QUANTIFYING ATHEROSCLEROSIS IN VASCULATURE USING ULTRASOUND IMAGING

A thesis submitted to Cardi↵ University in
candidature for the degree of Doctor of Philosophy

CARL AZZOPARDI

School of Engineering
Cardi↵ University
December 2020

Abstract

Cerebrovascular disease accounts for approximately 30% of the global burden associated with cardiovascular diseases [1]. According to the World Stroke Organisation, there are approximately 13.7 million new stroke cases annually, and just under six million people will die from stroke each year [2]. The underlying cause of this disease is atherosclerosis – a vascular pathology which is characterised by thickening and hardening of blood vessel walls. When fatty substances such as cholesterol accumulate on the inner linings of an artery, they cause a progressive narrowing of the lumen referred to as a stenosis.

Localisation and grading of the severity of a stenosis, is important for practitioners to assess the risk of rupture which leads to stroke. Ultrasound imaging is popular for this purpose. It is low cost, non-invasive, and permits a quick assessment of vessel geometry and stenosis by measuring the intima media thickness. Research is showing that 3D monitoring of plaque progression may provide a better indication of sites which are at risk of rupture. Various metrics have been proposed. From these, the quantification of plaques by measuring vessel wall volume (VWV) using the segmented media-adventitia boundaries (MAB) and lumen-intima boundaries (LIB) has been shown to be sensitive to temporal changes in carotid plaque burden. Thus, methods to segment these boundaries are required to help generate VWV measurements with high accuracy, less user interaction and increased robustness to variability in di↵erent user acquisition protocols.

This work proposes three novel methods to address these requirements, to ultimately produce a highly accurate, fully automated segmentation algorithm which works on intensity-invariant data. The first method proposed was that of generating a novel, intensity-invariant representation of ultrasound data by creating phase-congruency maps from raw unprocessed radio-frequency ultrasound information. Experiments carried out showed that this representation retained the necessary anatomical structural information to facilitate segmentation, while concurrently being invariant to changes in amplitude from the user. The second method proposed was the novel application of Deep Convolutional Networks (DCN) to carotid ultrasound images to achieve fully automatic delineation of the MAB boundaries, in addition to the use of a novel fusion of amplitude and phase congruency data as an image source. Experiments carried out showed that the DCN produces highly accurate and automated results, and that the fusion of amplitude and phase yield superior results to either one alone. The third method proposed was a new geometrically constrained objective function for the network's Stochastic Gradient Descent optimisation, thus tuning it to the segmentation problem at hand, while also developing the network further to concurrently delineate both the MAB and LIB to produce vessel wall contours. Experiments carried out here also show that the novel geometric constraints improve the segmentation results on both MAB and LIB contours.

In conclusion, the presented work provides significant novel contributions to field of Carotid Ultrasound segmentation, and with future work, this could lead to implementations which facilitate plaque progression analysis for the end-user.

To my wife, Roberta.

Acknowledgements

I would like to firstly thank my supervisor, Dr. Yulia Hicks, for guiding me throughout this project, for assisting me in developing my ideas, and most importantly, for steadying my resolve when I needed encouragement.

I would also like to express my gratitude towards Professor Kenneth Camilleri, who always willingly o↵ered his time to help me unfold the complexities of technical problems with a clear mind and with the correct approach.

I would like to thank my parents, my brother and my sister, who were always there for me to o↵er continuous encouragement.

And finally I would like to express my deepest, heartfelt gratitude towards my wife, Roberta, who was by my side during the highs, and patiently endured the lows. Her constant support and shelter from life's everyday problems, made the completion of this dissertation possible.

The research work disclosed in this publication is partially funded by the Malta Government Scholarship Scheme grant.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

ABI Ankle Brachial Index

AE Auto Encoders

ANN Artificial Neural Network

B-M Brightness Mode

CNN Convolutional Neural Networks

CVD Cardiovascular Disease

DCN Deep Convolutional Networks

DBN Deep Belief Networks

DSA Digital Subtraction Angiography

DUS Duplex Ultrasound

FIR Finite Impulse Response

FCN Fully Convolutional Networks

GA Genetic Algorithms

GAN Generative Adversarial Networks

GPU Graphical Processing Unit

LIST OF TABLES

HT Hough Transform

ICA Internal Carotid Artery

IMT Intima-Media Thickness

IQ In-phase Quadrature

LDL Low Density Lipoproteins

LIB Lumen-Intima Boundary

LI Lumen Intima

PZT Lead Zirconate Titanate

LSM Level Set Methods

MA Media-Adventitia

MAB Media-Adventitia Boundary

MAP Maximum A-Posteriori Estimation

MCC Maximum Correlation Coe cients

MHD Modified Hausdor↵ Distance

MLP Multi-Layered Perceptron

MRA Magnetic Resonance Angiography

MRF Markov Random Field

PCM Phase Congruency Map

PDM Phase Di↵erence Matrix

RF Radio Frequency

RBM Restricted Boltzmann Machines

RMS Root Mean Square

RNN Recurrent Neural Networks

SNR Signal-to-Noise Ratio

SAE Stacked Auto Encoders

SDK Software Development Kit

SGD Stochastic Gradient Descent

TIA Transient Ischaemic Attack

TGC Time Gain Compensation

TPV Total Plaque Volume

VWV Vessel Wall Volume

1

Introduction

1.1 Cardiovascular Disease - an overview

According to statistics presented by the World Health Organisation (WHO), cardiovascular diseases are the leading cause of death around the world, taking the lives of approximately 17.9 million people each year, and accounting for 31% of all deaths [3]. From all deaths associated with cardiovascular diseases, 85% are due to heart attacks and strokes [3]. The most common cause behind all this is a pervasive condition called *Atherosclerosis*, which lies silent and asymptomatic in the large majority of people for many years, before it finally surfaces with debilitating consequences [4; 5].

Cardiovascular diseases (CVDs) include within them diseases associated with the heart, vascular diseases of the brain, as well as diseases of blood vessels [1; 6]. The underlying conditions of CVDs may be di↵erent, with the most common being atherosclerosis. This may lead to ischaemic heart and coronary artery disease, such as heart attacks; cerebrovascular diseases like stroke; and diseases of the aorta and arteries, including those related to peripheral

vasculature and hypertension [1; 6]. Other CVDs may be caused by congenital or rheumatic conditions, cardiomyopathies, as well as cardiac arrhythmias.

Atherosclerosis is particularly common because it is mediated by several risk factors, including behaviour (tobacco use, physical inactivity), obesity and advancing age [1; 7]. It is a complex vascular pathology, developing over many years and characterised by the thickening and hardening of blood vessel walls [8]. When plaque deposits itself in the inner lining of vessels, it causes their inner surface to become irregular and narrow, and in turn makes it harder for blood to flow through [1; 6]. The vessels themselves also become less pliable, and if the plaque ruptures, an embolus may be carried by the blood further upstream, causing an occlusion in a narrower part of the vessel.

Rheumatic heart disease is caused by rheumatic fever, which leads to damage of the heart muscle and heart valves [1]. Other structural defects in the heart, such as holes in the septum, or abnormal valves and chambers, may be present at birth as congenital defects[1; 7]. Other disorders such as cardiomyopathies, or disorders of the electrical conduction system of the heart (cardiac arrhythmias) are also possible, although somewhat less common than heart attacks and strokes[1; 7].

1.2 Cerebrovascular Disease

Cerebrovascular disease accounts for approximately 30% of the global burden associated with cardiovascular diseases [1]. According to the World Stroke Organisation, there are approximately 13.7 million new stroke cases annually, and just under six million people will die from stroke each year [2].

The pathophysiology behind stroke may be either ischaemic or haemorrhagic in nature. Ischaemic events occur due to the formation of thrombi or emboli which occlude a vessel in the brain. Thrombotic cerebral infarctions result from atherosclerotic occlusion of cerebral and cervical arteries, with ischaemia taking place in part or all of the region of occluded artery. This may occur due to either the obstruction at the site of the main atherosclerotic plaque rupture, or due to a thromboembolism forming from this site and travelling to other distal cerebral arteries [9; 1]. A common example of the latter is due to atherosclerotic plaque forming and rupturing within the common carotid artery.

Embolic cerebral infarctions are due to embolism of a clot, plaque contents, or other particles in the cerebral arteries, originating from other parts of the arterial system. These may originate from atherosclerotic lesions in di↵erent vessels, cardiac lesions at the site of the heart or valves, or else due to rhythm disturbances such as atrial fibrillation, which may lead to stasis of blood and clotting within the heart [9; 10]. If the symptoms of an ischaemic event are short (less than 1 hour in duration), and show no evidence of infarction on imaging, then these are termed *Transient Ischaemic Events* (TIA) [11].

Intracerebral haemorrhages are commonly due to arteriolar hypertensive disease [9]. Less commonly, they may also occur due to coagulation disorders and vascular malformation within the brain, or due to the diet. Subarachnoid haemorrhages on the other hand mainly occur due to ruptures of aneurysms at the bifurcations of large arteries within the inferior surface of the brain [9; 1].

In countries which are demographically developed, the average age at which stroke occurs is approximately 73 years [9]. This reflects the older age structure exhibited by these countries. A patient who has experienced a stroke is at the highest risk of death in the first weeks after the event. Indeed, between 20% and 50% of patients die within the first month of the event, depending on the type of stroke, severity and presence of any co-morbidities [9]. Surviving patients may be left with no disability, but may also experience severe disability [9]. Risk factors which increase the incidence of stroke include inherent biological traits (age, sex, high blood pressure etc), behaviours such as smoking, alcohol consumption, poor diet and physical inactivity, as well as social characteristics such as education, social class and ethnicity [9; 1; 7].

1.3 Atherosclerosis

1.3.1 Anatomy and Physiology

The vascular system is an organ system which consists of arteries, veins and capillaries [12]. In combination with the lymphatic system, it is primarily responsible for circulating blood and other fluids around the body, in order to transport nutrients to other cells and to drain away tissue waste matter. Together, these mechanisms help to ensure that the body remains nourished, and that homeostasis is maintained [12].

Blood moves through the circulatory system through two main pathways of circulation, these being the venous and arterial routes [12]. In brief, the former route returns blood to the lungs for it to be oxygenated and for carbon dioxide to be removed; the latter then distributes the oxygenated blood throughout the rest of the system, in order for oxygen to be transported to cells [12].

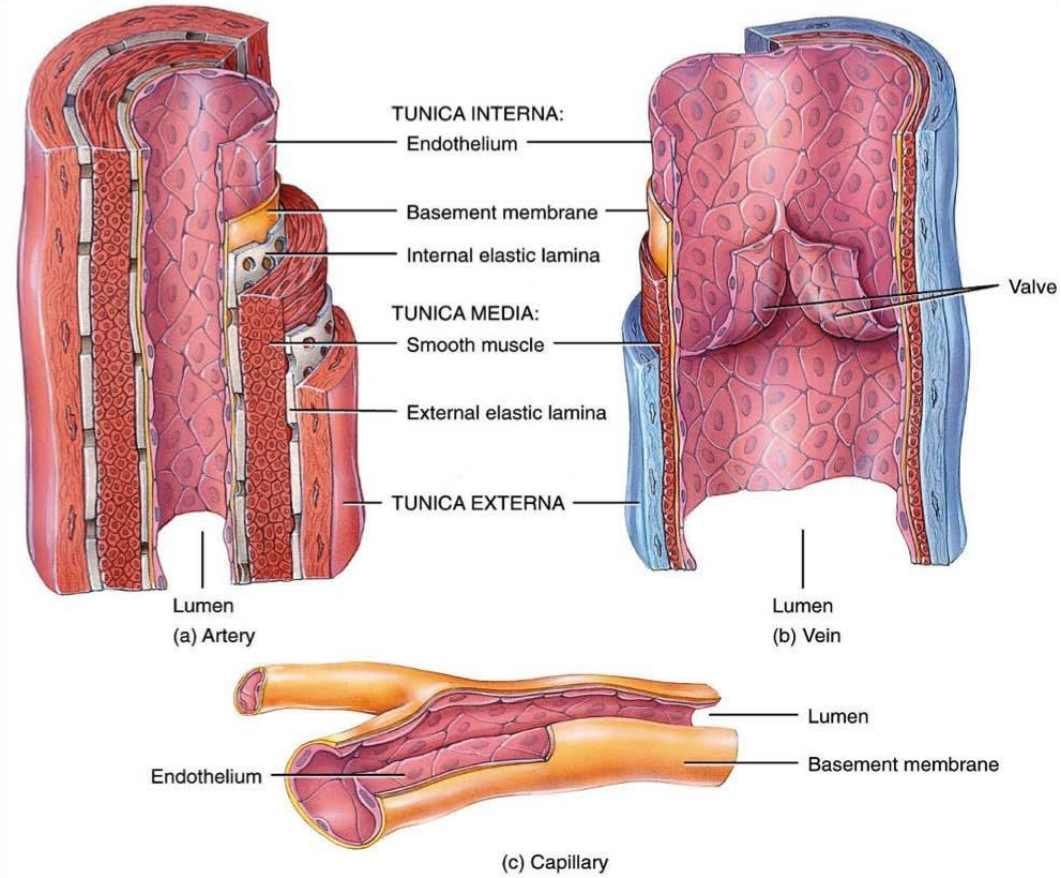

Figure 1.1: A cross-sectional cutout of the (a) arterial layers, (b) venous layers and (c) capillaries. Reproduced with permission from [13].

Capillaries themselves are very small blood vessels, which are only one cell in thickness, and which form the junction between arteries and veins [12].

The arteries are relatively large vessels made of three di↵erent layers. The innermost layer is called the tunica intima, and this is lined with simple squamous epithelial cells to form the endothelium. The latter provides for a very smooth surface in the arterial wall to permit normal flow of blood and to prevent abnormal clotting [12]. The middle layer is called the tunica media, and is made of smooth muscle cells and elastic connective tissue. Its role is to

maintain normal blood pressure, particularly the diastolic pressure when the heart is relaxed. The third and outer layer is called the tunica adventitia, and is made of fibrous connective tissue [12]. This layer is strong, and helps to prevent the rupture of the larger arteries that carry blood under higher pressures.

A graphical representation of these layers is given in Figure 2.1a.

Veins are composed of the same three tissue layers as those present in arteries. There are however some small di↵erences in structure [12]. Firstly, while the inner layer of veins is also made of smooth endothelium, the lining is folded at regular intervals to form valves. This is also shown in Figure 2.1b. These valves prevent the back-flow of blood, and are found in relatively large quantities in the veins of the legs, where the blood must return to the heart against the force of gravity [12]. The middle and outer layers of veins are thinner than those in arteries, because veins do not regulate blood pressure, and neither do they have to sustain high blood pressures like those found in arteries [12].

Throughout the circulatory network, both arteries and veins branch out to smaller vessels called arterioles and venules respectively. The latter two vessels are then joined together through capillaries, which are in e↵ect just an extension of the inner lining of arteries and veins, as shown in Figure 2.1c [12]. Since capillary walls are just one cell in thickness, this gives them a permeability, which allows for exchange of nutrients and oxygen between themselves and surrounding tissue. The blood flow through the capillary networks is regulated through smooth muscle cells called pre-capillary sphincters [12], which constrict or dilate accordingly, depending on whether the surrounding tissue needs more blood flow or otherwise [12].

During the process of exchange between capillaries and surrounding tissue, a

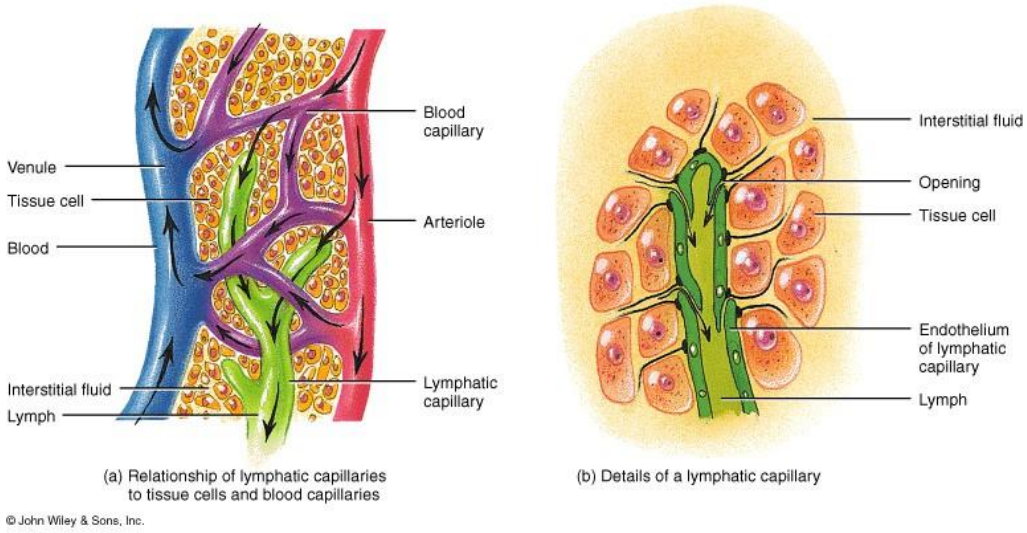

Figure 1.2: The lymphatic system drains away tissue waste from the extracellular space. Reproduced with permission from [13].

certain amount of tissue fluid is formed – only part of which is re-absorbed into the blood and back into the capillaries [12]. Thus if this process were to continue, the total amount of blood present would be depleted. The excess tissue fluid, therefore, is siphoned back through lymph capillaries, whereby it is eventually returned to the blood again to be recycled as plasma [12]. This is shown in Figure 2.2. The lymphatic system is composed of lymph vessels, lymph nodes and nodules, the spleen and the thymus gland. It is generally responsible for returning the said tissue fluid to the blood and for protecting the body against foreign materials [12].

