framework

For a quantitative evaluation of the proposed VW segmentation framework, different networks were trained with the same experimental data.

Table 4.3 shows the quantitative segmentation results achieved by the proposed VW segmentation framework, referred to as 2D DTML. The results are compared with a modified U-Net as a reference corresponding to the backbone architecture (section 3.2.2) of the proposed segmentation framework with only one output segmentation branch. Furthermore, to evaluate the effect of SSL with shape constraints in other frameworks the 2D version of a state-of-the-art shape aware semi-supervised framework (SASSnet [54]) with the same backbone architecture was also trained and evaluated with the same experimental data. Different from the implemented framework, SASSnet integrates an adversarial loss for consistency training.

Table 4.3: Quantitative evaluation. Comparison of SSL frameworks. N=4089.

Method	Labelled	Unlabelled	Dice $\pm \sigma$ (%)	Jaccard $\pm \sigma$ (%)	HD $\pm \sigma$ (%)
Mod.U-Net	2078	0	71.66 ± 15.11	57.70 ± 16.01	4.62 ± 4.63
2D SASSNet [54]	2078	8665	72.62 ± 15.21	58.92 ± 16.26	4.38 ± 4.19
2D DTML	2078	8665	72.96 ± 15.30	59.38 ± 16.40	4.30 ± 4.15

σ: standard deviation.

An ablation study was performed to individually evaluate the effect of shape constraints and semi-supervised strategies on the accuracy of segmentation (Table 4.4). L_{seg} corresponds to the backbone network only trained for binary mask segmentation. $L_{seg} + L_{SDF}$ corresponds to the backbone network with combined prediction of binary mask segmentation and SDM.

Table 4.4: Ablation study. N=4089

Method	Labelled	Unlabelled	Dice $\pm \sigma$ (%)	Jaccard $\pm \sigma$ (%)	HD $\pm \sigma$ (%)
L_{seg}	2078	0	71.66 ± 15.11	57.70 ± 16.01	4.62 ± 4.63
$L_{seg} + L_{SDF}$	2078	0	71.81 ± 15.53	58.00 ± 16.53	4.55 ± 4.58
2D DTML	2078	8665	72.96 ± 15.30	59.38 ± 16.40	4.30 ± 4.15

σ: standard deviation.

As can be observed in Table 4.3, compared with the fully supervised U-Net, both shape aware semi-supervised networks show improved segmentation performance on all evaluation metrics. This would suggest the capability of semi-supervised frameworks in taking advantage of unlabelled images and that shape priors are useful for semi-supervised image segmentation. Notably, the implemented 2D DTML achieves slightly better performance than 2D SASSNet which suggests that cross-task consistency learning with two individual networks is effective in drawing rich information from unlabelled data.

Table 4.4 shows the quantitative results of the ablation study performed. The slight accuracy improvement after the introduction of SDM regression to a fully supervised framework (second row) suggest that shape priors have the potential of increasing segmentation performance by imposing learnt shape constraints on the final segmentation prediction. Furthermore, the further improvement shown in the cross-task consistency learning strategy (DTML) further proves the importance of leveraging information from unlabelled data. This improvement could also be due to the task level regularization imposed by the consistency loss which guides the mutual learning of both networks and improves the generalization ability of the model.

4.2.2 Persistent homology post-processing module

Table 4.5 shows the quantitative result of the topological post-processing assessed against several baselines. The first and second row correspond to the evaluation of fully supervised U-Net and the proposed semi-supervised framework (2D DTML). The third row corresponds to the DTML framework results after the application of a largest connected component (LCC) algorithm that extracts the largest binary object as the target. The final row corresponds to the proposed post-processing step that applies a PH-based topological loss function [22]. The topological performance was evaluated by the proportion of predicted segmentations that demonstrated the correct topology: one connected component with one hole inside.

Table 4.5: Persistent homology topological constraint evaluation. N=4089

Model	% Correct topologies	Dice $\pm \sigma$ (%)	HD $\pm \sigma$ (%)
Mod. U–Net	82.2	71.66 ± 15.11	4.62 ± 4.63
2D DTML	91.8	72.96 ± 15.30	4.30 ± 4.15
2D DTML + LCC	94.9	72.98 ± 15.30	4.24 ± 4.08
2D DTML + LCC + PH	98.7	72.98 ± 15.33	4.23 ± 4.10

σ: standard deviation.

The effect of the PH post-processing step on the shape aware semi-supervised framework can be observed in Figure 4.4.