1.3.2 Natural History of Atherosclerosis

Atherosclerosis is a vascular pathology characterised by the thickening and hardening of walls of large and medium sized arteries, particularly in societies

where food, which is rich in cholesterol, is abundant and cheap [8], [14]. The condition commences in childhood. On the assumption that there are no external factors to accelerate its development, it will accumulate slowly until it is relatively widespread by old age [14]. There are however di↵erent genetic and environmental factors, which may a↵ect the progression of atherosclerosis, as well as the shape and size of vessel walls.

The vessel wall is continually subjected to biochemical and biomechanical stressors that a↵ect functional and adaptive responses [15]. If blood flow acutely increases for instance, the increases wall shear stress is sensed by the endothelial cells, causing relaxants to be released [16]. These lead to vessel dilation, which counteracts the increase in wall shear stress sensed initially [15]. If this e↵ect were to be sensed long-term, the response provokes diameter adaptation. Wall thickness changes (referred to as vascular remodelling) are seen in response to a multitude of biological processes, including changes in blood pressure, ageing, inflammation, oxidative stress and lipid accumulation [15; 17; 18].

In vasculature, two types of remodelling may be clearly seen: inward and outward remodelling. Both types may respectively exhibit hypertrophy (thickening) or hypotrophy (thinning) of the vessel walls[15]. This is shown in Figure 1.3. Aneurysm formation, for instance, is characterised by an increase in the diameter of the vessel, with a thinning of the vessel wall. This is termed outward hypothrophic remodelling [15; 19]. In early atherosclerosis, the initial remodelling tends to be an outward remodelling which preserves or increases lumen size [20]. However, although protective in the long-term, a matrix degradation takes place in this process, and this may predispose the atherosclerotic plaques to rupture. In advanced atherosclerosis, lesions may

cause inward remodelling, causing the vessels to shrink rather than enlarge, and thus exacerbating the stenosis [20].

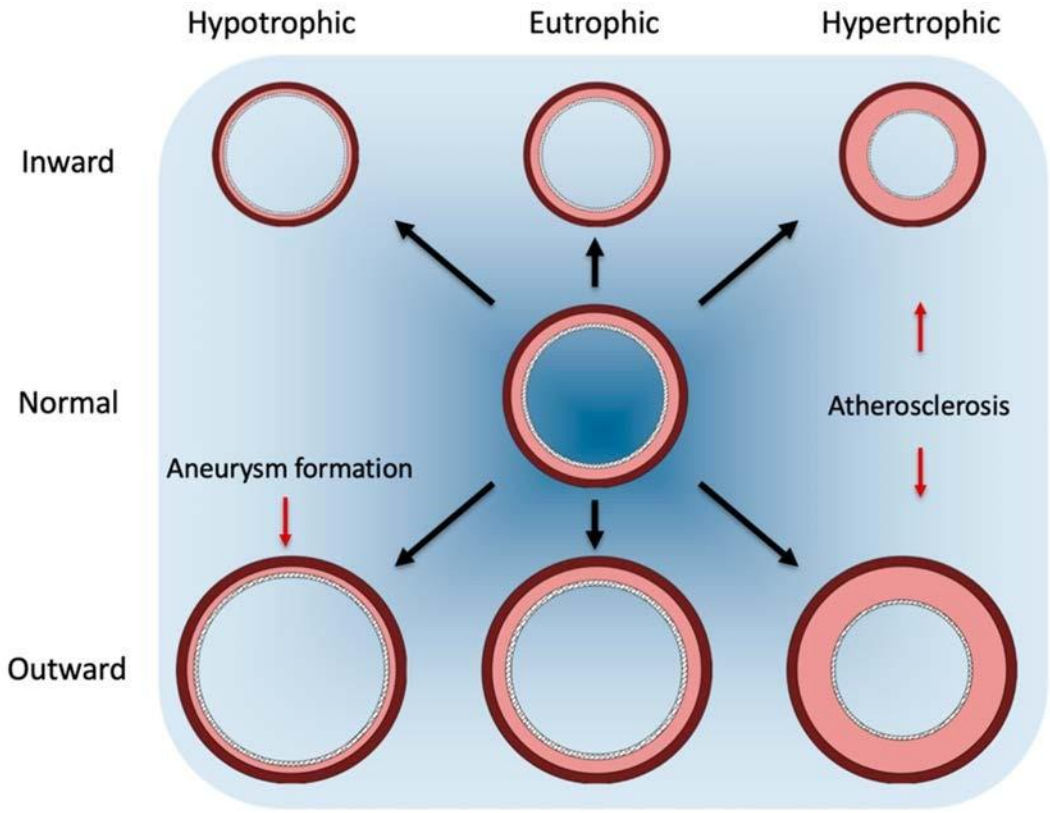

Figure 1.3: Di↵erent types of vascular remodelling. Reproduced with permission from [15].

Atherosclerosis is considered an inflammatory disease, commencing with endothelial stress and subsequent monocyte infiltration. When fatty substances such as cholesterol, triglycerides, or cellular waste products such as calcium and fibrin, start to accumulate on the inner linings of an artery, they cause a progressive narrowing of the lumen and consequently restrict the free flow of blood [17]. This build-up is referred to as plaque, and may eventually lead to the total or partial occlusion of a blood vessel. Atherosclerosis thus

leads to vascular insu ciency in the limbs, poor circulation in the renal system, or even rupture of large arteries. The formation of intravascular clots also potentially lead to life threatening conditions of the heart and brain. [14].

Atherosclerotic plaque develops from lesions caused by deposits of low density lipoproteins (LDL) in the arterial intima. Under normal circumstances, a healthy intimal surface will already contain deposits of low density lipoproteins, but these would be present at much lower concentrations [21]. The reason for the eventual increase in concentration of such lipid deposits is thought to arise from the absence of lymphatic vessels in the arterial intima [22; 23]. As described earlier, lymph vessels act as sumps, which drain away excess macromolecules and tissue waste within the extracellular space, as shown in Figure 2.3. However, since lymph vessels operate at low pressure, they would presumably fail to function correctly or collapse altogether when subjected to the high hydrostatic pressures present within arteries [22]. Thus, without e↵ective siphonage of macromolecules such as LDL, the stage is already set for an accumulation of these lipids over time. A secondary e↵ect, which exacerbates the situation, is the mechanical forces that certain arterial regions are exposed to. At points where arteries branch, there is a tendency for the endothelium, which lines the intima, to be pulled along or deformed by flowing blood. This results in shear stress of the endothelium, and is where the lipids generally tend to accumulate [14]. A greater length of exposure of the lumen surface to

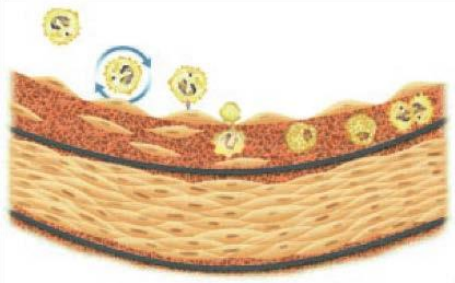

Figure 1.4: Recruitment of macrophages to engulf LDL. Reproduced with permission from [17].

circulating atherogenic particles also gives rise to a greater influx of lipids into the intima

[21].

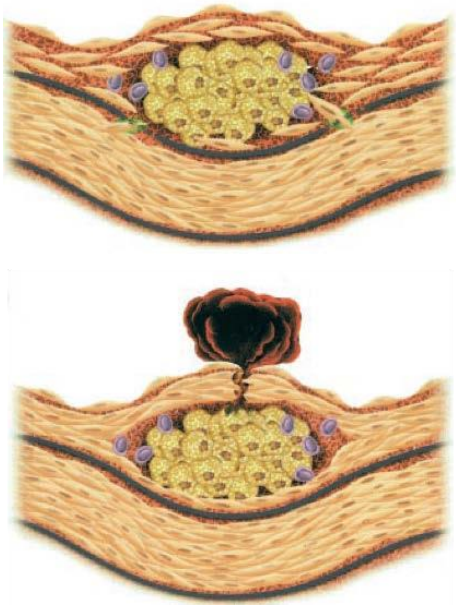

Figure 1.5: Accumulation of plaque individual LDL particles, leading to (top) and rupture (bottom). Reproduced with permission from [17].

The LDL particles traverse the arterial intima very slowly across the endothelial cells, while other LDL particles would slowly be crossing the endothelial barrier in the opposite direction. The slowness of this two-way migration causes the lipid particles to reside within the intima for weeks or months, and this maximises the opportunity for enzymatic and oxidative processes to take place. Such processes disrupt the integrity of the their chemical alteration and to two main consequences [22].

The first is that the enzymatic and oxidative processes break down the resistance of the lipid particles to coalescence, causing them to fuse together [22]. Taking this in conjunction with the progressive accumulation of LDL in the arterial intima will often lead to the formation of large lipid droplets, which then eventually lead to even larger lipid cores or pools [22]. These cores are less mobile than their constituents, and thus more di cult to remove from the arterial wall [22]. The second e↵ect caused by the chemical alteration of the LDL particles is that they are taken up by inflammatory scavenger cells called macrophages [17]. Part of the natural response of the body to foreign objects

is to cause inflammation to combat it. The macrophages are recruited into the intima, where they engulf the LDL deposits, convert them into cholesterolesters and become enlarged foam cells [17]. The foam cells then express the cholesterol-esters back into the intima, where they combine with high-density lipoproteins and traverse the endothelial barrier back into the blood stream [17]. The latter process is however slow, and the presence of cholesterol-esters in the intima slowly starts to cause the death of surrounding smooth muscle cells [17].

This process of smooth muscle cell death and accumulation of cholesterol creates a core of atherosclerotic plaque deep within the intima [17]. As a further inflammatory response therefore, fibrogenic mediators are released to promote the replication of smooth muscle cells at the surface, in between the atherosclerotic core and the endothelial surface [17]. Thus, a strong fibrous cap is formed, which keeps the atherosclerotic plaque contained and reduces the risk of it rupturing, as shown in Figure 2.4 (top) [17]. The downside to this is that excessive fibroproliferation of muscle cells will cause the fibrous cap to be too thick, thus choking o↵ the flow of blood within the artery [17]. This is the primary mechanism of how a stenosis is formed, leading to typical ischaemic side-e↵ects. Additionally, the cholesterol within the atherosclerotic core typically eats away at the fibrous cap from within, rendering it weak [17]. Thus a point may be reached where the fibrous cap ruptures and the contents spill into the blood stream, as shown in Figure 2.4 (bottom). This may result in the formation of a thrombus, or the contents of the plaque rupture themselves may form an embolism. Either of these may result in a transient ischaemic attack, or more seriously, in a stroke or heart attack.

1.3.3 Clinical Manifestations of Atherosclerosis

Atherosclerosis might not present any symptoms or sign of spread until it considerably narrows or totally blocks an artery. In many cases, people will be completely asymptomatic and unaware of having the disease until it is too late and they experience complications or a medical emergency, such as a stroke or a heart attack. Some signs and symptoms may, however, be present in some people, depending on which of the arteries are most severely a icted [24]. The carotid artery, for instance, supplies oxygen-rich blood to the brain. If accumulated plaque in an artery causes excessive narrowing, or a ruptured fibrous cap causes a thrombus or embolism, which partially or fully blocks the flow of blood, the ensuing event will be a transient ischaemic attack or, more severely, an ischaemic stroke, which deprives brain tissue of oxygen and causes death of brain cells within minutes.[24]. Amongst other symptoms, this would present as paralysis or numbness in the face, arm or leg - very commonly on just one side of the body; sudden weakness, loss of balance and confusion; trouble speaking and slurring of words or loss of consciousness.

Similarly, when coronary arteries supplying blood to the heart become occluded, and depending on the patency of vessels to the vascular territory, this may result in angina pectoris. This presents initially as chest pain, which would feel like pressure or squeezing of the chest [24], and comes on often during exertion and disappears again with rest. If plaque forms in the heart's smallest arteries, then the symptoms might include sleep problems, fatigue and lack of energy. If an atherosclerotic lesion results in a complete occlusion of the coronary artery, the myocardium supplied by that artery completely loses its blood supply and dies, causing a myocardial infarction [14] If the

occlusion is only partial, then the tissue is partially perfused, and may recover provided that the blockage is removed.

1.3.4 Risk Factors

A variety of environmental risk factors and genetic predispositions may accelerate the progression of atherosclerosis in an individual. Naturally, if one avoids conditions which are avoidable, or treats conditions for which a treatment is available, then such accelerating factors are minimised and the incidence of complications like stroke, infarctions etc. are reduced. Table 2.1 provides a non-exhaustive list of example conditions which accelerate the progression of atherosclerosis, as well as the mechanisms behind the action.

The presence of oestrogen for instance, is known to facilitate cholesterol removal by the liver, and this is evidenced by a slower progression of atherosclerosis in premenopausal women than in men. Additionally, there is epidemiologic evidence that shows that the cardiovascular system of postmenopausal women is protected when these are placed on oestrogen replacement therapy [14]. Cigarette smoking is one of the main predominant factors contributing to the progression of atherosclerosis in both men and women. Individuals who smoke 20 cigarettes a day increase their likelihood of death from ischaemic heart disease by approximately 70% when compared to non-smokers. Smoking is believed to damage the endothelial walls of vessels due to carbon monoxide induced hypoxia that is caused to the cells. The cessation of smoking is therefore highly e↵ective in slowing the progress of atherosclerosis [14].

The presence of hypertension also establishes an increased risk for atherosclerosis. An elevated blood pressure in the vessels causes increased

shear stress on the vessel walls, and this is thought to increase the progression of atherosclerosis [14]. Aside from primary causes of hypertension, which are idiopathic in nature, there are several secondary causes for which hypertension becomes a Table 1.1: List of conditions that accelerate progression of atherosclerosis [14].

Condition	**Mechanism of Action**
Gender (Male)	Lack of Estrogen, which has e↵ect of lowering LDL
Family History of heart disease / stroke	Genetic predisposition
Cigarette Smoking	Likely injury to endothelial cells due to hypoxia induced by carbon-monoxide intake
Hypertension	Damage to endothelium as a result of shear stress caused by increased blood pressure
Diabetes mellitus	Impaired hepatic function when removing LDL, which in turn causes increased LDL binding to blood vessel walls
Hypothyroidism	Impaired formation of LDL receptors in the liver

consequent e↵ect. Some examples include: renovascular hypertension, which have been described earlier, Cushing syndrome; hyperthyroidism; obesity; use of drugs such as decongestants, appetite suppressants and antidepressants; as

well as abuse of toxic substances such as cocaine, amphetamine and alchohol. Some secondary causes of hypertension may be simply avoided of course, but modern methods of treatment are also available to regulate the blood pressure to within normal levels [14].

When considering diabetes, one finds that there is a two-fold increase in the incidence of heart attacks in diabetics when compared to non-diabetics. Similarly, increased incidence of thrombotic strokes and circulatory deficiency leading to gangrene in peripheral areas, such as legs, are common in diabetics [14]. Treatment to control both blood pressure and glucose levels in patients su↵ering from diabetes is beneficial and e↵ective in reducing cardiovascular complications. Finally, in patients exhibiting hypothyroidism, the levels of plasma cholesterol and triglycerides levels increase. This is associated with a decreased formation of LDL receptors in the liver, and therefore with decreased lipoprotein lipase activity. The increased levels of cholesterol and triglycerides in turn accelerate the development of atherosclerosis [14; 25].

1.4 Clinical Diagnostics and Treament

1.4.1 NICE Pathways for managing Stroke/TIA

This section focuses on the diagnosis and treatment of stroke and transient ischaemic attacks, which are events which may typically occur to patients suffering from significant atherosclerosis in the carotid artery. The healthcare systems in di↵erent countries might follow di↵erent clinical pathways to diagnosing and treating occurrences of stroke and TIAs. In the United Kingdom, hospitals follow the pathways defined by the National Institute of Health and Care Excellence (NICE) [25]. If a person presents with sudden onset

of neurological symptoms, and stroke or TIA is suspected, a validated tool like FAST (Face Arm Speech Test) is used to carry out initial diagnosis [25].

In case a TIA is suspected, the patient is o↵ered medical treatment (such as Aspirin 300mg daily), followed by immediate referral to specialist TIA assessment within 24 hours of onset of symptoms [25]. The specialist assessment takes place in the form of urgent carotid ultrasound imaging in a TIA clinic, followed by possible brain imaging with MRI to determine the territory of ischaemia or to detect haemorrhage. During ultrasound carotid imaging, TIA patients presenting with carotid stenosis of between 50% to 99% according to North American Symptomatic Carotid Endarterectomy Trial (NASCET) criteria are o↵ered medical treatment and referred for urgent carotid endarterectomy [25]. This is a procedure whereby the carotid is opened and the plaque which is blocking the artery is removed [14]. In case endarterectomy is not possible, carotid artery stenting is considered [25]. In TIA patients who have symptomatic carotid stenosis of less than 50% according to NASCET criteria, or less than 70% according to European Carotid Surgery Trial (ECST) criteria, these are o↵ered medical treatment to control blood pressure and cholesterol as well as anti-platelet agents and lifestyle advice [25].