Figure 4.5 shows the qualitative evaluation of 4 patients with the fully proposed VW segmentation module. First, after the application of a shape aware semi-supervised framework (DTML) with respect to a reference modified U-Net and then, after the application of PH post processing module in slices with topologically incorrect predictions (columns 3 and 4).
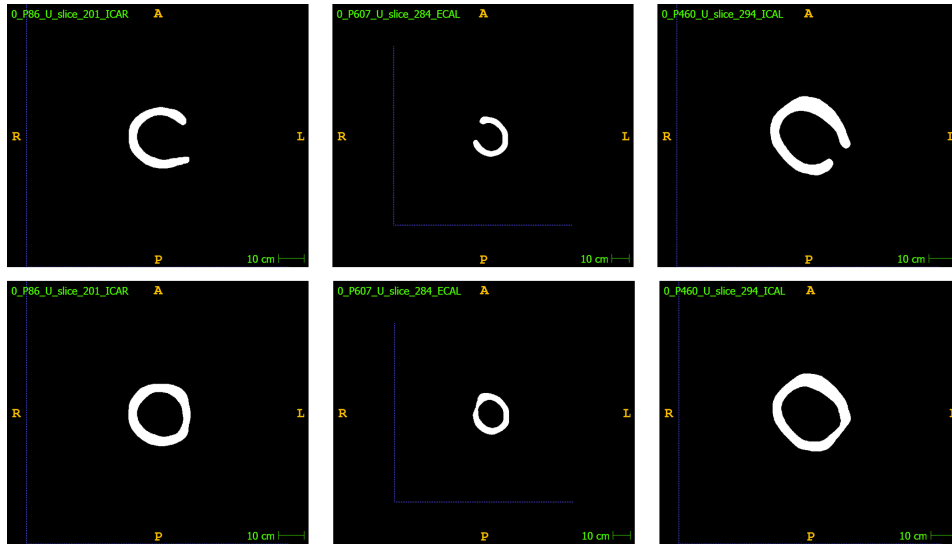
Figure 4.4: Qualitative evaluation of PH module in 3 sample slices.Above: 2D DTML prediction. Below: 2D DTML + PH post-processing prediction
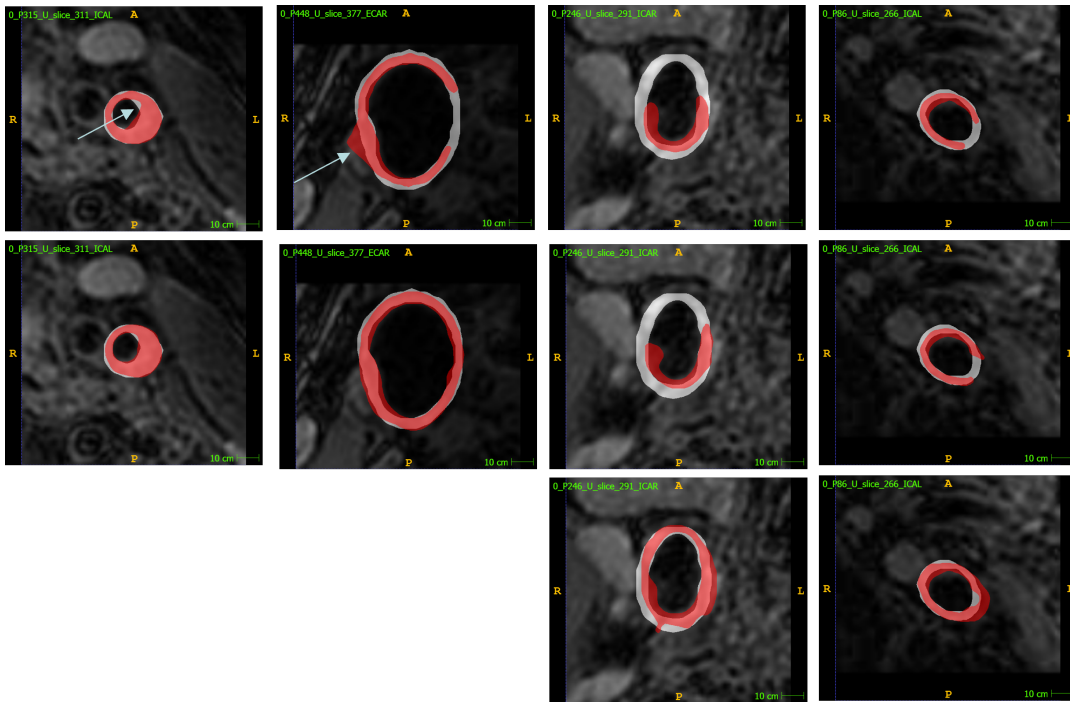
Figure 4.5: Qualitative evaluation of complete segmentation framework in 4 sample slices (columns). First row: Reference modified U-Net, Second row: 2D DTML framework, Third row: 2D DTML + PH (only applied in cases with incorrect topology). Ground-truth shown in gray, prediction in red and intersection in orange. The white arrow indicates highlighted regions.

As can be observed in Table 4.5 , the proposed 2D DTML showed superior percentage of correct topologies with respect to the reference modified U-Net. This could be due to the increased shape awareness provided by the combined learning of binary segmentations and SDM. The third row shows the results for the proposed DTML framework after the application of a LCC algorithm. As seen in the table, the number of correct topologies increases as LCC only keeps one connected component (the largest) in the binary image. However, the percentage of correct topologies is still not optimal. The last row shows the segmentation results after the introduction of the PH based loss function to the post processing module. As can be seen, after its application the proportion of correct topologies significantly increases reaching to a near complete percentage of correctness (98.7% of the predictions are topologically correct). Different from the LCC algorithm, the PH topological loss function imposes further topological constraints such as the expectation that VW predictions must be closed structures with a hole inside. Therefore, the increased topological correctness could be due to the correction of C-shaped or broken-down VW predictions. This is further illustrated in Figure 4.4 .

It must be remarked that the improvement of the PH post-processing module cannot be observed through the Dice metric. This could be due to several reasons. Firstly, the PH module is only applied on output segmentations with incorrect topologies, which after the application of the LCC algorithm only accounts for 5% of the testing slices. Furthermore, correcting C-shaped predictions of the VW could generate a difference of only a small number of pixels which would not cause observable improvements with respect to the spatial overlap between prediction and ground-truth segmentation. Although regional metrics such as the Dice do not show significant improvement, we consider that the application of PH as a post-refinement step is meaningful, especially for the task of VW segmentation. Clinical metrics such as the plaque burden, require a closed lumen and outer-wall prediction for their extraction [8, 107].

Figure 4.5 illustrates the visualization results of the entire segmentation module proposed in this work, first after the implementation of the shape aware DTMLframework (second row) and then after the application of the PH-based post-processing module (third row) compared with the output results of the reference modified U-Net (first row). This was shown for 4 sample patients (columns). In the first two columns, we can observe that DTML predictions have a higher overlap with the ground-truth mask compared with the reference modified U-Net. The increased overlap (orange) can be seen both for the inner and outer-wall contours, as signalled by the white arrows. This further suggests the effectiveness of introducing knowledge from unlabelled slices and applying shape constraints to improve the accuracy of segmentation. Furthermore, as seen in Figure 4.5-column 2, this improvement also occurs in carotid BF regions which shows that applying shape constraints to DL models could aid in the segmentation of heterogeneous regions. Exploiting unlabelled data for additional information could have also aided in the segmentation of heterogeneous VW morphologies due to the increased the generalization ability of the model. In the last two columns, the DTML failed to output topologically meaningful predictions, this could be due to the complex signal characteristics of the image. In those cases the PH-based post-processing module

was applied.As can be observed, the PH module successfully corrects the broken-down VW prediction, and it does so by generating predictions with a high degree of overlap with respect to the ground-truth.

Finally, the Dice VW segmentation results achieved with the proposed VW segmentation and post-processing module (72.98%) were above the average of the submitted works to the CAVWSC 2021 segmentation challenge [10] (71.70 %) which further proves the effectiveness of semi-supervised frameworks and of the integration of SDM and topological shape constraints for the task of VW segmentation. Achieving high accuracy and topologically correct VW predictions is required for the extraction of quantitative metrics that characterize luminal stenosis and plaque morphology in clinical applications.