In case an acute stroke is suspected, the patient is referred immediately for brain imaging using non-enhanced CT within one hour. The patient is also provided blood sugar control and may be o↵ered supplemental oxygen therapy if blood saturation levels dip below 95% [25]. The brain imaging is used to determine if the patient has su↵ered an ischaemic stroke or a haemorrhagic stroke. In case of the former, the patient is treated with anti-platelet medication, together with either thrombolysis, thrombectomy, or

both. In thrombolysis, the patient is given an intra-venous administration of recombinant tissue plasminogen activator (tPA) which breaks up the clot [26]. In some cases, the tPA may be administered directly inside the brain close to where the stroke is happening, via a catheter inserted through an artery in the groin. For thrombectomy, the patient needs to have a confirmed occlusion demonstrated by CT Angiography or MR angriography. In this procedure, a mesh device is inserted into the artery in one's groin, moved up to the brain, and used to pull out clots in large arteries [25]. In case the stroke su↵ered was non-disabling, the patient is also referred for assessment of carotid stenosis. In a manner similar to that for TIAs, the NASCET / ECST criteria are used to determine the level of stenosis, and whether the patient is referred for medical or surgical treatment.

If following initial CT scans, the patient is diagnosed with a haemorrhagic stroke, they are o↵ered medical or surgical treatment, depending on history, and patient condition [25]. Typical priorities would also be to reverse e↵ects of anticoagulation medical treatment in patients who were taking this, in order to return clotting levels to normal as soon as possible. Additionally, the blood pressure is typically reduced rapidly, unless there are conditions for which this is contraindicated [25].

1.4.2 Assessment of Carotid Stenosis in Stroke Units

As mentioned in the preceding section, in case of non-disabling strokes or TIAs, patients are referred to specialist assessment, which includes carotid ultrasound imaging and grading of any stenosis according to criteria defined in the NASCET or ECST trials. In both cases, the trials sought to grade the degree of stenosis in order to stratify risk, however the methodologies used

for measuring stenosis severity were di↵erent [27; 28]. In the NASCET method, the diameter of the residual lumen in the stenotic segment was compared to the diameter of the normal Internal Carotid Artery (ICA), distal to the bulb. In the ECST method, the residual lumen in the stenotic segment was compared to an estimate of the diameter of the whole artery at the point of stenosis [28]. Both these methods may be visualised in Figure 1.6.

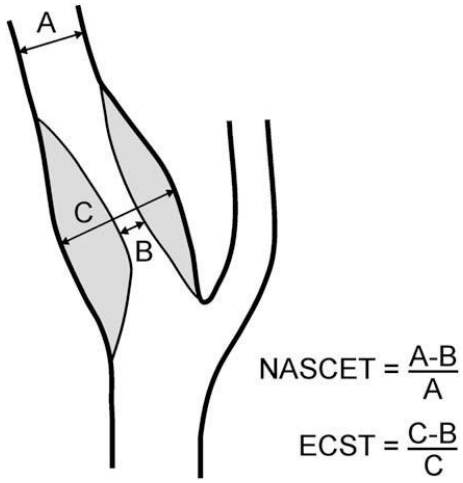

Figure 1.6: Measurement of stenosis using NASCET and ECST methods. Reproduced with permission from [28].

All the randomised patients under the NASCET / ECST trials were imaged using intra-arterial angiography [27; 28]. This practice has since been abandoned, due to the risk of angiography-related stroke [27]. Present day practice is to use Duplex ultrasound (DUS), due to its low cost and accessibility. B-Mode imaging and colour flow imaging are initially used to locate the stenosis, and DUS is then used to undertake Doppler flow velocity measurements, and determine stenosis based on the peak systolic velocity (PSV), end-diastolic velocity (EDV) and their ratios in the ICA and common carotid artery (CCA) [27]. The PSV and EDV measurements are taken in both the distal common carotid artery, and in the ICA at the site where the highest PSV is observed [28]. These four measurements allow the calculation of the recommended indices, shown here based on the NASCET criteria, in Table 1.2. A second operator repeats the examination to confirm the findings or to detect variation. [28].

Trials such as NASCET and ECST have clarified the role of carotid endarterectomy in the management of patients with carotid stenosis [29]. According to the NICE guidelines, patients presenting with a NASCET graded stenosis greater than 50% (or equivalent ECST graded stenosis greater than 70%) should be referred for urgent carotid endarterectomy. A study in 2006 by Walker and Nayor [29] however showed, following an audit of UK practice, that 43% of clinics did not know what criteria they were reporting against

Table 1.2: NASCET Diagnostic Criteria. Reproduced with permission from [28].

% Stenosis (NASCET)	**ICA PSV cm/s**	**PSV ratio** ICA_{PSV}/CCA_{PSV}	**St Mary's ratio** ICA_{PSV}/CCA_{EDV}
< 50	< 125	< 2	< 8
50 - 59	> 125	2 - 4	8 - 10
60 - 69	> 230	> 4	11 - 13
70 - 79	> 230	> 4	14 - 21
80 - 89	> 230	> 4	22 - 29
> 90 but less than near occlusion	> 400	> 5	> 30
Near occlusion	High, low - string flow	Variable	Variable
Occlusion	No flow	Not applicable	Not applicable

[29]. The same study stated that while North America predominantly used the NASCET criteria, some centres in Europe preferred to use the ECST indices

while others preferred the NASCET indices. In an attempt to standardise practice in the UK, a joint working group was set up between the Vascular Society of Great Britain and Ireland and the Society for Vascular Technology of Great Britain and Ireland [28]. In a report issued by Oates *et al.* in 2009 [28], the recommendations of that working group were presented, suggesting the use of NASCET criteria. As of 2017, the clinical practice guidelines of the European Society of Vascular Surgery [27] continue to adopt the NASCET criteria throughout the guidelines unless stipulated otherwise [27].

1.4.3 Intima Media Thickness and Plaque Burden

Ultrasound technology also provides for additional metrics that are often associated with monitoring of atherosclerosis progression. Intima media thickness (IMT) is a metric which measures the thickness of the tunica intima and tunica media, the innermost two layers present on the wall of an artery [6]. Figure 1.7 shows a typical example being assessed via brightness mode imaging.

The measure of IMT is often used in research studies to investigate the e↵ect of pharmacotherapeutic agents [31; 6]. Researchers hypothesise that longitudinal changes in carotid IMT are associated with CVD risk factors, but also that race, ethnicity, and use of anti-hypertensive and lipid-lowering medications

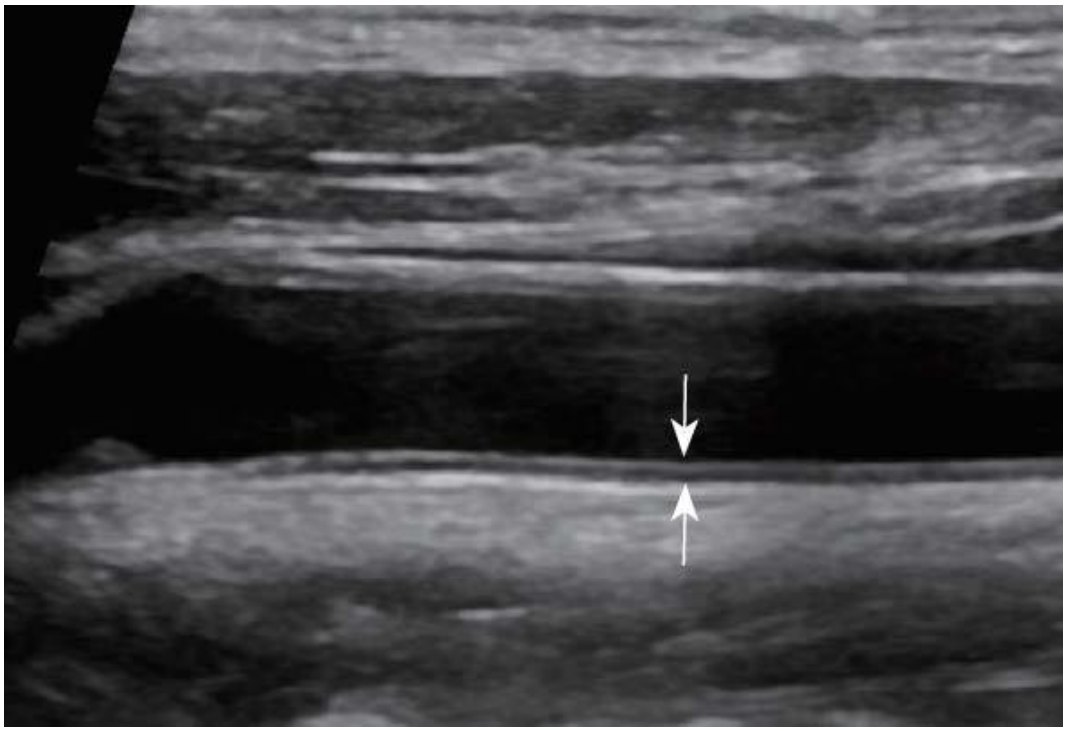

Figure 1.7: Measurement of IMT in longitudinal view of common carotid artery. Image reproduced with permission from [30].

a↵ect the progression of carotid IMT [31]. From a clinical perspective, a number of studies have demonstrated that increase in cardiac IMT independently predicts CVD events and increased risk [32; 33; 34; 35]. Concurrently however, conflicting results are obtained by many other studies as well [36; 37; 38]. This inconsistent predictive value of IMT may likely be associated with the variability in the image acquisition, IMT processing methods, operator errors, and also hardware. The annual change in carotid IMT is believed to be in the region of 0.01 to 0.04 mm, and is lower than the current-generation ultrasound pixel resolution of 0.1 to 0.2 mm [6]. This therefore makes it di cult to follow up carotid IMT change.

Given the inconsistent results related to predictive power of carotid IMT for CVD risk, it is unsurprising to see that recommendations for its use vary. The European guidelines on CVD prevention in clinical practice supported its use in the publication issued in 2012 [39], however in the latest version in 2016, its use is no longer recommended [40], with the task force citing that recent metaanalyses failed to demonstrate any added value of IMT in comparison to the Framingham Risk Score. Similarly, the 2013 American College of Cardiology / American Heart Association guidelines on Cardiovascular Risk Assessment do not advocate routine measurement of cardiac IMT, in view of concerns in meta-analyses of quality and standardisation of its measurement [41].

Now, although increasing evidence shows that IMT provides little added-value in terms of risk prediction of CVD events, the inclusion of plaque along cardiac IMT has been consistently shown to improve the predictive power for CVD and coronary events [42; 43; 44]. To improve the predictive power even further, the quantification of plaque burden (as opposed to simple plaque presence), has been developed as an interesting replacement to cardiac IMT measurement

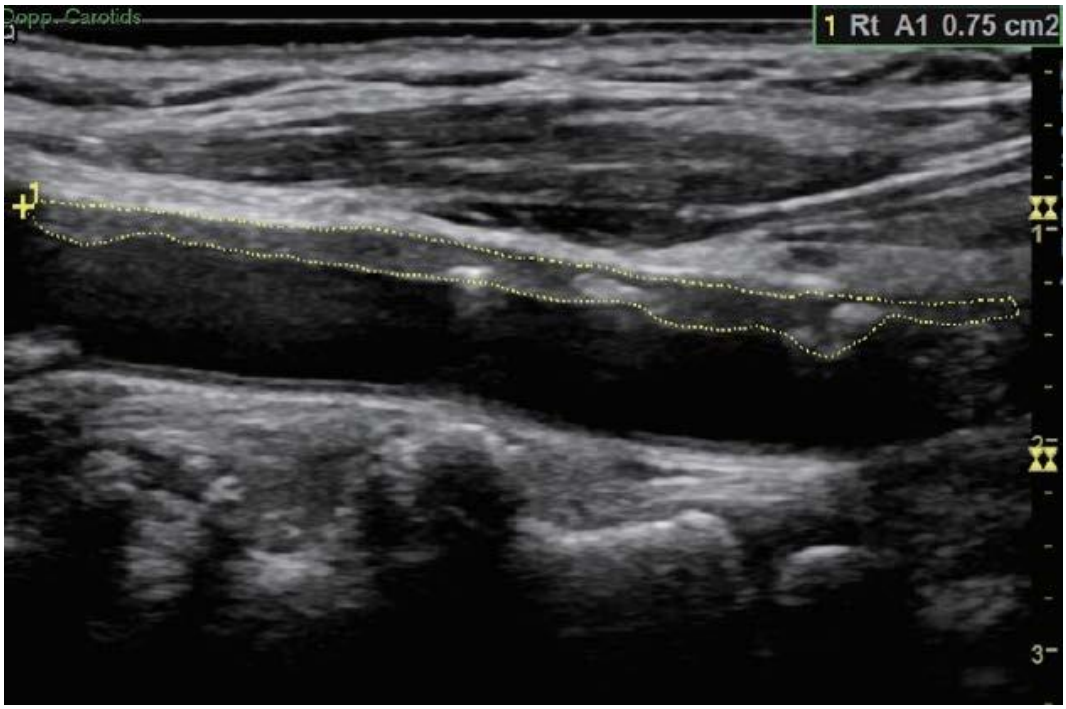

Figure 1.8: Measurement of TPA in longitudinal view of common carotid artery. Image reproduced with permission from [30].

[45; 46]. Carotid plaque burden is measured as either the total plaque area (TPA) using 2D ultrasound planimetry, or using total plaque volume (TPV) by 3D ultrasound (described in the next section). TPA is the total cross-sectional area of detected plaque when observed in longitudinal views as shown in Figure 1.8, whereas TPV is obtained by detecting plaque on axial views over a number of image frames as shown in Figure 1.9 [45].

Plaque burden quantification has been shown to be superior to IMT, since its progression strongly predicts cardiovascular events and high risk patients. Cardiac IMT changes by subtle values of 0.01 to 0.04 mm per year [6], and therefore is weak in predicting CVD risk or monitoring medical therapeutic effects. In contrast, TPA and TPV changes are significantly more noticeable,

with values around 10 mm^2/year and 50 to 100 mm^3/year respectively [47]. This makes it easier to monitor changes, and allows for smaller sample sizes and shorter durations of follow ups required to observe e↵ects of new therapies [48]. Plaque burden may also be used to manage patients with asymptomatic carotid stenoses, whereby intensive medical therapy can be given to patients based on their plaque measurement, and thus resulting in a marked reduction in CVD events [30]. Finally, plaque burden may also be used in genetics research, since it allows for genotyping the sampled patients. Carotid ultrasound phenotypes are biologically di↵erent, signifying that IMT, plaque burden, stenosis and plaque rupture are a↵ected by di↵erent genetic factors
[30]. [45].

The review paper by Stella *et al.* [30] discusses additional areas of research on

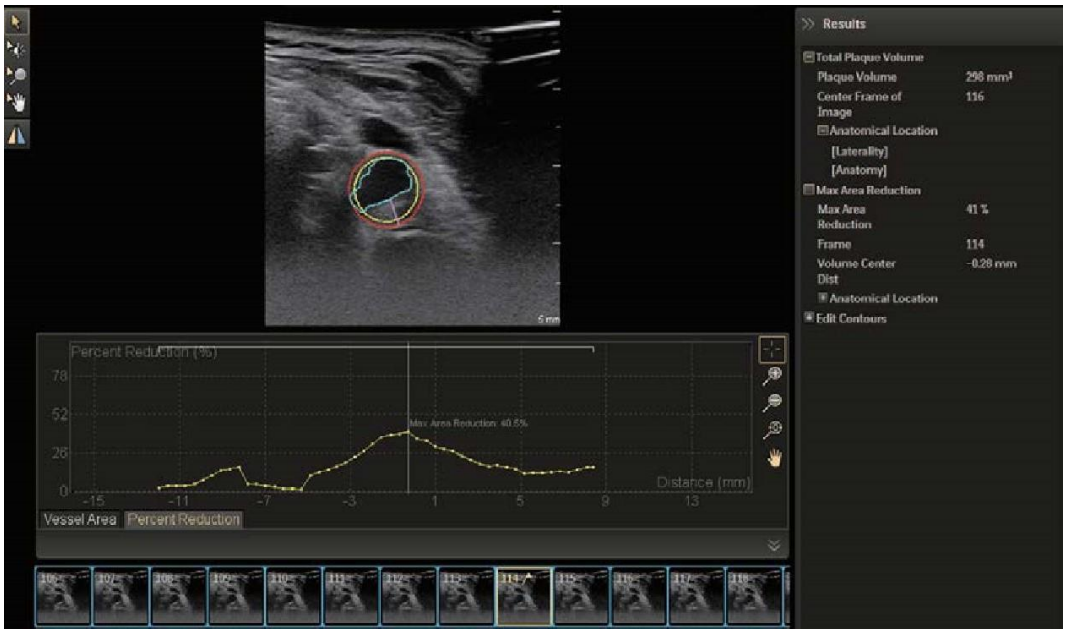

Figure 1.9: Measurement of TPV in axial view from a number of US slices. Image reproduced with permission from [30].

plaque characterisation as a means of risk stratification. Characterisation of plaque by means of plaque echogenicity and morphology provides additional information to categorise vulnerable and non-vulnerable plaque, and this in turn may help risk stratification [30]. Plaque echolucency has been shown to be a strong marker for risk of ischemic stroke, and has been proven by means of histology to represent lipid rich necrotic cores or intraplaque haemorrhage [49]. Contrast enhanced ultrasound can be used to detect neovascularization and to quantify neovessels in the entire plaque. This technique has been suggested to be more reliable to predict risk of plaque rupture or intraplaque haemorrhage. [50]. Finally, abnormalities in plaque motion have been observed with high resolution ultrasound on symptomatic carotid stenoses, and these have subsequently been proven by histology to represent high-risk lesions from plaque rupture to ulcers [51].

1.5 Ultrasonography

1.5.1 Physics of Ultrasound Imaging

Ultrasound waves are an acoustic type of waves generated with an ultrasound probe at frequencies between 20 KHz and 50 MHz [52]. They are produced when a probe (often also referred to as a transducer) makes contact with a material and vibrates to produce minute push-pull actions in rapid succession. The push-pull actions are what cause mechanical vibration and which in turn generate the waves. Waves, which oscillate at a frequency of below 20 KHz, are referred to as audible acoustic waves, whereas waves oscillating at frequency above 20 KHz are termed ultrasonic, because they are too high pitched to be heard by the human ear [52; 53].

Very commonly, the acoustic waves that we are familiar with are produced by a vibrating source in air, such as by vocal cords, machinery or loudspeaker. In ultrasonic transducers used in medical ultrasound, the vibrations are caused by one or several piezoelectric crystals mounted next to each other in the probe, and made to vibrate by the application of an electric voltage signal [52]. Similarly, when ultrasonic waves, which are reflected from material, are incident on the same piezoelectric crystals, they cause them to vibrate and to generate electrical voltages, which are proportional to the magnitude of vibration. Most ultrasound machines are not made to generate continuous waves, but rather they operate in short bursts to create pulses of three to four cycles in duration. The structures which are being examined are then visualised by detecting the reflected wave echoes, which occur when the transmitted waves are reflected or scattered by changes in tissue structure [52].

The vibrating action of the ultrasound transducers against the face of tissue causes regions of compression and rarefaction, which in turn result in regions of increased and decreased tissue density. This is shown in Figures 1.10a and 1.10b. These regions of compression and rarefaction may be represented as ultrasound waves, having the typical characteristics of wavelength and amplitude [52; 53]. The wavefront will be planar if the ultrasound waves are generated by a transducer having a flat face. Conversely, if the transducer has a convex face, then the wavefronts generated will also be convex. These might be used, for instance, to focus on a particular region at a particular distance from the transducer [52].

The frequency of the ultrasound wave refers to the number of oscillations per second that a particular point in the medium is made to oscillate at rapidly, back and forth. Additionally, the ultrasound wave traverses the medium at a

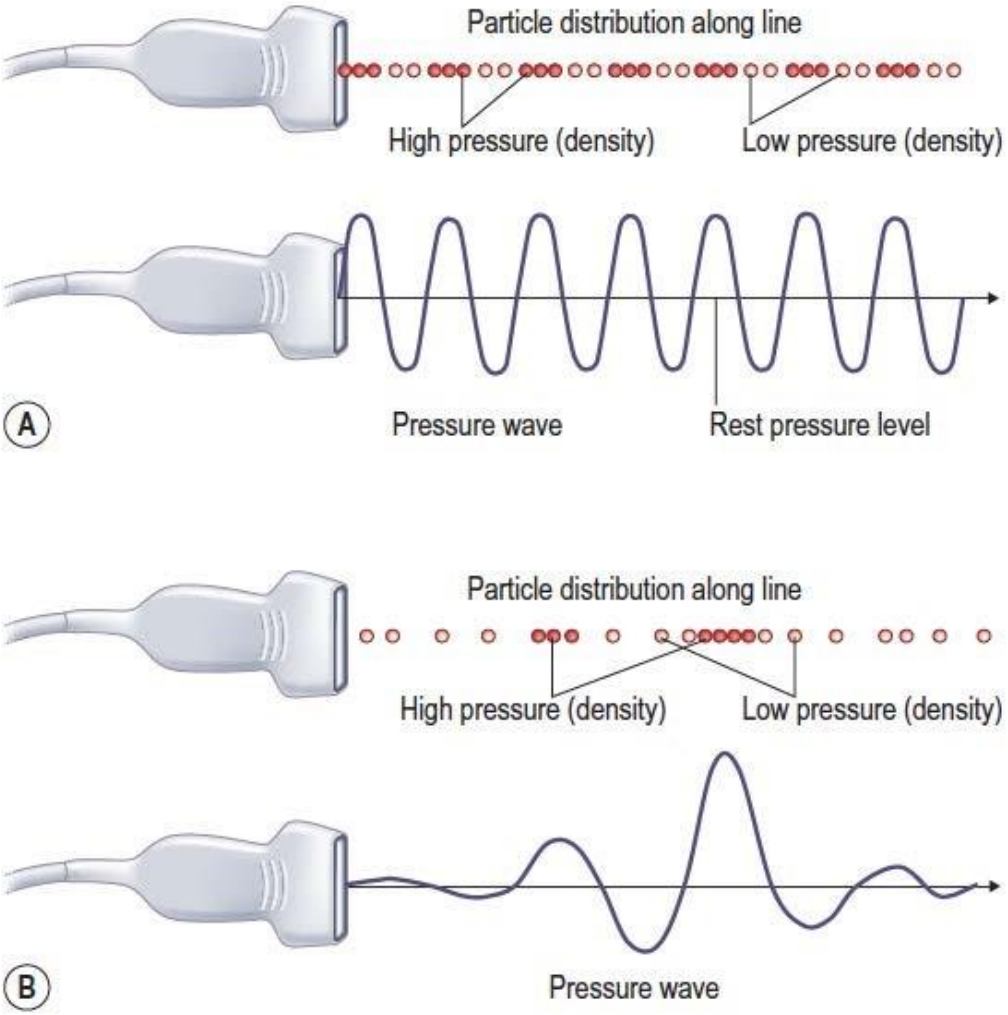

Figure 1.10: The waveform description of ultrasound pressure fluctuations: a) showing compressions and rarefactions; b) showing pulsed waves. Images reproduced with permission from [52].

certain speed, which in most human soft tissue is close to around 1540 m/s [52; 53]. This enables ultrasound pulses to be transmitted and the echoed reflections collected at very high speed. This, in turn, allows us to generate a vast number of images per second. An important formula to consider which Table 1.3: Speed of ultrasound in di↵erent tissues. [53]

Type of Material	**Speed of ultrasound (m/s)**	**Acoustic Impedance (g/cm^2)**
Water	1480	1.48 ⇥ 10^5
Blood	1570	1.61 ⇥ 10^5
Bone	3500	7.80 ⇥ 10^5
Fat	1450	1.38 ⇥ 10^5
Liver	1550	1.65 ⇥ 10^5
Muscle	1580	1.70 ⇥ 10^5
Polythene	2000	1.84 ⇥ 10^5
Air	330	0.0004 ⇥ 10^5
Soft tissue (average)	1540	1.63 10^5 ⇥

relates the speed of sound, frequency of sound wave and wavelength is:

$$c = f \tag{1.1}$$

whereby c represents the speed of sound, f represents the frequency of the wave and represents the wavelength. The rigidity and density of soft tissue has a bearing on the speed of the wave travelling through it. The less rigid a material is, the lower the speed of the wave travelling through it, as can be seen from Table 1.3. One observes, however, that the speed of ultrasound in human soft tissue may be averaged to around 1540 m/s [52; 53]. This value is important to diagnostic ultrasound, as in conjunction with equation 1.1, is used to determine the depth of tissue reflecting the wave, as shall be seen later.

Ultrasound transducers are one of the central pieces of hardware behind clinical ultrasound imaging. Their performance is critical to the generation of

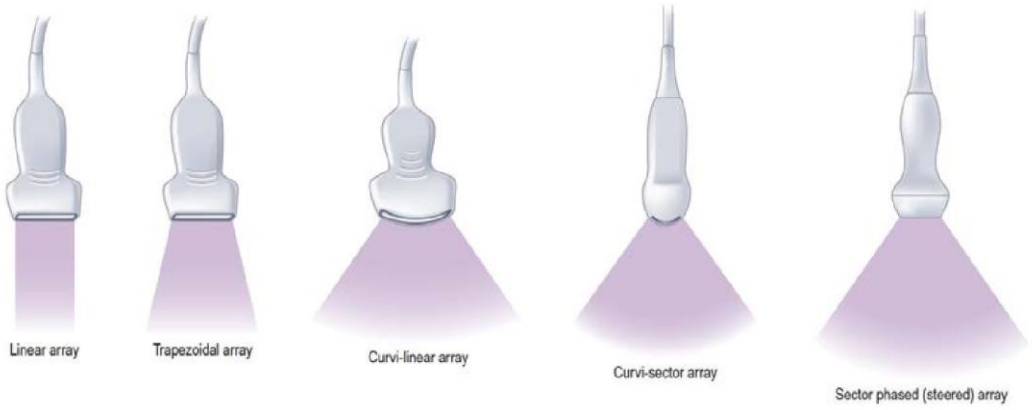

Figure 1.11: Di↵erent types of transducers and the ultrasound beam they produce. Images reproduced with permission from [54].

a quality image. An ultrasound transducer receives an electrical signal from the processor of the machine and converts the latter into an ultrasound pulse, which travels along the medium as a wave [54]. The main component of a transducer is a very thin layer of piezoelectric material, to which metal electrodes are connected on either side.

To form 2D ultrasound images, an array of PZT crystals are present along the length of the transducer and a beam is produced by each crystal, moving in sequence along the length of the transducer face. Array transducers come in various shapes, producing di↵erent imaging formats for di↵erent applications [54]. Figure 1.11 depicts some of the most common types of array transducers, which are used for transcutaneous imaging.

1.5.2 Principles of Image Formation

The previous section describes how a typical 2D image is created from an array of piezoelectric crystals, each individually producing single but adjacent ultrasound beams in sequence. Assuming a very simple image forming process,

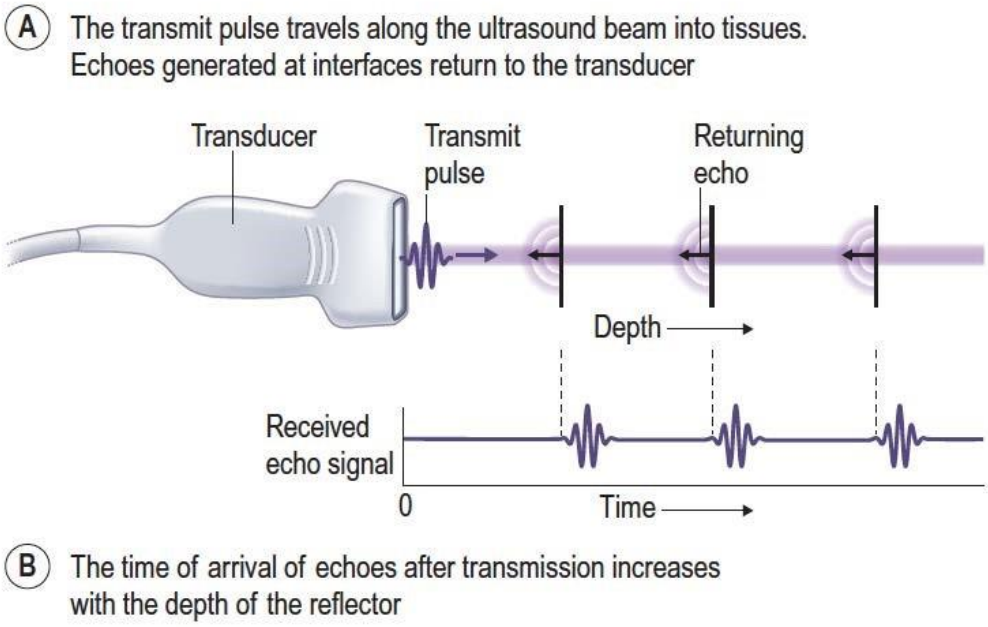

Figure 1.12: [A] The transmitted pulse travels into the tissue and is reflected at various interfaces. [B] The time it takes for echoes to arrive is proportional to depth of the interface. Images reproduced with permission from [54].

each such beam, or 'line' within the eventual ultrasound image, is formed by a single pulse-echo cycle, as depicted in Figure 1.12. Each crystal in the array produces a pulse of a few wave cycles at a particular transmission frequency (for example, 5 MHz). The pulse travels into the tissue at the speed of sound, *c*, indicated for that type of tissue, until it eventually reaches an interface between two di↵erent tissues having di↵erent acoustic impedances. This

interface causes the pulse to be partially reflected and partially transmitted onwards. The reflected pulse travels backwards at the same speed of sound, *c*, until it reaches the transducer. The pulse, which is transmitted onwards, travels further down the tissue, where it is reflected at various other interfaces. Each time it produces echoes, which reach the transducer at di↵erent, delayed times. The image processor assumes that the time, *t*, taken between pulse transmission and receipt, is the time taken for a pulse to 'go and return' between a distant interface at depth *d*. Thus the time is defined as $2d/c$. Knowledge of time and speed allows the processor to determine the depth of the reflecting interface.

In practice, the number of interfaces in the path of the transmitted pulse are many, and these generate a constant series of echoes following transmission. Each line of echoes in the beam forms a line in the image, which is called a 'Brightness-Mode Image', or 'B-mode' image. The strength of the echo is represented by the brightness of each point along the B-mode line, whereas the time delay of the echo is represented by the distance along the line. This

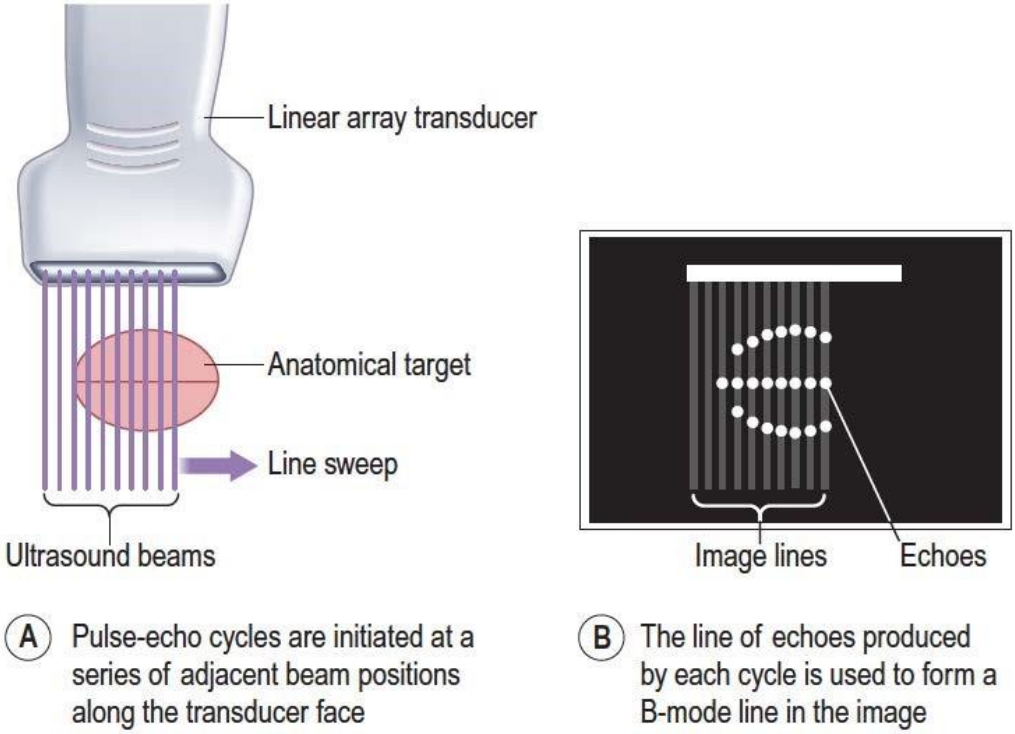

Figure 1.13: Sequence of lines which form a Brightness mode image. Images reproduced with permission from [54].

is shown in Figure 1.13. As the last echo reaches the transducer from the maximum depth possible, the next adjacent crystal fires in a similar manner, to start producing the next B-mode line. The process is repeated until all crystals fire and produce an entire image made of several B-mode lines. After the last line is created, the process is repeated from the beginning in order to have a series of image frames forming a real-time moving clip.