5 Conclusions

In this thesis, inspired by the CAVWSC 2021 challenge, we propose an automatic framework for 2D carotid VW MRI segmentation based on deep SSL with shape priors. The aim was to achieve high accuracy of segmentation for 3D-MERGE images resliced in the axial plane. The main challenges encountered where the variability of the carotid vessel wall morphology as difference carotid branches (CCA, ICA and ECA) were considered in the segmentation task, the small number of annotated slices provided and the varying degrees of image quality.

The proposed VW segmentation framework consists of a artery localization and segmentation module. The localization module enabled the automatic localization of the carotid vessel wall for posterior ROI extraction. For this the YOLOv4 object algorithm was custom trained for carotid center localization and the initial predictions were refined with a vessel centerline refinement method. The proposed method tracked the artery center localization in consecutive slices and allowed the identification and correction of miss-detections. As shown in the results, after the refinement, high accuracy of detection was achieved for different carotid types.

The second part consisted of a VW segmentation module that enforces shape constraints at two levels. First, inspired by [55] a DTML framework was implemented that imposed geometrical shape constraints with the integration of SDM regression. The DTML framework consisted of two individual dual-task networks (for segmentation and SDM prediction) that learnt collaboratively, similar to mutual learning methods, and were joined through a cross-task consistency loss. The cross-task consistency loss leveraged rich information from unlabelled slices and imposed a task-level regularization that increased the generalization ability of the network. Then, test-time adaptation was performed with a PH-based topological loss function [50] that generated topologically meaningful predictions.

The results obtained show that challenging problems such as VW segmentation could benefit from the integration of shape constraints into the DL framework both during training, through SDM regression, and at a post-processing stage through topological priors. Furthermore, the results further show that task-level consistency regularization enables the extraction of meaningful information from unlabelled slices that can further improve the VW segmentation accuracy in different carotid branches.

High accuracy of segmentation is necessary for the integration of DL-based VW segmentation models in the clinical practice. Risk predictors such as plaque burden require for the VW segmentation to be topologically correct, with no broken-down predictions for their extraction. On the other hand, due to the challenging nature of VW-MRI segmentation, high accuracy segmentation is not achieved in some cases. To further improve the performance of the model we propose, as future work, to smooth the outer-wall contour predictions with the integration of a curvature-based loss function.

6 Future work

The results presented in this thesis suggest that the integration of shape priors can improve the segmentation accuracy of DL models for the task of carotid VW-MRI segmentation. Achieving high accuracy and topologically correct VW segmentation predictions is crucial for the extraction of quantitative metrics that characterize luminal stenosis and plaque morphology in clinical applications. However, the challenging nature of VW segmentation problems such as neighbouring vessels and the weak signal characteristics of MRI images have led, in several instances, to output segmentations with non-smooth contours even after the integration of the proposed shape constraints. This has been especially observed in outer-wall contour predictions. Further work should be done to increase the accuracy of the VW segmentation model, with the final aim of its implementation in the clinical practice.

Given the encouraging results obtained after the integration of shape priors we propose, as future work, an additional shape constraint that encourages the smoothness of contour predictions such as the outer wall. For this, we propose the minimization of a curvature-based loss function. The curvature measures the angle of a curve with a fixed direction at a unit length. In a circular shape, given the regularity or smoothness of the curve the curvature measurements at different points should not vary, therefore, the standard deviation of the curvature should tend to 0 [108, 109].

Figure 6.1 illustrates the outer-wall curvature for different VW segmentation predictions. The graphs represent the magnitude and direction of the curvature at different points in the curve, which are further colour-coded in the binary mask. The standard deviation of the curvature is shown in the purple boxes above each prediction segmentation. The two figures above correspond to predictions with non-smooth or inaccurate contours whilst the figures below correspond to smooth VW contour predictions. As can be observed in the curvature graphs, the variation of the curvature is significantly higher in the non-smooth predictions. This is further shown in the increased value of the corresponding standard deviation of the curvature which appears significantly higher in predictions with non-smooth contours.

Therefore, to encourage the prediction of smooth contours we aim to constrain the standard deviation of the curvature to values that tend to 0. For this, we propose as a loss function the standard deviation of the curvature function. This is defined in Equation 6.1 [109].

$$Loss_{curvature} = stdev \left(\frac{x'' * y' - x' * y''}{(x' * x' + y' * y')^{1.5}} \right) \tag{6.1}$$

Where x' and y' represent the first derivation of the x and y contour coordinates, respectively. x'' and y'' represents the second derivation and $stdev()$ represents the standard deviation.
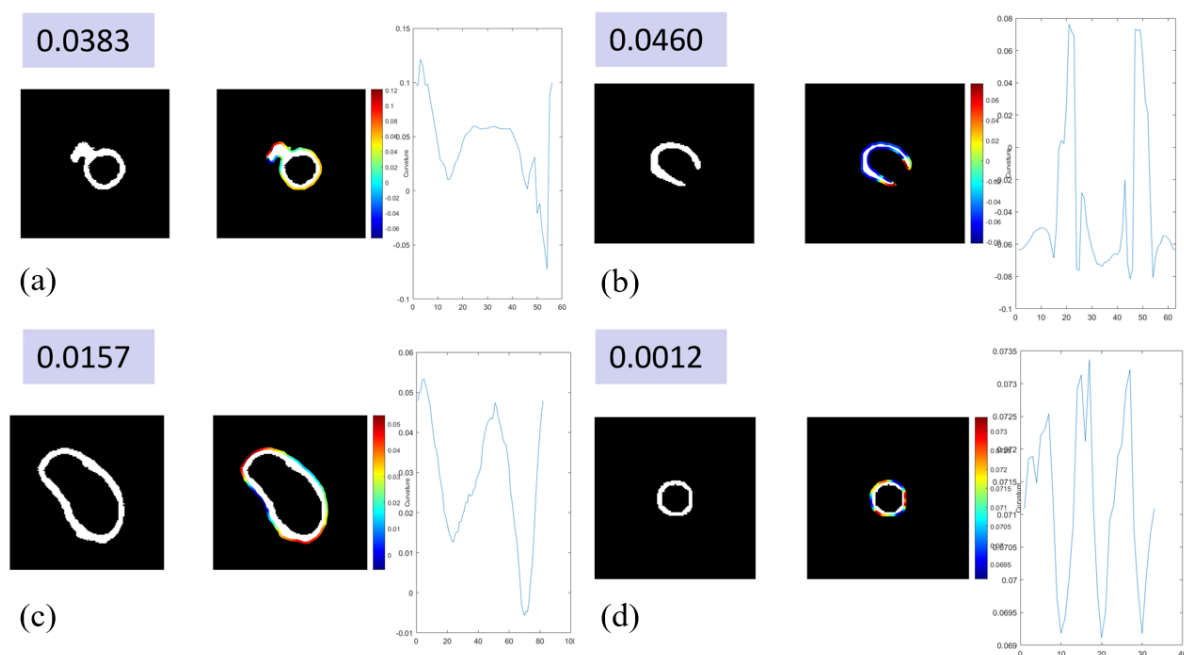
Figure 6.1: 4 sample segmentation predictions with corresponding outer-wall curvature graphs. (a)-(b): non-smooth contour predictions. (c)-(d): smooth contour predictions. The purple box shows the standard deviation of the curvature.