A typical ultrasound B-mode image taken from the abdomen may be seen in Figure 1.14. At the top of the image, the curved boundary corresponds to the surface of the skin, with which the curved transducer makes contact. Moving

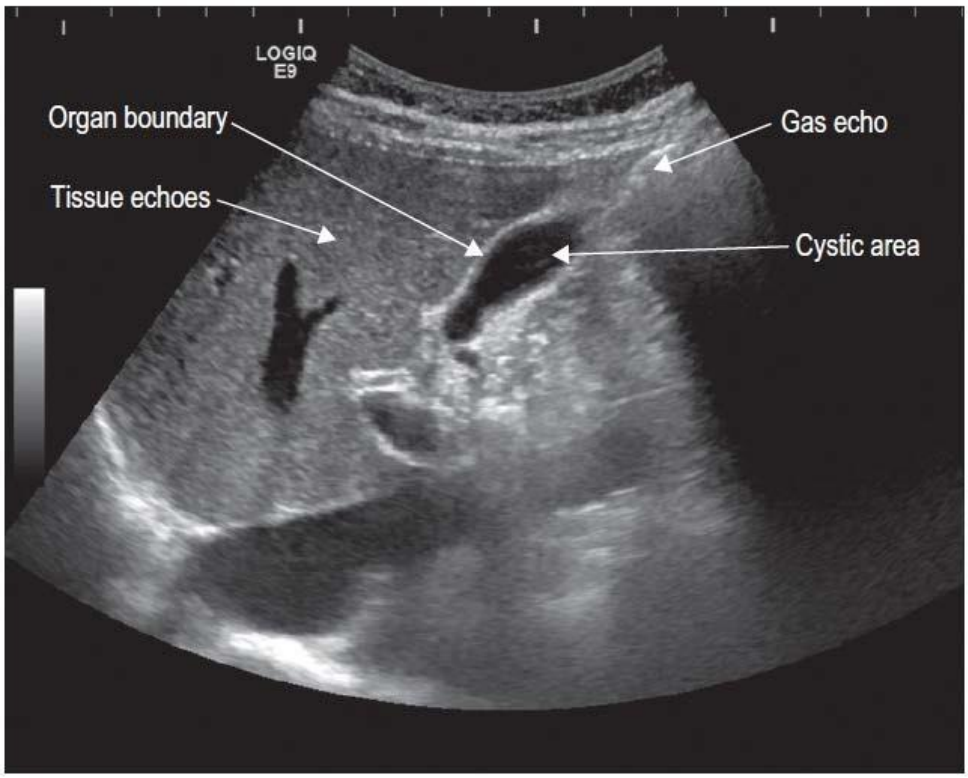

Figure 1.14: An example of a B-mode image of the abdomen. Images reproduced with permission from [54].

away from this edge and down the image corresponds to an increased depth within the abdomen. Each beam is transmitted in a straight line, but normal to the curved surface of the transducer. All the beams are also constrained to lie on a same plane, therefore resulting in the image being two-dimensional. The image is produced in greyscale. Brighter, whiter patches correspond to strong reflections, particularly at interfaces between tissue and gas or tissue and bone. Regions which have no change in acoustic properties result in dark regions.

1.5.3 The Image Processing Chain

Aside from the transducer, an additional component used in modern ultrasound systems to form the B-mode lines described previously is the beam former. This is a programmed component which activates groups of piezoelectric elements to transmit focussed pulses. The B-mode lines produced by the transducer and beam former then need to undergo a series of image processing steps prior to the final image being produced. Figure 1.15 shows a very simplified block diagram of the additional steps in the image processing chain. Although not shown, some hardware signal conditioning steps take place right after the data is acquired by the transducer elements, in preparation for the signals to be digitised. The remaining processing steps, including the beamformer are therefore carried out digitally.

Time Gain Compensation (TGC). When ultrasound waves propagate through tissue and are reflected, the echoes reflected from deep tissue are weaker in strength than those which are closer to the transducer. This is due to the attenuation that takes place while propagating through tissue. However, the final B-mode image produced by the ultrasound needs to show similar brightness of

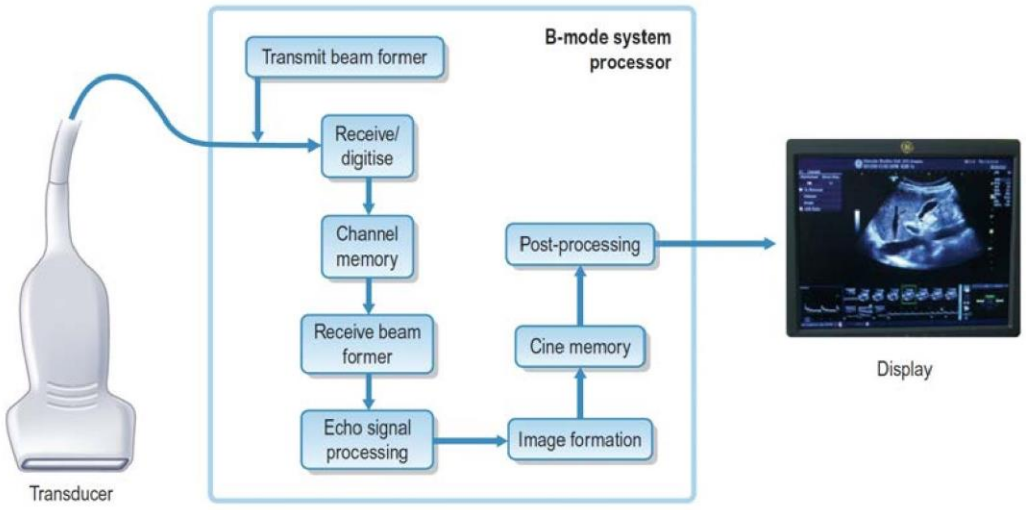

Figure 1.15: A simplified block diagram showing the components of a B-Mode System. Images reproduced with permission from [54].

similar types of tissue, irrespective of their depth. Thus, the attenuation due to depth is compensated for, by increasing the amplification applied in proportion to the delay of signal arrival following transmission. This technique is called time gain compensation (TGC). The ultrasound machine will always apply a baseline level of TGC, however the user is allowed to manually adjust the TGC applied to di↵erent levels of depth using special slide controls.

Dynamic Range. A variety of di↵erent interfaces within tissue may result in a vast range of di↵erent echo strengths, with some being a hundred times stronger than others. Moreover, many echoes forming in the image result from small scattering structures within the tissue, which generate very weak signals in comparison to interface echoes. These can sometimes be weaker by a factor of 1 million, but may still contain important diagnostic information. The ultrasound machine must be capable of displaying this large variety of echo strengths at the same time on the display monitor. However, typical flat

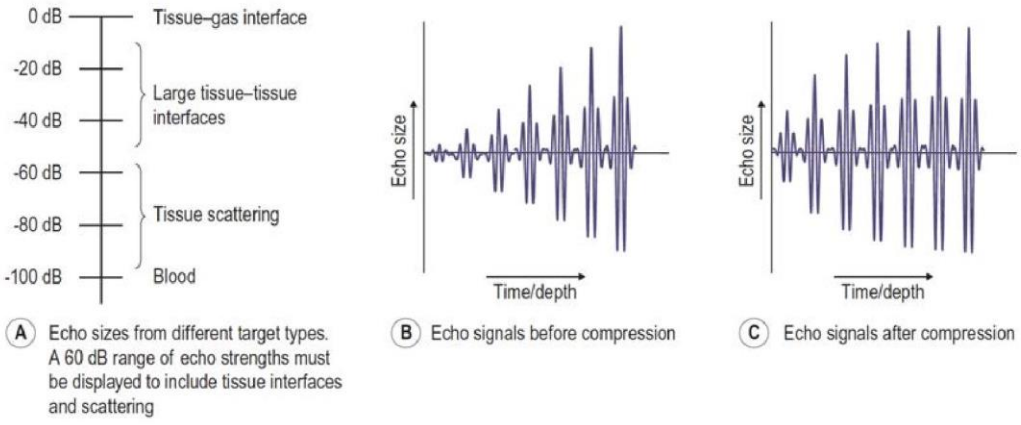

Figure 1.16: [A] Signal strength of di↵erent targets. [B] Echo signals before compression. [C] Echo signals after compression. Images reproduced with permission from [54].

screen monitors have a dynamic range which is much lower than that required to display all the levels of intensities. Thus, in order to correctly display all intensities at the same time, the dynamic range of the received echoes is compressed by the system processor, such that weak signals are amplified by a large factor, whereas large signals are amplified less. The di↵erent echo strengths of signals are normally expressed on a logarithmic scale in dB, given the large variety in order of magnitude. Figure 2.12a gives some examples of strength of di↵erent targets, while Figure 2.12b and 2.12c show the amplitude of signals before and after compression.

Demodulation. The signal processing steps that have been described so far are always applied on the raw Radio-Frequency (RF) signal. The RF signal is that which still contains the original ultrasound transmission frequency. Since the brightness levels shown in the final image are related to the amplitude of this RF signal, the original transmission frequency is not required. Thus, a Hilbert

Transform is typically applied on the RF signal to remove the radio-frequency component and obtain just the envelope. This step is referred to as demodulation.

Image Memory. After the preceding image processing steps described, the B-mode image is stored in memory as an array of pixels, with each pixel capable of taking on 256 di↵erent grey levels. Additional and successive images are stored sequentially in a cine memory, until the available memory is full. At this point, new and successive images are stored by overwriting the oldest frame, such that the most recent images are always retained. The user may then be able to review a number of frames in a cine loop for diagnostic purposes.

Additional Processing. Many other image processing steps, which are not described here, may be applied to the images by the machine or by the user if chosen. Manufacturers are constantly developing such post-processing steps to enhance the display image further. These may include techniques which smoothen the image to reduce noise and speckle, or else accentuate edges and boundaries, making them sharper.

1.5.4 The Doppler E□ect

The ultrasound techniques described so far were based on the assumption that the reflecting targets during the formation of each image remain stationary. However, through the use of the Doppler e↵ect, one may detect the presence, direction and velocity, of a moving target such as blood inside vessels [53]. In turn, the velocity of blood inside a vessel may be used to infer information regarding the presence of stenosis, as described in Section 1.4.2. The Doppler e↵ect is a phenomenon which sees a change in the frequency of sound

observed, if there is a relative motion between the the source of the sound and the receiver [53]. In the stationary situation, a receiver views the same number of pressure waves as that emitted by the source. However, if there is a motion of the source, this distorts the pattern of observed wavefronts seen by the receiver. The observed frequency is influenced by the rate of movement of the source or receiver towards each other or away from each other [53]. More pressure waves per unit time strike the receiver, if the source / receiver are moving towards each other. This causes an elevation in the frequency observed. Conversely, the opposite is true if they are moving away from each other. The change in frequency between the actual transmitted frequency and the received frequency is called the Doppler frequency shift [53].

The magnitude of the doppler frequency shift depends on how rapidly the source, receiver, or both, are moving. In other words, an increase in the relative velocity between the two will cause a greater di↵erence between transmitted and received frequency (and hence a greater Doppler frequency shift). The Doppler frequency shift is also a↵ected by the speed of sound in the medium concerned, and also by the initial transmitted frequency. Thus, the equation to calculate the Doppler frequency shift f_D is given as follows [53]:

$$f_D = \frac{2vf}{c} \tag{1.2}$$

where c is the velocity of sound in tissue, v is the velocity of the interface and f is the transmission frequency of the transducer. Equation 1.2 assumes that the velocity of the interfaces for biological systems is in fact relatively small in comparison to that of speed of sound in tissue (1540 m/s) [53]. Using this equation, given that we know the speed of sound in tissue, the probe's transmission frequency, and that the doppler frequency shift can be measured

by the machine from the reflected pulse, it is possible to determine the velocity *v* of the moving medium. An increase in the velocity of the interface will

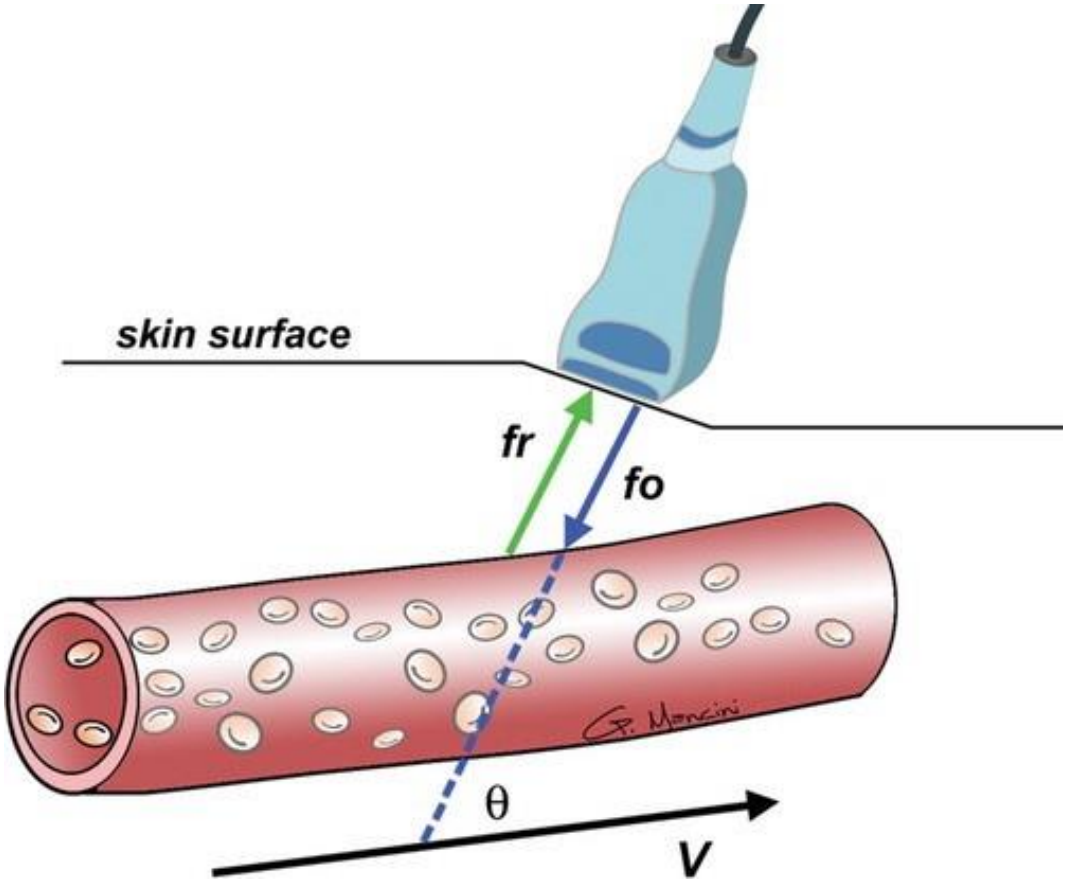

Figure 1.17: Diagram showing the doppler e↵ect with a sound beam which is incident to the reflecting moving interface at a doppler angle '. Images reproduced with permission from [55].

correspond to an increase in the measured doppler frequency shift and vice versa [53].

Furthermore, equation 1.2 is only valid in circumstances where the direction of motion between sound source and the moving reflector is parallel to the

propagation of the sound wave [53]. If, as shown in Figure 1.17 the angle of incidence of the sound beam is di↵erent from 0 degrees with respect to the motion of the reflecting interface, then the equation changes to include the cosine of this insonification angle ' (also referred to as the Doppler angle), as follows [53]:

$$f_D = \frac{2vf cos\varphi}{c} \tag{1.3}$$

1.5.5 Modern Vascular Ultrasound Systems

Modern ultrasound systems have various imaging techniques available when it comes to visualise and diagnose conditions related to the vascular system. In 2D imaging, the sonographer usually starts by using standard brightness mode imaging to investigate the vascular anatomy of the site of interest [56]. As explained in section 1.5.2, a brightness mode image is constructed from many individual B-mode lines, which encode the strength of reflections from tissue interfaces with single amplitude (brightness) [54]. In vascular imaging,

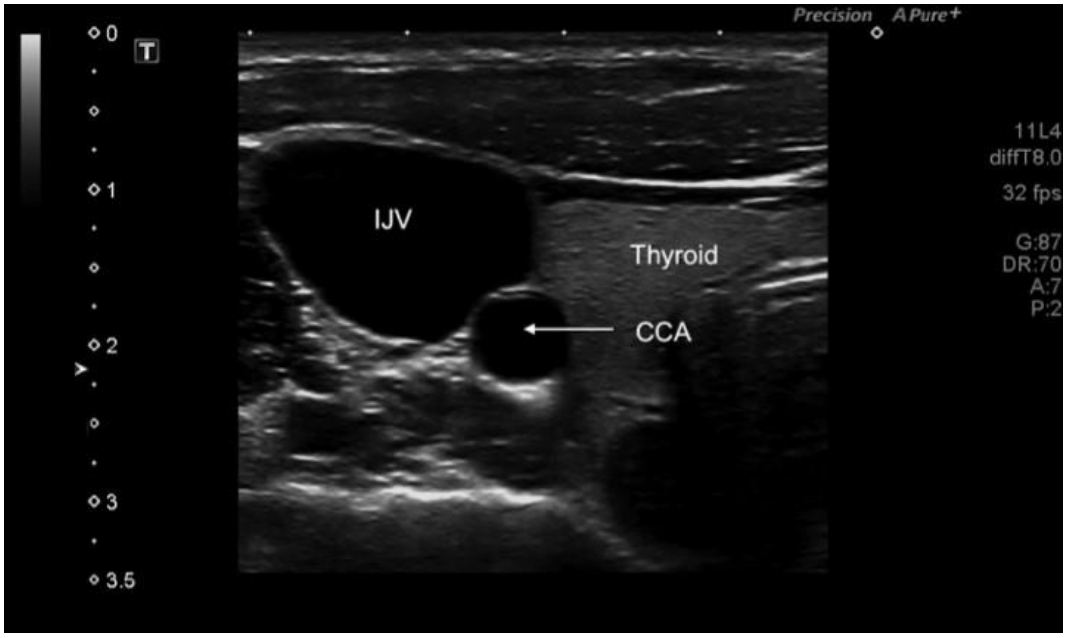

Figure 1.18: A typical brightness mode image visualising the carotid ultrasound in axial view.

brightness mode imaging is initially used to locate the the vascular structure within the field of view of the ultrasound's screen, as shown in Figure 1.18, and to determine the orientation of imaging (axial vs longitudinal) [56]. In case measurements of IMT are being taken, these would also be taken using brightness mode imaging.