Through the integration of the curvature-based loss function in the proposed framework, the smoothness of the outer-wall contours could be enforced. This would lead to higher accuracy carotid VW segmentation predictions and to step forward towards the integration of DL-based algorithms in the clinical practice for the evaluation of carotid atherosclerosis.

Acronyms

3D-MERGE three-dimensional - Motion-Sensitized Driven Equilibrium(MSDE)-prepared Rapid Gradient Echo. iv, viii, 1, 2, 7, 8, 12–14, 59

AGE advanced glycation end products. 4

BF bifurcation. 2, 49–51, 57

CARE II Chinese Atherosclerosis Risk Evaluation. viii, 13, 14, 45, 47

CASCADE Computer-Aided System for Cardiovascular Disease Evaluation. 14, 49

CAVWSC Carotid Artery Vessel Wall Segmentation Challenge. 1, 13, 45, 47, 58, 59

CCA common carotid artery. iv, 1, 2, 4, 11, 13, 14, 38, 40, 47, 49, 52, 59

CEA carotid endarterectomy. 1, 4, 5

CNN Convolutional Neural Network. 11, 15–18, 24, 26, 35, 44

CRF Conditional Random Field. 21

CT Computed Tomography. 6

CTA Computed Tomography Angiography. 5

CVD Cardiovascular disease. 1

DL deep learning. iv, 1, 2, 9, 18, 24, 57, 59, 61

DML deep mutual learning. 23, 24, 41

DTML dual-task mutual learning. ix, 24, 33, 41, 54–57, 59

ECA external carotid artery. iv, 1, 2, 4, 13, 14, 37, 38, 40, 47, 49–52, 59

EMA exponential moving average. 23

FC fibrous cap. 3, 6

FCN Fully Convolutional Network. 11, 18

HD Haussdorf Distance. 47, 48, 54

ICA internal carotid artery. iv, viii, 1, 2, 4, 11, 13–15, 37, 38, 40, 47, 49–52, 59

IoU Intersection over Union. 17, 46, 47, 49

IPH intraplaque haemorrhage. 1, 5, 6

IVUS intravascular ultrasound. 11

LCC largest connected component. 55, 57

LDL low-density lipoprotein. 3

LRNC lipid-rich necrotic core. viii, 6, 7

MAD mean absolute distance. 47, 49, 52

MDCT Multidetector Computed Tomography. 5, 6

MRA Magnetic Resonance Angiography. viii, 5, 7, 11

MRI Magnetic Resonance Imaging. iv, 1, 2, 5–8, 11, 13, 59, 61

MSDE Motion Sensitized Driven Equilibrium. 8

MSE mean square error. 22, 43, 45

PDW proton density-weighted. 11

PH persistent homology. iv, ix, 12, 26–28, 33, 44–46, 55–59

R-CNN Region based Convolutional Neural Network. 16

ReLU Rectified Linear Unit. 19

ROI region of interest. iv, 2, 11, 12, 33, 35, 40, 46, 59

SDM signed distance map. iv, 12, 20, 24–26, 31–33, 41–43, 54, 55, 57–59

SGD Stochastic Gradient Descent. 46

sRAGE soluble form of receptor for advanced glycation end products. 4

SSL semi-supervised learning. iv, xi, 2, 12, 21, 24, 33, 54, 59

T1W T1-weighted. viii, 6, 7

T2W T2-weighted. viii, 6, 7

TIA transient ischemic attack. iv, 1, 6

TOF time-of-flight. viii, 6, 7, 11

US Ultrasound. 5, 6

VW vessel wall. iv, viii, ix, 1, 2, 4, 6, 7, 9, 11–14, 21, 24, 26, 28, 31–33, 35, 40, 42, 44–48, 54, 55, 57–59, 61, 62

VWI vessel wall imaging. 5, 6

VW-**MRI** vessel wall-Magnetic Resonance Imaging. iv, 1, 2, 7, 11, 12, 61

YOLO you only look once. 16, 17, 29, 30, 35, 36, 38, 46, 49, 59

Bibliography

[1] *World Health Organization (WHO). Cardiovascular Diseases.* `https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)`. Accessed: 2021-08-10.

[2] L. Saba, C. Yuan, T. Hatsukami, N. Balu, Y. Qiao, J. DeMarco, T. Saam, A. Moody, D. Li, C. Matouk, et al. "Carotid artery wall imaging: perspective and guidelines from the ASNR vessel wall imaging study group and expert consensus recommendations of the American Society of Neuroradiology". In: *American Journal of Neuroradiology* 39.2 (2018), E9–E31.

[3] K. Prasad. "Pathophysiology and medical treatment of carotid artery stenosis". In: *International Journal of Angiology* 24.03 (2015), pp. 158–172.

[4] S. Howell. "Carotid endarterectomy". In: *British journal of anaesthesia* 99.1 (2007), pp. 119–131.

[5] C. C. Phatouros, R. T. Higashida, A. M. Malek, P. M. Meyers, T. E. Lempert, C. F. Dowd, and V. V. Halbach. "Carotid artery stent placement for atherosclerotic disease: rationale, technique, and current status". In: *Radiology* 217.1 (2000), pp. 26–41.

[6] J. K. DeMarco and J. Huston. "Imaging of high-risk carotid artery plaques: current status and future directions". In: *Neurosurgical focus* 36.1 (2014), E1.

[7] N. Balu, V. L. Yarnykh, B. Chu, J. Wang, T. Hatsukami, and C. Yuan. "Carotid plaque assessment using fast 3D isotropic resolution black-blood MRI". In: *Magnetic resonance in medicine* 65.3 (2011), pp. 627–637.

[8] K. Murata, N. Murata, B. Chu, H. Watase, D. S. Hippe, N. Balu, J. Sun, X. Zhao, T. S. Hatsukami, C. Yuan, et al. "Characterization of carotid atherosclerotic plaques using 3-dimensional merge magnetic resonance imaging and correlation with stroke risk factors". In: *Stroke* 51.2 (2020), pp. 475–480.

[9] X. Liu, L. Song, S. Liu, and Y. Zhang. "A review of deep-learning-based medical image segmentation methods". In: *Sustainability* 13.3 (2021), p. 1224.

[10] *Carotid Artery Vessel Wall Segmentation Challenge.* `https://vessel-wall-segmentation.grand-challenge.org/`. Accessed: 2021-08-10.

[11] P. Castelnuovo, I. Dallan, and M. Tschabitscher. *Surgical anatomy of the internal carotid artery.* Springer, 2016.

[12] *TeachMeAnatomy. Major Arteries of the Head and Neck.* `https://teachmeanatomy.info/neck/vessels/arterial-supply/`. Accessed: 2021-08-10.

[13] H. A. Hall and H. S. Bassiouny. "Pathophysiology of carotid atherosclerosis". In: *Ultrasound and Carotid Bifurcation Atherosclerosis*. Springer, 2011, pp. 27–39.

[14] Z. Ozgur, F. Govsa, and T. Ozgur. "Anatomic evaluation of the carotid artery bifurcation in cadavers: implications for open and endovascular therapy". In: *Surgical and Radiologic Anatomy* 30.6 (2008), p. 475.

[15] N. d. B. de la Arteria and C. C. en Relación. "Common carotid artery bifurcation levels related to clinical relevant anatomical landmarks". In: *Int. J. Morphol* 24.3 (2006), pp. 413–416.

[16] R. Cobiella, S. Quinones, M. Konschake, P. Aragones, X. León, T. Vazquez, J. Sanudo, and E. Maranillo. "The carotid axis revisited". In: *Scientific Reports* 11.1 (2021), pp. 1–9.

[17] R. Ross. "The pathogenesis of atherosclerosis—an update". In: *New England journal of medicine* 314.8 (1986), pp. 488–500.

[18] J. W. Gofman, F. Lindgren, H. Elliott, W. Mantz, J. Hewitt, B. Strisower, V. Herring, and T. P. Lyon. "The role of lipids and lipoproteins in atherosclerosis". In: *Science* 111.2877 (1950), pp. 166–186.

[19] *Atherosclerosis. University of Michigan.* `https://www.uofmhealth.org/health-library/ue4664abc`. Accessed: 2021-08-10.

[20] *Carotid Artery Disease. University of California San Francisco.* `https://surgery.ucsf.edu/conditions--procedures/carotid-artery-disease.aspx`. Accessed: 2021-08-10.

[21] H. J. Barnett, D. W. Taylor, M. Eliasziw, A. J. Fox, G. G. Ferguson, R. B. Haynes, R. N. Rankin, G. P. Clagett, V. C. Hachinski, D. L. Sackett, et al. "Benefit of carotid endarterectomy in patients with symptomatic moderate or severe stenosis". In: *New England Journal of Medicine* 339.20 (1998), pp. 1415–1425.

[22] T. M. Freilinger, A. Schindler, C. Schmidt, J. Grimm, C. Cyran, F. Schwarz, F. Bamberg, J. Linn, M. Reiser, C. Yuan, et al. "Prevalence of nonstenosing, complicated atherosclerotic plaques in cryptogenic stroke". In: *JACC: Cardiovascular Imaging* 5.4 (2012), pp. 397–405.

[23] J. Lovett, P. Gallagher, and P. Rothwell. "Reproducibility of histological assessment of carotid plaque: implications for studies of carotid imaging". In: *Cerebrovascular diseases* 18.2 (2004), pp. 117–123.

[24] A. E. Zavodni, B. A. Wasserman, R. L. McClelland, A. S. Gomes, A. R. Folsom, J. F. Polak, J. A. Lima, and D. A. Bluemke. "Carotid artery plaque morphology and composition in relation to incident cardiovascular events: the Multi-Ethnic Study of Atherosclerosis (MESA)". In: *Radiology* 271.2 (2014), pp. 381–389.

[25] B. A. Wasserman, W. I. Smith, H. H. Trout, R. O. Cannon, R. S. Balaban, and A. E. Arai. "Carotid artery atherosclerosis: in vivo morphologic characterization with gadolinium-enhanced double-oblique MR imaging—initial results". In: *Radiology* 223.2 (2002), pp. 566–573.

[26] N. Altaf, S. T. MacSweeney, J. Gladman, and D. P. Auer. "Carotid intraplaque hemorrhage predicts recurrent symptoms in patients with high-grade carotid stenosis". In: *Stroke* 38.5 (2007), pp. 1633–1635.

[27] T. Saam, M. Ferguson, V. Yarnykh, N. Takaya, D. Xu, N. Polissar, T. Hatsukami, and C. Yuan. "Quantitative evaluation of carotid plaque composition by in vivo MRI". In: *Arteriosclerosis, thrombosis, and vascular biology* 25.1 (2005), pp. 234–239.

[28] M. E. Porambo and J. K. DeMarco. "MR imaging of vulnerable carotid plaque". In: *Cardiovascular diagnosis and therapy* 10.4 (2020), p. 1019.

[29] J. Sun, H. R. Underhill, D. S. Hippe, Y. Xue, C. Yuan, and T. S. Hatsukami. "Sustained acceleration in carotid atherosclerotic plaque progression with intraplaque hemorrhage: a long-term time course study". In: *JACC: Cardiovascular imaging* 5.8 (2012), pp. 798–804.

[30] H. R. Underhill, T. Hatsukami, J. Cai, W. Yu, J. DeMarco, N. Polissar, H. Ota, X. Zhao, L. Dong, M. Oikawa, et al. "A noninvasive imaging approach to assess plaque severity: the carotid atherosclerosis score". In: *American Journal of Neuroradiology* 31.6 (2010), pp. 1068–1075.