Next, the sonographer will typically use Colour Doppler to identify regions of high flow. Colour Doppler uses a superimposition of colour maps over the brightness mode image, to indicate the direction and velocity of blood flow [56].

Figure 1.19: A colour doppler image showing flow in an internal carotid artery. Images reproduced with permission from [54].

Typically, shades of blue signify movement in one direction, whereas shades of red indicate movement in the opposite direction. The brighter the colour in each direction, the higher the velocity being registered at that spatial location [56; 57]. An example is shown in Figure 1.19. Colour doppler is useful for two principal reasons. Firstly, to help in identifying branches of the external carotid artery, and secondly, to locate the fastest flowing stream of blood in a vessel. In this technique, however, the dependence of the Doppler frequency shift on the rate of change of phase angle presents problems with background noise. Since the latter noise also has a random phase angle, it is mistakenly mapped as a background colour, and therefore appears as a flow with random velocity and direction [58]. Additionally, given that the magnitudes of noise and true signals are often similar, it is easy to have the true signal masked out by noise unless instrument settings are properly selected. Other limitations with the technique are its dependency on the angle of insonation and the presence of frequency aliasing [58].

After identifying locations with the fastest flowing streams of blood in the vessel, the sonographer uses a technique called Pulsed-Wave Doppler (or Spectral Doppler Imaging) to calculate the velocity of moving reflectors based on the reflected Doppler frequency shift [56]. The information is visualised using the Doppler spectral waveform underneath the B-mode and colour doppler as shown in Figure 1.20. The doppler spectral waveform includes information about three variables which are: frequency, magnitude and time [53]. The frequency is displayed along the vertical axis, whereas the magnitude of di↵erent frequency components is represented by the brightness of the

points making up the waveform. The x-axis is reserved to show time information, in order to show how the velocity changes over time [53]. This doppler spectral waveform allows the user to calculate the peak systolic velocity (PSV), which then as

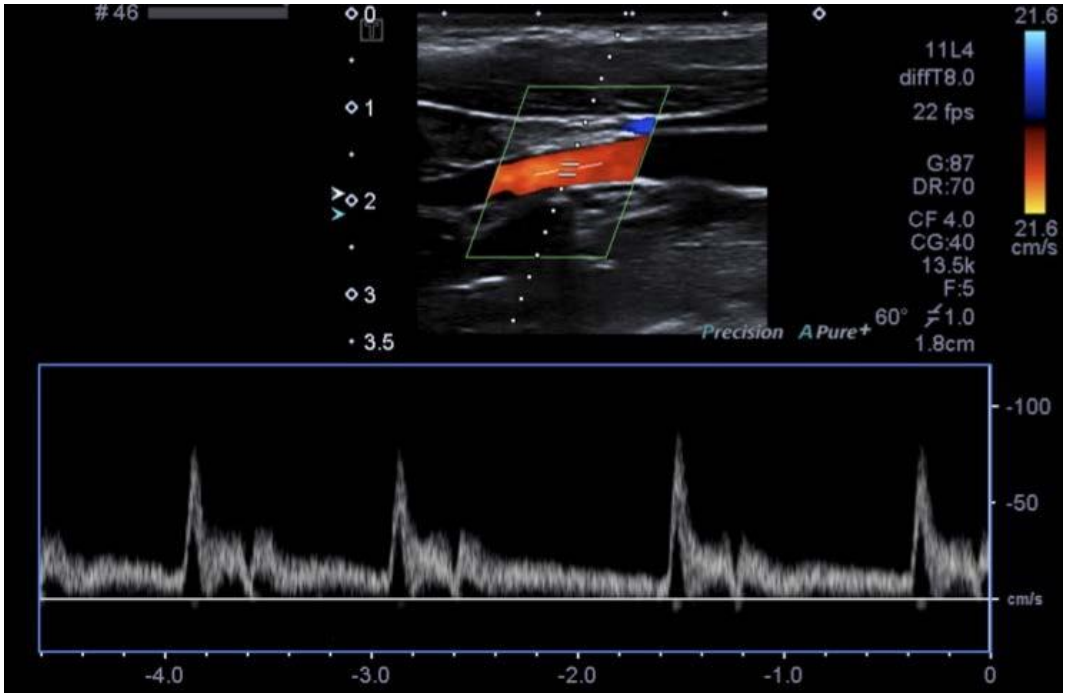

Figure 1.20: A doppler spectral waveform shown underneath the color doppler. Images reproduced with permission from [56].

described in section 1.4.2, is used to determine the ratio of PSVs from di↵erent sites to yield information about the amount of stenosis in a vessel.

Another technique which is sometimes used is called Power Doppler Imaging. This was developed in response to the issues encountered with colour doppler, whereby intravascular colour signals were generated from the amplitude of the echo signals reflected o↵ red blood cells [59]. Like Colour Doppler Imaging,

this technique may be used to display vascular blood flow in colour. However, rather than displaying the mean Doppler frequency shift, its colour map displays the integrated power of the Doppler signal, therefore making it independent of velocity and angle of insonification, as well as aliasing-free [58], [60]. Furthermore, it was shown to be more sensitive to smaller degrees of

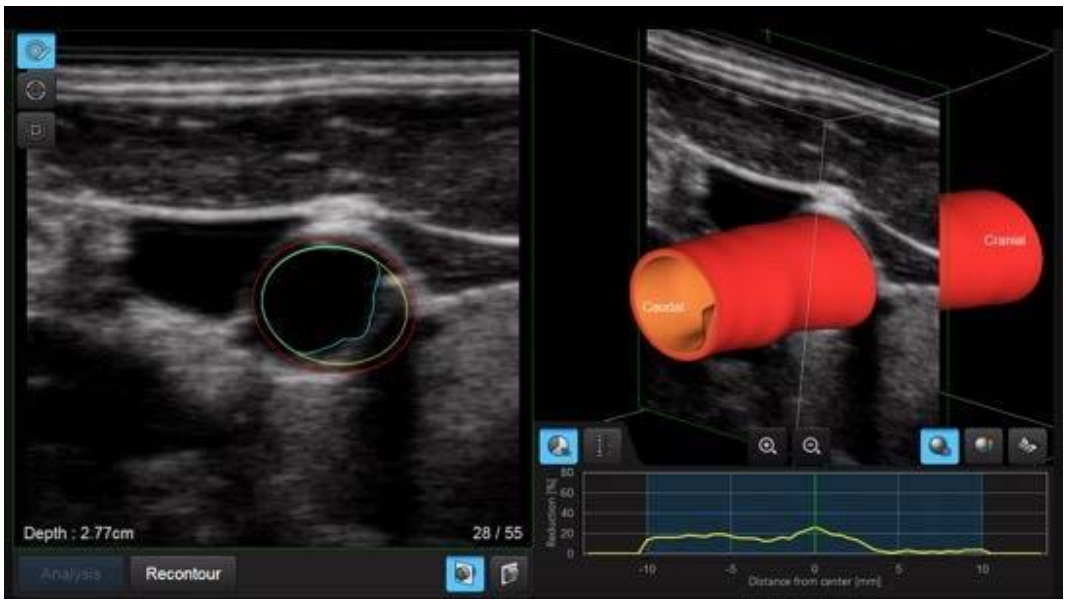

Figure 1.21: A three dimensional reconstruction of an artery from manual delineation in 2D. Taken from a commercial manufacturer.

blood flow, and also allowed for better di↵erentiation of atheromatous plaque surface morphology. On the downside the technique does not permit the determination of velocity per se, nor the direction of the movement of blood [59].

1.6 Three Dimensional Measurement of Plaque

Although Spectral Doppler Imaging and Colour Doppler Imaging are both widely used nowadays in the quantification of atherosclerosis and arterial stenoses, there are studies which have cast doubt on both their e cacy when gauging low to moderate degrees of occlusion, where di↵use disease is present, or when the patient su↵ers from cardiac arrhythmias [61; 58; 62]. Additionally, both techniques also have an inherent disadvantage, due to them being two-dimensional imaging modalities. When only one plane is being imaged, there exists the possibility of either mis-grading a stenosis or even missing it entirely [58]. This happens because stenoses may be seen well in one particular plane, but might not show up at all in another. Additionally, plaque is known to progress quicker along the vessel wall in comparison to the rate at which it thickens the vessel wall [63]. Thus, although carotid stenosis measurement remains the tool of choice to assess patients following stroke or TIAs, when it comes to monitoring disease progression and prediction of cardiovascular events, it remains a weak indicator [64]. A solution thought to possibly resolve such problems lies in three-dimensional imaging.

1.6.1 Three Dimensional Ultrasound Systems

Three-dimensional ultrasound has drawn a lot of interest over the past decades [65; 66]. Classically, 3D ultrasound systems would reconstruct 3D structures by first acquiring a number of 2D images throughout the volume of interest. The 2D images are then collated as a set of parallel slices, in shapes such as cuboids, wedges or ones. The orientation and position must be accurately known to avoid imaging artefacts arising from geometric distortions [53]. Mechanical and freehand scanning methods have been

developed to sample the volume of interest. More recently, the approach has been to use a native matrix 2D array probe which can acquire the 3D dataset without motion of any part of the transducer [53].

Mechanical Scanning Methods

Mechanical scanning methods use either external mechanical assemblies or else moving mechanical assemblies within the probe itself to scan the acquisition through a range of motion [53]. In the external assemblies, the transducer is mounted on a mechanical assembly that translates the transducer linearly

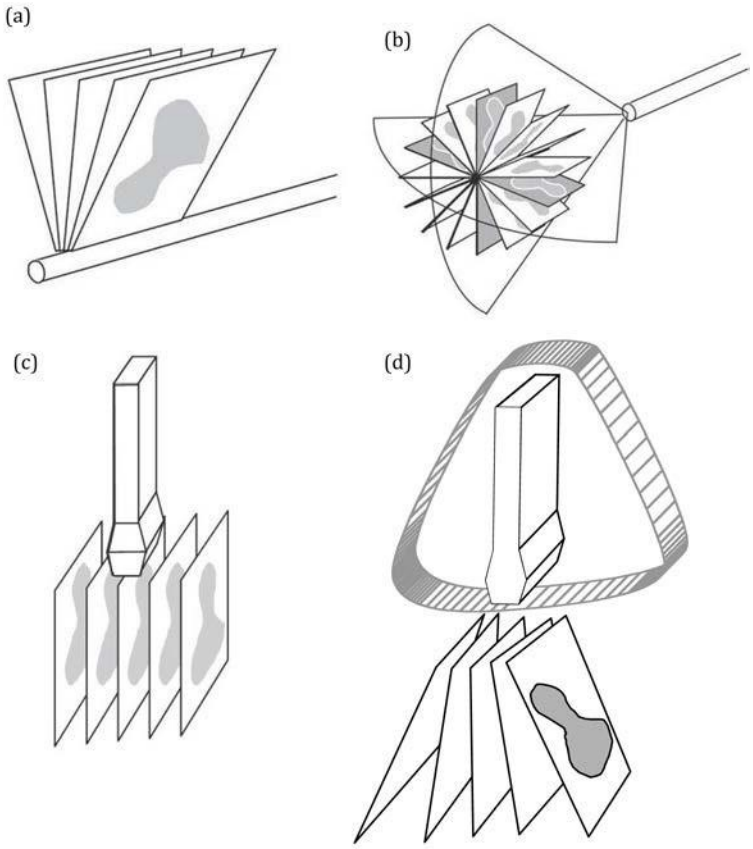

Figure 1.22: A) and B) Rotationally swep 3D transducers; C) Linearly translated transducers (external assembly); D) Fan-swept 3D transducers. Reproduced under open access agreement.

over the patient's surface. The 2D images are acquired at defined intervals, but most importantly in a parallel fashion. In internally-swept mechanical assemblies, the transducer is tilted or rotated about an axis, and the system automatically acquires images at constant angular intervals [53]. The resulting 3D volumes are arranged in a fanlike wedge or cone. Examples of all three types of methods are shown in Figure 1.22.

Freehand Scanning Methods

Freehand scanning methods have the user manually move the transducer across the patient, with some position tracking technology which senses the position and orientation of the probe [53]. The most common position tracking techniques include: A) Mechanical field sensors, whereby a transmitter nearby the transducer generates a varying magnetic field, and a receiver mounted on the ultrasound probe, containing a set of orthogonal coils, measures the field strength in three directions and allows determination of position and angulation. An example is shown in Figure 1.23; B) Inertial Sensors, whereby 3D accelerometers and 3D gyroscopes measure angular velocity and earth's gravitational force, and through a series of mathematical transforms allow position and orientation information to be obtained; C) Image-based correlation, whereby successive images are analysed with respect to their energy pattern. The shift in relative position of adjacent images is determined from the change image characteristics [53].

2D Matrix Array Methods

The third type of 3D image reconstruction involves using native rectangulararray (matrix) type transducers [53]. This acquires the image natively in three dimensions, since several rows of piezoelectric transducers are able to each acquire a 2D frame, and the ultrasound system then collates them into a 3D array of voxels straight away. The ultrasound beam sampling direction may still be controlled electronically via the system using the appropriate sequence of crystal firing [53]. Parallel processing architectures then reduce the sampling and processing time. Matrix probes also o↵er the additional advantage of acquiring data in orthogonal planes [53]. These types of probes were once

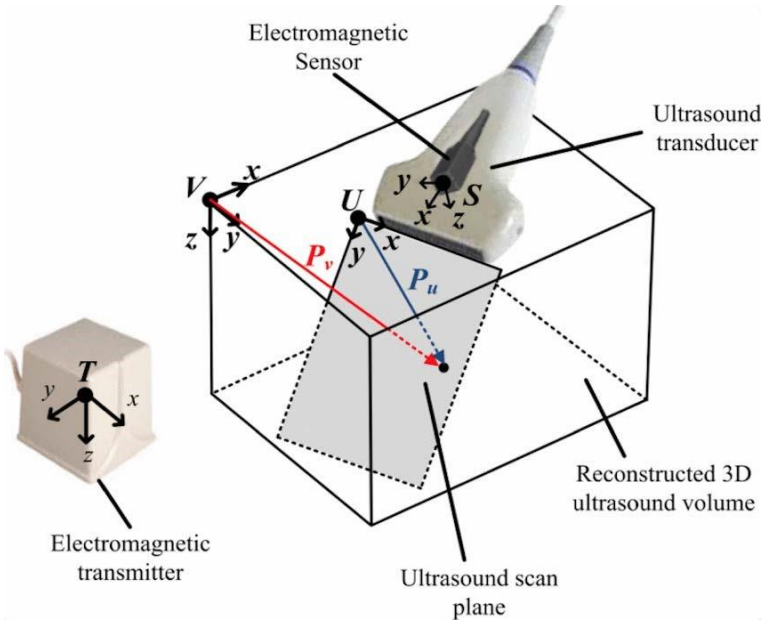

Figure 1.23: An example of a freehand type of position tracking using an electromagnetic transmitter and a receiving sensor on the ultrasound probe. Reproduced with permission from [67].

prohibitively expensive, but advances in technology has made their price accessible to the commercial market, and they are nowadays widely used. An example may be seen in Figure 1.24.

1.6.2 Three Dimensional Carotid Plaque Measurement

Along with the proposed reconstruction of 3D vascular structures using three dimensional ultrasound, novel volumetric indices derived from 3D images have also recently been proposed to help with the characterisation of plaque burden, as a potential replacement to disease monitoring using carotid stenosis. The total plaque volume (TPV) was initially reported by Ainsworth *et al.* in [48] to be e↵ective at measuring the e↵ects of anti-atherosclerotic treatments, by gauging the e↵ect on TPV in three dimensions. predictive of cardiovas-

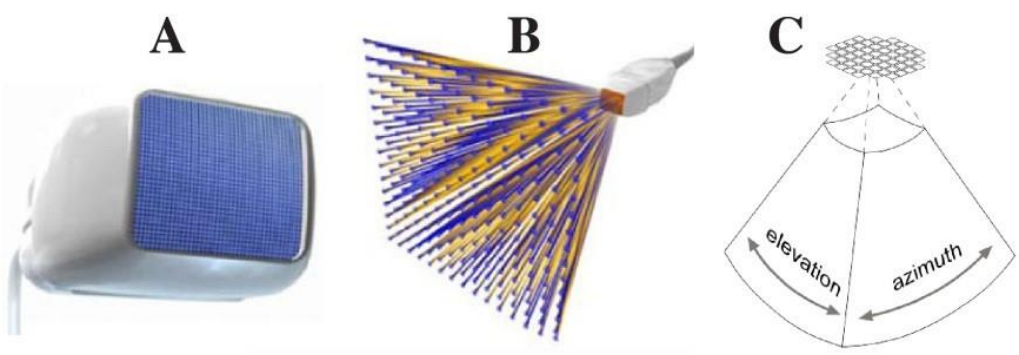

Figure 1.24: A) An example of a 2D matrix array probe. B) The spread of the ultrasound beam and C) the captured 3D volume. Reproduced under open access agreement.

cular events than total plaque area (TPA) or IMT measurement in patients attending vascular prevention clinics. In 2007, the group of Fenster *et al.* [68]

however noted that while TPV provides valuable quantitative information, it was di cult to measure and provided no spatial information regarding where the wall was undergoing changes. They therefore proposed a new index: Vessel Wall Volume, which measures vessel wall thickness and plaque together between the Lumen-Intima and Media-Adventitia. The authors showed that VWV is easier to interpret and more reproducible during measurement. In 2013, the group of Wannarong *et al.* confirmed the findings of Ainsworth *et al.* but also showed that TPV could be used as a predictor of cardiovascular events [45].