[31] X. Zhao, R. Li, D. S. Hippe, T. S. Hatsukami, C. Yuan, C.-I. Investigators, et al. "Chinese Atherosclerosis Risk Evaluation (CARE II) study: a novel cross-sectional, multicentre study of the prevalence of high-risk atherosclerotic carotid plaque in Chinese patients with ischaemic cerebrovascular events—design and rationale". In: *Stroke and vascular neurology* 2.1 (2017).

[32] C. Yuan, E. Lin, J. Millard, and J.-.-N. Hwang. "Closed contour edge detection of blood vessel lumen and outer wall boundaries in black-blood MR images". In: *Magnetic resonance imaging* 17.2 (1999), pp. 257–266.

[33] H. M. Ladak, J. B. Thomas, J. R. Mitchell, B. K. Rutt, and D. A. Steinman. "A semi-automatic technique for measurement of arterial wall from black blood MRI". In: *Medical Physics* 28.6 (2001), pp. 1098–1107.

[34] H. R. Underhill, W. S. Kerwin, T. S. Hatsukami, and C. Yuan. "Automated measurement of mean wall thickness in the common carotid artery by MRI: a comparison to intima-media thickness by B-mode ultrasound". In: *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine* 24.2 (2006), pp. 379–387.

[35] G. Unal, S. Bucher, S. Carlier, G. Slabaugh, T. Fang, and K. Tanaka. "Shape-driven segmentation of the arterial wall in intravascular ultrasound images". In: *IEEE Transactions on Information Technology in Biomedicine* 12.3 (2008), pp. 335–347.

[36] W. Liu, N. Balu, J. Sun, X. Zhao, H. Chen, C. Yuan, H. Zhao, J. Xu, G. Wang, and W. S. Kerwin. "Segmentation of carotid plaque using multicontrast 3D gradient echo MRI". In: *Journal of Magnetic Resonance Imaging* 35.4 (2012), pp. 812–819.

[37]   R. van't Klooster, P. J. de Koning, R. A. Dehnavi, J. T. Tamsma, A. de Roos, J. H. Reiber, and R. J. van der Geest. "Automatic lumen and outer wall segmentation of the carotid artery using deformable three-dimensional models in MR angiography and vessel wall images". In: *Journal of Magnetic Resonance Imaging* 35.1 (2012), pp. 156–165.

[38]   A. M. Arias-Lorza, J. Petersen, A. van Engelen, M. Selwaness, A. van der Lugt, W. J. Niessen, and M. de Bruijne. "Carotid artery wall segmentation in multispectral MRI by coupled optimal surface graph cuts". In: *IEEE transactions on medical imaging* 35.3 (2015), pp. 901–911.

[39]   L. Luo, S. Liu, X. Tong, P. Jiang, C. Yuan, X. Zhao, and F. Shang. "Carotid artery segmentation using level set method with double adaptive threshold (DATLS) on TOF-MRA images". In: *Magnetic resonance imaging* 63 (2019), pp. 123–130.

[40]   K. Simonyan and A. Zisserman. "Very deep convolutional networks for large-scale image recognition". In: *arXiv preprint arXiv:1409.1556* (2014).

[41]   J. Long, E. Shelhamer, and T. Darrell. "Fully convolutional networks for semantic segmentation". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015, pp. 3431–3440.

[42]   O. Ronneberger, P. Fischer, and T. Brox. "U-net: Convolutional networks for biomedical image segmentation". In: *International Conference on Medical image computing and computer-assisted intervention*. Springer. 2015, pp. 234–241.

[43]   F. Shi, Q. Yang, X. Guo, T. A. Qureshi, Z. Tian, H. Miao, D. Dey, D. Li, and Z. Fan. "Intracranial vessel wall segmentation using convolutional neural networks". In: *IEEE Transactions on Biomedical Engineering* 66.10 (2019), pp. 2840–2847.

[44]   J. Wu, J. Xin, X. Yang, J. Sun, D. Xu, N. Zheng, and C. Yuan. "Deep morphology aided diagnosis network for segmentation of carotid artery vessel wall and diagnosis of carotid atherosclerosis on black-blood vessel wall MRI". In: *Medical physics* 46.12 (2019), pp. 5544–5561.

[45]   L. Chen, J. Sun, G. Canton, N. Balu, X. Zhao, R. Li, T. S. Hatsukami, J.-N. Hwang, and C. Yuan. "Automated artery localization and vessel wall segmentation of magnetic resonance vessel wall images using tracklet refinement and polar conversion". In: *arXiv preprint arXiv:1909.02087* (2019).

[46]   C. Zhu, X. Wang, Z. Teng, S. Chen, X. Huang, M. Xia, L. Mao, and C. Bai. "Cascaded residual U-net for fully automatic segmentation of 3D carotid artery in high-resolution multi-contrast MR images". In: *Physics in Medicine & Biology* 66.4 (2021), p. 045033.

[47]   R. El Jurdi, C. Petitjean, P. Honeine, V. Cheplygina, and F. Abdallah. "High-level prior-based loss functions for medical image segmentation: A survey". In: *Computer Vision and Image Understanding* 210 (2021), p. 103248.

[48]   Y. Xue, H. Tang, Z. Qiao, G. Gong, Y. Yin, Z. Qian, C. Huang, W. Fan, and X. Huang. "Shape-aware organ segmentation by predicting signed distance maps". In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 34. 07. 2020, pp. 12565–12572.

[49]   F. Navarro, S. Shit, I. Ezhov, J. Paetzold, A. Gafita, J. C. Peeken, S. E. Combs, and B. H. Menze. "Shape-aware complementary-task learning for multi-organ segmentation". In: *International Workshop on Machine Learning in Medical Imaging*. Springer. 2019, pp. 620–627.

[50]   J. R. Clough, N. Byrne, I. Oksuz, V. A. Zimmer, J. A. Schnabel, and A. P. King. "A topological loss function for deep-learning based image segmentation using persistent homology". In: *arXiv preprint arXiv:1910.01877* (2019).

[51]   J. E. Van Engelen and H. H. Hoos. "A survey on semi-supervised learning". In: *Machine Learning* 109.2 (2020), pp. 373–440.