Volumetric indices like both TPV and VWV require the segmentation / delineation of the Media-Adventitia boundary (MAB) and the Lumen-Intima boundary (LIB) in the 2D images forming the 3D US volume. These parameters have already been used in a number of research studies and are gaining traction [69; 70; 71]. However, they have not yet gained widespread clinical acceptance due to certain challenges, amongst which are the tediousness and inter/intra-observer variability when manually delineating the LIB and MAB interfaces [63]. Subsequently therefore, a number of other technical research studies sought to address this by proposing automated or semi-automated methods of computing these indices. In 2012, Yang *et al.* [72] investigated the use of Active Shape Models to segment both inner and outer walls of the common carotid artery. The group of Ukwatta *et al.* proposed in 2013 [73] to use sparse field level sets to segment both the LIB and MAB. Similarly, in 2015, Hossain *et al.* [63] proposed to use a distance regularized level set method with various energy terms and a novel stopping criterion to semi-automatically segment the LIB and MAB.

Following the work by Wannarong *et al.*, many other studies have shown the usefulness of measuring carotid plaque burden as a predictor of cardiovascular risk [74; 75; 76; 44]. In 2020, Spence [77] states that carotid plaque burden is highly correlated with coronary calcium scores, and as predictive of risk, while at the same time being less costly and not requiring the use of radiation. When used to monitor patients who were high-risk and with asymptomatic carotid stenoses, it was associated with more than 80% reduction of stroke and myocardial infarction over a period of 2 years [77].

1.6.3 Commercial Platforms measuring Plaque Burden

Most commercial ultrasound platforms include tools for carotid stenosis measurement (spectral doppler ultrasound, colour flow imaging) as standard functions. Many commercial manufacturers also include the option of automated intima-medial thickness measurement. An example is that provided by Siemens

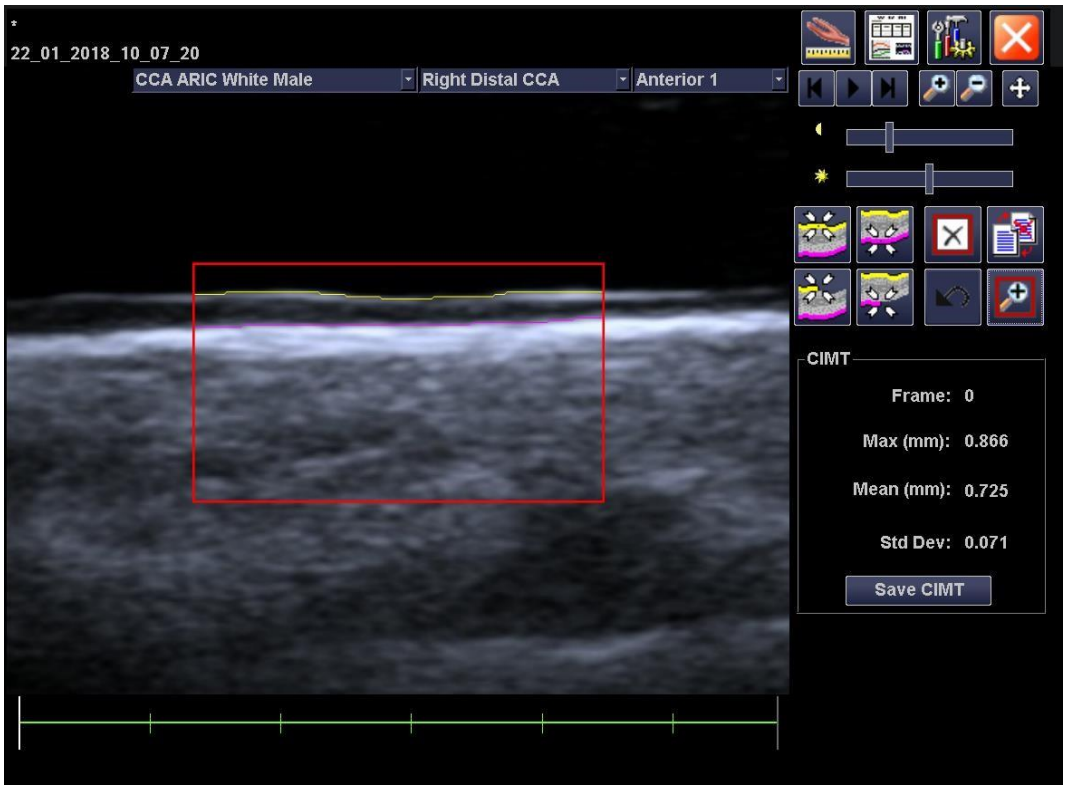

Figure 1.25: Siemens include under their Arterial Health Package automated tools for carotid IMT measurement. Reproduced under open access agreement.

Healthcare under their Arterial Health Package, as shown in Figure 1.25.

As of 2020 however, commercial manufacturer Philips Healthcare systems has also adopted tools for automated plaque burden measurement, under a vascular assessment suite of measurement tools called QLAB. This particular tool, called ”Vascular Plaque Quantification, requires the user to first acquire a three dimensional volume of carotid using a 3D matrix probe. The user is asked to mark two reference frames as beginning and end-frames. While in 2D transverse mode, the user then manually delineates the plaque in at least one

key frame, and the system automatically propagates the segmentation model to subsequent frames, to ultimately generate the estimated overall

Figure 1.26: Philips include a tool specifically for measurement of vascular plaque volume under their QLAB software. Reproduced under open access agreement.

plaque volume detected. An example is shown in Figure 1.26. Philips is the only manufacturer to include tools for plaque volume measurement at present. This however indicates that commercial manufacturers are starting to accept the usefulness of plaque burden as a measurement tool.

1.7 Problem Definition and Analysis

In the United Kingdom, the use of Spectral Doppler ultrasound and Colour Flow imaging remain predominant in diagnosing atherosclerotic lesions following a cerebrovascular event like stroke or TIA [25]. Screening of patients for subclinical atherosclerosis or asymptomatic carotid stenosis does take place. However it is based on a free NHS risk assessment of cohorts who are

considered to be at risk, and involves a physical examination, questionnaire and blood tests by a health professional [78; 79]. No imaging is involved in this screening process. In the United States and in Europe, previous guidelines used to include screening of such patients with carotid IMT, but this is no longer the case [39; 40; 41]. In the US, the 2019 guidelines from the American College of Cardiology and American Heart Association include calcium scoring but not carotid plaque imaging as a risk modifier [80; 81].

In Europe however, the European guidelines on cardiovascular disease prevention [40] list both calcium scoring and carotid plaque assessment using ultrasound (though not IMT) as potential risk modifiers. In addition, many other studies and clinical groups have published their recommendations to include carotid plaque burden as a tool for assessment and screening of atherosclerosis [82; 77; 74; 80; 83]. The practice guidelines of the European Society for Vascular Surgery [27] indicate that there may be potential benefits of selective screening of high-risk patients. Thus, the take-away message indicates that it would be beneficial to have a tool which adequately modifies risk factors of high-risk patients to better stratify their position, and which can be used to assess therapeutic e↵ects on patients undergoing medical therapy for atherosclerosis. In both these instances, volumetric metrics such as TPV or VWV are gaining momentum [80], and this justified the need for the research in this work to study the problem.

Both three-dimensional reconstruction of wall volumes or plaque volumes require the user to manually segment the interface between the lumen and adjacent vascular wall structures. This is a time consuming process, and may be subject to user variability and interpretation across di↵erent

radiographers. A number of studies have sought to develop algorithms to segment these interfaces [84; 73; 72; 63], to reduce processing time and to reduce inter-user variability. The majority of these are however semi-automatic in nature, still requiring some degree of user intervention to initialise anchor points for the technique to work. Many are also subject to strict acquisition protocols on behalf of the end user, assuming that images of a minimum level of quality are available. Thus, there appears to be a clear need for developing techniques for obtaining highly accurate, fully automatic segmentation of vascular wall interfaces, such as the the Media-Adventitia (MA) interface and the Lumen-Intima (LI) interface, in a manner which is robust to changes in end-user acquisition protocols. This would eliminate user intervention during the segmentation process, leaving the role of the user to be that of providing human supervision to the algorithm. Automated segmentation then provides the basis for additional algorithms to reconstruct three dimensional structures or compute indices of plaque burden.

At the time of research and publication of this work (as opposed to the time of writing of this dissertation), the state-of-the-art in terms of research was the work carried out by Hossain *et al.*. In their work, the authors acquired 3D volumetric ultrasound datasets from 10 patients, and from these they extracted 2D image slices. A distance regularised level set method was used to semi-automatically segment the MAB and LIB. In other previous similar works, researchers similarly acquired 3D volumes, but then extracted 2D slices [84; 72; 73; 85] and applied segmentation strategies on these 2D images, before reconstructing the segmented result in three dimensions. On the commercial front, the Vascular Plaque Quantification software o↵ered by Philips, also acquires the data in 3D, but has the user initialise the segmentation model in

two dimensions on key 2D frames before extending this through the volume. No information is available on the implementation methodology used by Philips, which is the only manufacturer to include plaque quantification. However it is reasonable to assume that the segmentation algorithm also operates natively in 2D on the individual 2D slices making up the volume. On the back of this, and given that the hardware available during this project included only 2D acquisition probes, a native 2D acquisition strategy was chosen in order to develop and refine the segmentation algorithm. Position tracking and the subsequent 3D volume reconstruction was tested as a proof of concept at later stages of the research.

Furthermore, at the time of research, the use of artificial intelligence and deep networks in the field of image segmentation had garnered much interest, with the number of publications related to medical image segmentation rising considerably. As at 2017 however, the survey by Litjens *et al.* [86] cited only 20 contributions to the field of ultrasound imaging, with only 6 of these being related to cardiac / vascular applications. The problem of segmenting both the MAB and LIB in a fully automated manner and with good accuracy remained unsolved at the time. With deep networks o↵ering improved performance over classical methods in many applications, and with their ability to be used in a fully automated manner, this drove the decision for this work to study their application to segmenting the MAB and LIB in carotid ultrasounds.

At the time of writing of this dissertation, the author retrospectively notes that, following the second publication of this work concerning the use of deep neural networks for automated carotid segmentation, a number of additional studies by other research groups followed suit in using di↵erent variants of

deep neural networks for the problem of carotid artery segmentation [87; 88; 89]. These studies are described and compared later in the thesis. Additionally, the author also notes that as of 2019, commercial manufacturers such as General Electric (GE Healthcare) have also adopted the use of deep neural networks for segmentation tasks within ultrasound imaging, although these are presently intended for generic lesion segmentation. The latter, together with the additional research studies mentioned, continue to further justify the relevance of the work in this dissertation.

1.8 Research Objectives

The overarching aim of this project is to devise a novel, ultrasound segmentation technique, which permits accurate and robust delineation of the MediaAdventitia interface and Lumen-Intima interface for the eventual purpose of computation of vessel wall volume. It is desirable that the proposed technique be: highly accurate, and thus obtaining good agreement with delineations obtained manually; fully automated, and therefore requiring no user intervention during the segmentation process; and robust, thus providing a degree of resilience against variability in some basic user settings (such as gain).

In more quantified terms therefore, the general research objectives forming part of the projects aims shall be as follows:

1. Design a novel technique, which presents an amplitude-invariant image representation, which still contains anatomical structural information that lends itself towards segmentation.

2. Design a novel ultrasound segmentation technique, which is able to segment vascular interfaces / borders in a fully automated manner without user intervention.

3. Generalise the technique proposed in point number 2 further, such that this may be implemented to provide automatic segmentation of both MA and LI interfaces.

4. Validate the proposed technique developed by measuring Vessel Wall Volume, and compare its performance against recent studies.

1.9 Project Contributions

This project produced the following contributions, which fulfil the research objectives set out in section 1.3 above. The contributions are in the form of novel developed algorithms or systems:

1. **Radio Frequency derived Phase Congruency Maps of Carotid Ultrasound images**

   This algorithm proposes a novel, intensity-invariant representation of carotid ultrasound data, which may be used for segmentation, through the creation of phase congruency maps from raw radio-frequency ultrasound data. The novelty of the algorithm stems from being the first to obtain phase congruency maps from RF-based carotid ultrasound data. Phase congruency maps are arrays of phase information, which are invariant to amplitude / gain settings, and which still provide structural information about the underlying anatomy. The phase congruency maps

are used in conjunction with amplitude information for subsequent segmentation.

2. **Carotid Artery Segmentation using Deep Convolutional Networks and Phase Congruency Maps**

   This algorithm proposes a novel segmentation technique, which delineates the outer Media-Adventitia layer of carotid arteries in ultrasound images in a fully automated manner, without requiring setting of seed points or other forms of user-intervention. The novelty of the algorithm stems from being the first to apply deep convolutional networks (DCN) to transverse carotid ultrasound images, and also from the DCN being applied to a novel fusion of input data, which combines amplitude phase congruency information.

3. **Bimodal Automated Carotid Ultrasound Segmentation using Geometrically Constrained Convolutional Neural Networks**

   This algorithm implements a novel deep convolutional U-NET to segment both the outer Media-Adventitia interface and inner Lumen-Intima interface of carotid arteries in ultrasound images. The main novelty of the proposed algorithm is in the modification of the network's objective function, which integrates a-priori knowledge regarding the carotid structures in the form of geometric constraints. The network is applied to the previously proposed fusion of data, which combines amplitude data and phase congruency maps.

1.10 Structure of the Dissertation

The structure of the dissertation generally follows the project methodology in terms of the topic of discussion of the various chapters. This first chapter has given a detailed introduction to put atherosclerosis in context of the burden it causes through cardiovascular disease and cerebrovascular disease. It has briefly described clinical pathways for its diagnosis and treatment, as well as the shortcomings related to present methods monitoring its progression. It has then briefly described the physics of ultrasound imaging and 3D quantification of atherosclerotic plaque, and has provided a critique to define the problem studied in this work. Consequently, it has laid out the main objectives and contributions of this project.

The second chapter provides a detailed review of the technical literature pertaining to the project. It provides a comprehensive review of medical image segmentation methods, and a description of deep neural networks. It provides some background on deep neural networks and also gives an overview of different network architectures used in literature for medical image segmentation problems. This is followed by a discussion and justification for the proposed methodology.

The third chapter discusses initially the data acquisition process and the raw radio-frequency format of data acquired from the ultrasound machine. It also describes the mathematics behind obtaining a typical Brightness mode image based on amplitude information. It then proceeds to propose the novel phase-based and amplitude-invariant representation of the carotid data. Two alternatives are introduced: Phase di↵erence matrices and Phase congruency maps. The underlying theory is explained and a set of experiments are then

described, which compare their performance in producing a segmentation of the outer media-adventitia interface.

The fourth chapter proposes the novel application of DCNs to a fusion of amplitude and phase congruency carotid ultrasound data. It then proceeds to describe the experiments carried out to identify the ideal DCN structure used to segment the MA interface from amplitude and phase information. The results obtained from the experiments are then provided and their significance is discussed.

Chapter five describes the proposed novel objective function used within an expanded U-NET architecture to segment both the MA and LI interfaces in the carotid ultrasound images. It provides the physiological basis behind the proposed geometric constraints / penalties applied to the objective function used, as well as a mathematical derivation behind the said geometrically constrained penalties. A set of experiments are described, which test and compare the performance of the proposed modifications versus the original unmodified deep networks. The results obtained are presented and their significance discussed.

The sixth chapter provides a proof of concept for a) utilising the developed algorithms to segment the MA and LI interfaces in a short segment of carotid artery b) reconstructing a three-dimensional model of the said carotid artery and c) quantifying the vessel wall volume in this segment. The performance in quantifying the vessel wall volume is compared against techniques proposed in recent literature.

Chapter seven will finally summarise the main results obtained from this study. It will highlight the main contributions provided to the body of knowledge concerning ultrasound imaging. It will also provide some concluding thoughts, describe some limitations present for the project and some suggestions for future avenues of research.

1.11 List of Publications

1. C. Azzopardi, K. P. Camilleri and Y. A. Hicks, ”Carotid ultrasound segmentation using radio-frequency derived phase information and gabor filters,” 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Milan, 2015, pp. 6338-6341. [90].

2. C. Azzopardi, Y. A. Hicks and K. P. Camilleri, ”Automatic Carotid ultrasound segmentation using deep Convolutional Neural Networks and phase congruency maps,” 2017 IEEE 14th Intern. Symposium on Biomedical Imaging (ISBI 2017), Melbourne, VIC, 2017, pp. 624-628. [91].

3. C. Azzopardi, Y. A. Hicks and K. P. Camilleri, ”A Bimodal Automated Carotid Ultrasound Segmentation using Geometrically Constrained Convolutional Neural Networks,” April 2020, IEEE Journal of Biomedical and Health Informatics, vol. 24, no. 4, pp. 1004-10105. [92].

2

Technical Literature Review

2.1 Introduction

This chapter shall provide a brief and broad overview of the leading medical imaging segmentation techniques available in literature, followed by a more focused review of image segmentation techniques that have been applied to the problem of carotid ultrasound imaging. A discussion, which justifies the choice of methodologies for the problem of segmenting the MAB and LIB interfaces will follow. The basic theory underlying DCNs is then described, along with the various general neural network structures used in literature for object detection, classification and semantic segmentation problems. The chapter concludes with a discussion which justifies the choice of the proposed Fully Convolutional Network (FCN) used for semantic segmentation application to carotid images.