[52]   X. Li, L. Yu, H. Chen, C.-W. Fu, L. Xing, and P.-A. Heng. "Transformation-consistent self-ensembling model for semisupervised medical image segmentation". In: *IEEE Transactions on Neural Networks and Learning Systems* 32.2 (2020), pp. 523–534.

[53]   X. Luo, J. Chen, T. Song, and G. Wang. "Semi-supervised medical image segmentation through dual-task consistency". In: *arXiv preprint arXiv:2009.04448* (2020).

[54]   S. Li, C. Zhang, and X. He. "Shape-aware semi-supervised 3d semantic segmentation for medical images". In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2020, pp. 552–561.

[55]   Y. Zhang and J. Zhang. "Dual-Task Mutual Learning for Semi-Supervised Medical Image Segmentation". In: *arXiv preprint arXiv:2103.04708* (2021).

[56]   A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao. "Yolov4: Optimal speed and accuracy of object detection". In: *arXiv preprint arXiv:2004.10934* (2020).

[57]   L. Chen, H. Zhao, H. Jiang, N. Balu, D. B. Geleri, B. Chu, H. Watase, X. Zhao, R. Li, J. Xu, et al. "Domain adaptive and fully automated carotid artery atherosclerotic lesion detection using an artificial intelligence approach (LATTE) on 3D MRI". In: *Magnetic Resonance in Medicine* 86.3 (2021), pp. 1662–1673.

[58]   W. Kerwin, D. Xu, F. Liu, T. Saam, H. Underhill, N. Takaya, B. Chu, T. Hatsukami, and C. Yuan. "Magnetic resonance imaging of carotid atherosclerosis: plaque analysis". In: *Topics in Magnetic Resonance Imaging* 18.5 (2007), pp. 371–378.

[59]   Z. Cao, T. Simon, S.-E. Wei, and Y. Sheikh. "Realtime multi-person 2d pose estimation using part affinity fields". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017, pp. 7291–7299.

[60]   H. Peng, S. Guo, and X. Zuo. "A Vehicle Detection Method Based on YOLOV4 Model". In: *2021 2nd International Conference on Artificial Intelligence and Information Systems*. 2021, pp. 1–4.

[61]   Z. Yang and R. Nevatia. "A multi-scale cascade fully convolutional network face detector". In: *2016 23rd International Conference on Pattern Recognition (ICPR)*. IEEE. 2016, pp. 633–638.

[62] P. F. Jaeger, S. A. Kohl, S. Bickelhaupt, F. Isensee, T. A. Kuder, H.-P. Schlemmer, and K. H. Maier-Hein. "Retina U-Net: Embarrassingly simple exploitation of segmentation supervision for medical object detection". In: *Machine Learning for Health Workshop*. PMLR. 2020, pp. 171–183.

[63] S. Albahli, N. Nida, A. Irtaza, M. H. Yousaf, and M. T. Mahmood. "Melanoma lesion detection and segmentation using YOLOv4-DarkNet and active contour". In: *IEEE Access* 8 (2020), pp. 198403–198414.

[64] C.-M. Kim, K. Chung, and R. C. Park. "Anomaly detection model of mammography using YOLOv4-based histogram". In: *Personal and Ubiquitous Computing* (2021), pp. 1–12.

[65] R. Girshick, J. Donahue, T. Darrell, and J. Malik. "Rich feature hierarchies for accurate object detection and semantic segmentation". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2014, pp. 580–587.

[66] S. Ren, K. He, R. Girshick, and J. Sun. "Faster r-cnn: Towards real-time object detection with region proposal networks". In: *Advances in neural information processing systems* 28 (2015), pp. 91–99.

[67] K. He, G. Gkioxari, P. Dollár, and R. Girshick. "Mask r-cnn". In: *Proceedings of the IEEE international conference on computer vision*. 2017, pp. 2961–2969.

[68] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi. "You only look once: Unified, real-time object detection". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 779–788.

[69] P. Mahto, P. Garg, P. Seth, and J. Panda. "Refining Yolov4 for vehicle detection". In: *International Journal of Advanced Research in Engineering and Technology (IJARET)* 11.5 (2020).

[70] Z. Dong, H. Chang, X. Pu, P. Luo, J. Weng, and Z. Liu. "Aircraft segmentation algorithm based on Unet and improved Yolov4". In: *2021 International Conference on Intelligent Transportation, Big Data & Smart City (ICITBS)*. IEEE. 2021, pp. 614–617.

[71] Z. Zheng, P. Wang, W. Liu, J. Li, R. Ye, and D. Ren. "Distance-IoU loss: Faster and better learning for bounding box regression". In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 34. 07. 2020, pp. 12993–13000.

[72] T. Lei, R. Wang, Y. Wan, B. Zhang, H. Meng, and A. K. Nandi. "Medical image segmentation using deep learning: a survey". In: *arXiv preprint arXiv:2009.13120* (2020).

[73] F. Milletari, N. Navab, and S.-A. Ahmadi. "V-net: Fully convolutional neural networks for volumetric medical image segmentation". In: *2016 fourth international conference on 3D vision (3DV)*. IEEE. 2016, pp. 565–571.

[74] S. Sharma and S. Sharma. "Activation functions in neural networks". In: *Towards Data Science* 6.12 (2017), pp. 310–316.

[75] R. Yamashita, M. Nishio, R. K. G. Do, and K. Togashi. "Convolutional neural networks: an overview and application in radiology". In: *Insights into imaging* 9.4 (2018), pp. 611–629.

[76] C. Zhang, S. Bengio, M. Hardt, B. Recht, and O. Vinyals. "Understanding deep learning (still) requires rethinking generalization". In: *Communications of the ACM* 64.3 (2021), pp. 107–115.

[77] X. Yang, Z. Song, I. King, and Z. Xu. "A Survey on Deep Semi-supervised Learning". In: *arXiv preprint arXiv:2103.00550* (2021).

[78] Y. Zhang, T. Xiang, T. M. Hospedales, and H. Lu. "Deep mutual learning". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2018, pp. 4320–4328.

[79] O. Chapelle, B. Scholkopf, and A. Zien. "Semi-supervised learning. 2006". In: *Cambridge, Massachusettes: The MIT Press View Article* (2006).

[80] W. Bai, O. Oktay, M. Sinclair, H. Suzuki, M. Rajchl, G. Tarroni, B. Glocker, A. King, P. M. Matthews, and D. Rueckert. "Semi-supervised learning for network-based cardiac MR image segmentation". In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2017, pp. 253–260.

[81] Y. Zhang, L. Yang, J. Chen, M. Fredericksen, D. P. Hughes, and D. Z. Chen. "Deep adversarial networks for biomedical image segmentation utilizing unannotated images". In: *International conference on medical image computing and computer-assisted intervention*. Springer. 2017, pp. 408–416.

[82] Y. Ouali, C. Hudelot, and M. Tami. "An overview of deep semi-supervised learning". In: *arXiv preprint arXiv:2006.05278* (2020).

[83] Y. Ouali, C. Hudelot, and M. Tami. "Semi-supervised semantic segmentation with cross-consistency training". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2020, pp. 12674–12684.

[84] A. R. Zamir, A. Sax, N. Cheerla, R. Suri, Z. Cao, J. Malik, and L. J. Guibas. "Robust learning through cross-task consistency". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2020, pp. 11197–11206.

[85] G. Hinton, O. Vinyals, and J. Dean. "Distilling the knowledge in a neural network". In: *arXiv preprint arXiv:1503.02531* (2015).

[86] C. Li, C. Xu, C. Gui, and M. D. Fox. "Level set evolution without re-initialization: a new variational formulation". In: *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*. Vol. 1. IEEE. 2005, pp. 430–436.

[87] R. Tsai, S. Osher, et al. "Level set methods and their applications in image science". In: *Communications in Mathematical Sciences* 1.4 (2003), pp. 1–20.

[88] J. Luo, Z. Luo, L. Chen, L. Tong, and M. Y. Wang. "A semi-implicit level set method for structural shape and topology optimization". In: *Journal of Computational Physics* 227.11 (2008), pp. 5561–5581.

[89]   Y. Wang, X. Wei, F. Liu, J. Chen, Y. Zhou, W. Shen, E. K. Fishman, and A. L. Yuille. "Deep distance transform for tubular structure segmentation in ct scans". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2020, pp. 3833–3842.

[90]   J. Ma, Z. Wei, Y. Zhang, Y. Wang, R. Lv, C. Zhu, C. Gaoxiang, J. Liu, C. Peng, L. Wang, et al. "How distance transform maps boost segmentation CNNs: an empirical study". In: *Medical Imaging with Deep Learning*. PMLR. 2020, pp. 479–492.

[91]   B. Bischke, P. Helber, J. Folz, D. Borth, and A. Dengel. "Multi-task learning for segmentation of building footprints with deep neural networks". In: *2019 IEEE International Conference on Image Processing (ICIP)*. IEEE. 2019, pp. 1480–1484.

[92]   H. Edelsbrunner, D. Letscher, and A. Zomorodian. "Topological persistence and simplification". In: *Proceedings 41st annual symposium on foundations of computer science*. IEEE. 2000, pp. 454–463.

[93]   M. Li, K. Duncan, C. N. Topp, and D. H. Chitwood. *Persistent homology and the branching topologies of plants*. 2017.

[94]   A. Adcock, E. Carlsson, and G. Carlsson. "The ring of algebraic functions on persistence bar codes". In: *arXiv preprint arXiv:1304.0530* (2013).

[95]   J. Redmon and A. Farhadi. "Yolov3: An incremental improvement". In: *arXiv preprint arXiv:1804.02767* (2018).

[96]   D. M. Mish. "A self regularized non-monotonic neural activation function.(2019)". In: *arXiv preprint arXiv:1908.08681* (1908).

[97]   C.-Y. Wang, H.-Y. M. Liao, Y.-H. Wu, P.-Y. Chen, J.-W. Hsieh, and I.-H. Yeh. "CSPNet: A new backbone that can enhance learning capability of CNN". In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*. 2020, pp. 390–391.

[98]   S. Liu, L. Qi, H. Qin, J. Shi, and J. Jia. "Path aggregation network for instance segmentation". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018, pp. 8759–8768.

[99]   K. He, X. Zhang, S. Ren, and J. Sun. "Spatial pyramid pooling in deep convolutional networks for visual recognition". In: *IEEE transactions on pattern analysis and machine intelligence* 37.9 (2015), pp. 1904–1916.

[100] *PyTorch*. https://pytorch.org/. Accessed: 2021-08-10.

[101] *NVIDIA Tesla T4*. https://www.nvidia.com/es-es/data-center/tesla-t4/. Accessed: 2021-08-10.

[102] R. Brüel-Gabrielsson, B. J. Nelson, A. Dwaraknath, P. Skraba, L. J. Guibas, and G. Carlsson. "A topology layer for machine learning". In: *arXiv preprint arXiv:1905.12200* (2019).

[103] L. R. Dice. "Measures of the amount of ecologic association between species". In: *Ecology* 26.3 (1945), pp. 297–302.

[104] P. Jaccard. "Nouvelles recherches sur la distribution florale". In: *Bull. Soc. Vaud. Sci. Nat.* 44 (1908), pp. 223–270.

[105] D. P. Huttenlocher, G. A. Klanderman, and W. J. Rucklidge. "Comparing images using the Hausdorff distance". In: *IEEE Transactions on pattern analysis and machine intelligence* 15.9 (1993), pp. 850–863.

[106] Y. R. Limbu, G. Gurung, R. Malla, R. Rajbhandari, and S. R. Regmi. "Assessment of carotid artery dimensions by ultrasound in non-smoker healthy adults of both sexes." In: *Nepal Medical College journal: NMCJ* 8.3 (2006), pp. 200–203.

[107] C. Zhu, Z. Teng, U. Sadat, V. E. Young, M. J. Graves, Z. Y. Li, and J. H. Gillard. "Normalized Wall Index Specific and MRI-Based Stress Analysis of Atherosclerotic Carotid Plaques–A Study Comparing Acutely Symptomatic and Asymptomatic Patients–". In: *Circulation Journal* (2010), pp. 1009030883–1009030883.

[108] D. Coeurjolly, S. Miguet, and L. Tougne. "Discrete curvature based on osculating circle estimation". In: *International Workshop on Visual Form*. Springer. 2001, pp. 303–312.

[109] *Underground Mathematics- Curvature*. `https://undergroundmathematics.org/glossary/curvature`. Accessed: 2021-08-10.