2.1.1 General Medical Image Segmentation Methods

The purpose of medical image segmentation is to extract regions that constitute a meaningful part of an object within an image, in order to identify

that part or to analyse it further. A universally accepted method of image segmentation is not available to date, due to the large variety of data available and the individual issues arising with each. Variation in intensity, non-homogeneity of images, di↵erent spatial characteristics and textures, and lack of distinction between grey levels, are only but a few of the many issues that need to be addressed by image segmentation techniques.

Segmentation techniques may be broadly categorised into five groups as shown in Figure 2.15. These categories are: 1. Thresholding-based methods; 2. Region-based methods; 3. Edge-based methods; 4. Clustering-based methods; and 5. Miscellaneous methods based on principles not falling under the other categories. The following sections briefly describe the principal methods within each category.

Thresholding-based methods

Grey level thresholding methods are one of the most widely used and quickest segmentation techniques. They are based on the principle that images are made up of multiple grey level regions, and that based on a histogram examination, the objects within may be classified according to the histogram peaks and valleys [93; 94]. The thresholding method globally di↵erentiates the image into a foreground area, where pixel intensities equal or exceed a threshold value, and a background area, which encompasses pixels with intensities below this value. If the choice of threshold value is incorrect, this method leads to poor segmentation results. Global thresholding methods do not take into ac-

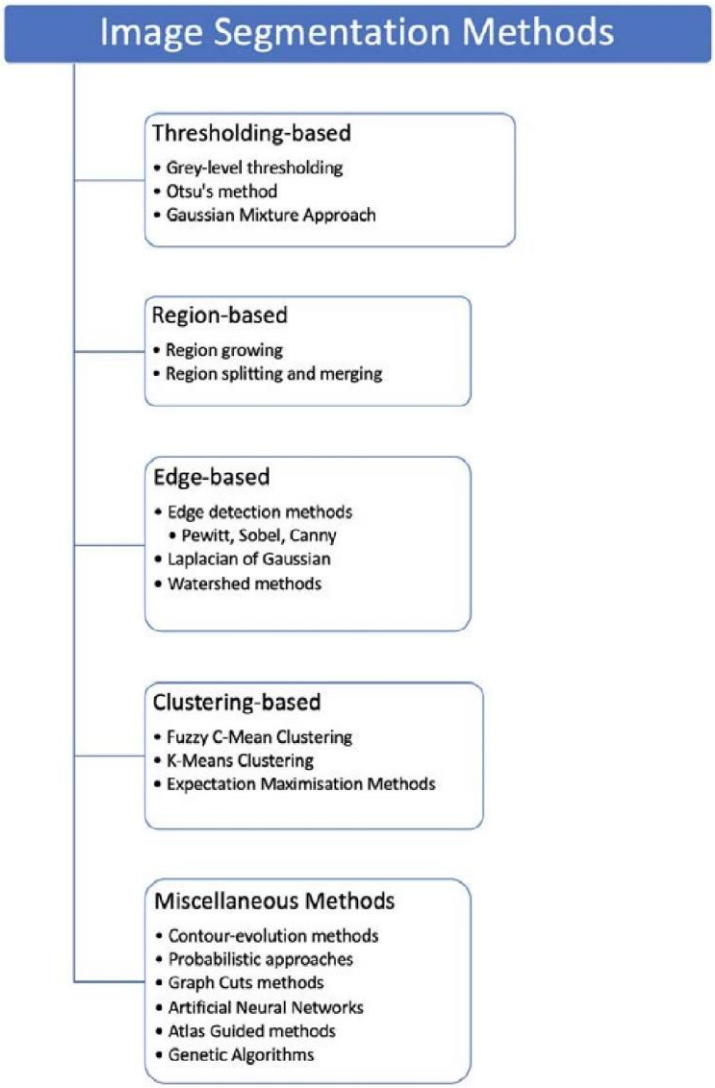

Figure 2.1: Di↵erent categories of image segmentation methods.

count spatial information, and this may also lead to incorrect results. A local thresholding technique improves on this by dividing the image into multiple sub-regions based on some spatial relationship, and applies thresholding within each region [93; 95].

Otsu's method works similarly to a global thresholding technique, however, it determines the optimal thresholding value by considering inter-class and intraclass variation within an image [96]. In order to minimise intra-class variance or maximise inter-class variance, the image is considered as a bimodal source of only 2 classes. Initially, the intra-class variance is obtained by calculating the weighted sum of variances of each cluster, followed by determination of the mean value of each cluster. Subsequently, the individual class variance is calculated. The optimal threshold is the value which maximises the inter-class variation [97]. In the Gaussian Mixture Approach, the posterior probability and maximum likelihood is calculated for a number of components in an image. The mean, covariance and mixing coe cients are also computed sequentially, without requiring any initialisation [98]. Then, from a single seed mixture component that covers the whole dataset, the data is sequentially and incrementally split during expectation maximisation steps. The technique shows its e↵ectiveness after several runs [98].

Regions-based methods

Region growing methods seek to extract regions from an image based on a set of specific characteristics. Typically, prior information must be provided, in the form of a seed pixel selected by the operator [99]. The algorithm then iteratively grows the region to encompass pixels around the seed, which share a common characteristic, until an edge is detected. Occasionally, region growing methods may be sensitive to noise, which ends up with mergers of separate regions or holes in regions that should not be there [100].

Region splitting and merging techniques operate by splitting an image into multiple sub-images or regions, and then merge them back together according

to regions with similar characteristics. The approach uses a quad-tree generation which produces four branches. The branches of the quad-tree represent the sub-images, which are re-merged into regions until no further partitioning or splitting is possible [101].

Edge-based methods

Edge-based segmentation techniques function by identifying and locating edges and borders, by seeking out sharp discontinuities in the intensity values of an image. The edges are detected using masks or filters, which are superimposed over the image to detect the discontinuities. The filters used may be first order filters, such as the Prewitt operator, the Sobel operator or the Canny operator. The Prewitt operator was proposed by Judith Prewitt [102], and is a gradient based operator which determines the magnitude and one of eight directions of an edge by calculating the gradient of image intensity. The Sobel mask uses discrete di↵erentiation to determine vertical and horizontal orientations [103]. An example is shown in Figure 2.2. It has good noise suppression characteristics, which makes it popular. The Canny operator was developed by John Canny, and operates a multi-stage edge detector which can be used to detect various edges and orientations [104].

The Laplacian of Gaussian operator is a second order operator, proposed by Marr and Hildreth in [105]. It is a di↵erential and adaptive operator, based on a combination of Laplacian and Gaussian techniques, and used to detect blurry edges and sharply focussed fine details. It can thus be adapted according to the application area. The Watershed segmentation algorithm, proposed by Beucler and Mayor in [106], is used to separate overlapping objects and to address the issue of over-segmentation in natural images in image processing. It uses the

watershed transform together with homotrophy modification, and is applied on grey level images. It benefits from being fast, simple and intuitive.

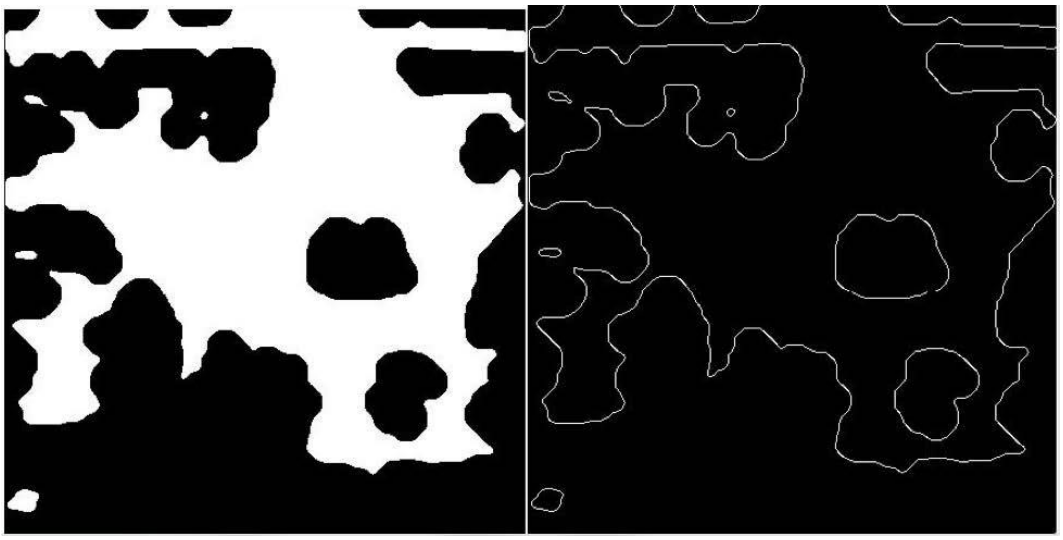

Figure 2.2: An example of sobel edge detection used to extract edges of cartorid artery.

Clustering-based methods

Clustering approaches employ a technique, which groups objects (or pixels) sharing similar characteristics and properties, to form group classes called clusters. For an e↵ective performance, the objects should as much as possible, have a great similarity to one type of cluster, but be highly dissimilar to other clusters – since the objective of the algorithm is to maximise intra-class similarity and to minimise interclass similarity [107]. The clustering algorithms are usually unsupervised in nature, and learn through several iterations on the training data. The K-means algorithm is one of the simplest and most popular types of clustering techniques. It classifies the 'n' datasets into k-clusters iteratively [108]. The mean intensity is obtained for each

cluster, and then pixels having the closest value to the mean are classified accordingly.

The Fuzzy C-Means algorithm is a generalised version of the k-means algorithm, with the main di↵erence being that the data is classified into clusters based on the degree of membership of an object [109]. The algorithm is based on Fuzzy Set theory and allows for soft-segmentation. This makes it popular in segmentation tasks due to it taking into account the uncertainty and vagueness of certain data. The Expectation Maximization (EM) algorithm is also an unsupervised method that is based on a Gaussian Mixture Model. The technique involves iterating several times to calculate posterior probabilities and maximum likelihood values for special parameters [110]. The parameters are mean, covariance, and mixing coe cients of the mixture model. The Gaussian Mixture Model has a reduced sensitivity over K-means and Fuzzy C-means algorithms.

Miscellaneous methods

Contour-evolution methods seek to initialise a contour and evolve it to some desired boundary through an energy minimisation approach. Level set methods (LSMs), initially proposed by Osher and Sethian [111], involve the use of level sets for numerical analysis of surfaces and shapes, and are particularly useful for tracking of evolving contours. Although the original method does not involve energy minimization [112], later implementations of the technique [113] apply the level set formulation as a contour energy minimisation problem, using region-based image features, with or without edge-based features to construct the energy. LSMs can be performed on curves and surfaces on a fixed Cartesian grid, and do not require us to parametrise the

objects being segmented. Somewhat similar to LSM techniques are those based on Active Contours (Snakes), which seek to initialise a seed contour and then actively deform this contour to the desired border by iteratively reducing a defined energy function until convergence is achieved. The energy function may also, in this case, be defined on region-based or edge-based features, as well as a priori knowledge to constrain the deformation process.

In image segmentation, probabilistic approaches seek to use probabilistic models to assign the most likely label to a pixel in an image, given a particular set of features. A popular method is to model label dependencies using a Markov Random Field (MRF) and to determine the optimal labelling by Bayesian estimation, in particular, maximum a posteriori (MAP) estimation [114]. One of the main advantages of MRF models is that prior information can be applied and imposed locally through the use of 'clique' (subset of graph vertices) potentials. The algorithm for MAP initially defines a neighbourhood of each feature and sets the initial probabilities. The class statistics for each label are computed from some available training data. Using Bayes theorem, the probability of each class label is computed, given a certain neighbourhood, and the process is repeated iteratively using optimisation algorithms, until such probabilities are maximised [114].

The Graph Cut technique is an e cient manner in which to minimise a larger class of energy functions, which correspond to the maximum a posteriori solution. The optimal or minimal cut in a graph is the subset of edges, that if removed, divide the graph into two regions: foreground and background [116]. A schematic diagram is shown in Figure 2.3. The cost of the cut which must be minimised, is computed as the sum of weights of the cut edges, and

the minimal cut is that with the minimal sum of weights from all possible cuts. The scoring inside the graph is applied as follows: pixels inside the object are given values depending on how well their intensity values match the object's appearance model, with low values signifying better matches. Similarly, pixels in the background are given values depending on how well their intensity matches the background's appearance model [117]. Adjacent pixels on the

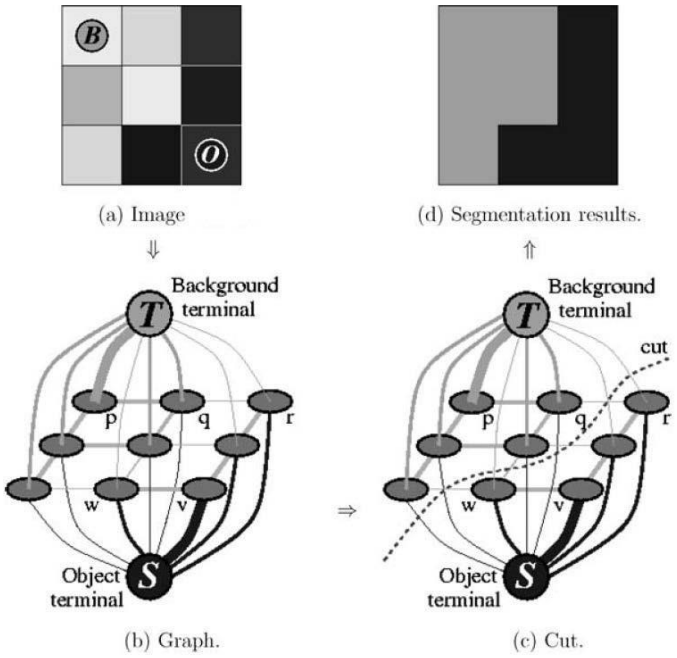

Figure 2.3: An example of image segmentation using graph cuts. Reproduced with permission from [115].

edge of the object are given values according to whether both have similar intensities. A low value corresponds to contrasting intensities [117].

Atlas-guided approaches operate on the basis that human organs vary significantly in shape and size, and that this variability makes it di cult to have a technique which can extract a good general representation. To mitigate this,

techniques based on atlases have been devised, which create a standard 'template' to model the organ in question. An example is shown in Figure 2.4. These atlases contain important information, such as local image statistics and probability of assigning particular labels to particular spatial locations. The new images may then be easily mapped onto the new atlases through atlasbased segmentation. Atlas-based approaches are divided in two: parametric and non-parametric methods. In the former, the new images are combined

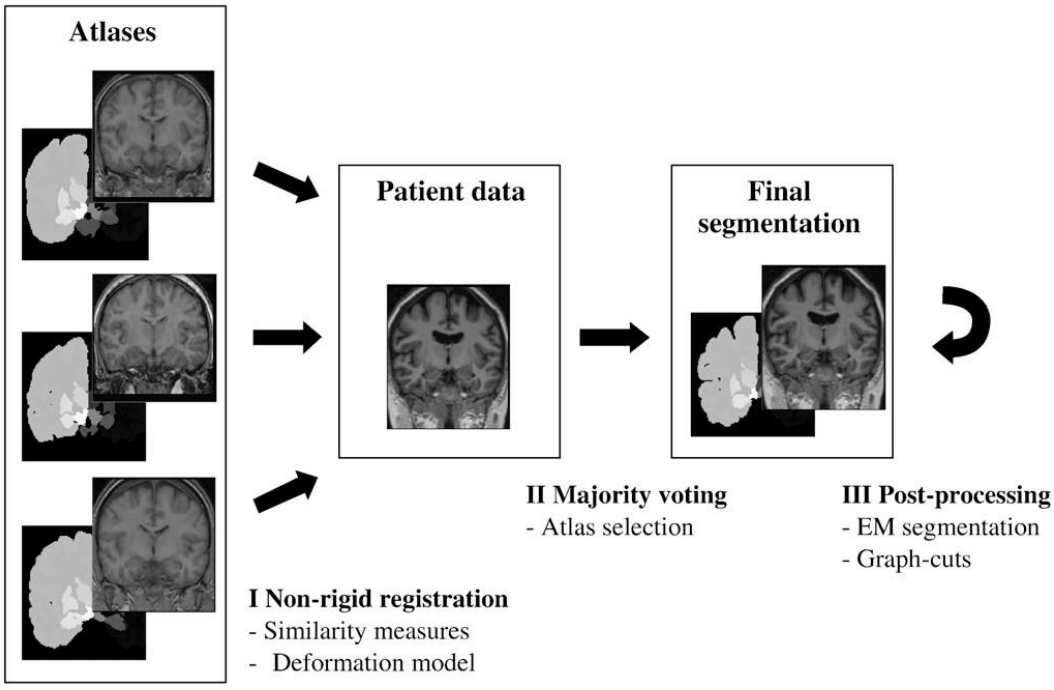

Figure 2.4: An example block diagram of a typical atlas guided segmentation process. Reproduced with permission from [118].

with the trained images to create a single atlas. In the latter, all images are used separately for training purposes.

Genetic algorithms (GA) apply an optimisation technique based on evolutionary biological sciences. The technique employs three steps: mutation, crossover and selection. An initial population of solutions is first considered and evaluated for fitness of each individual in the population. The best individual solutions are combined to produce 'o↵spring' with better characteristics. The mutations resulting from the combination generate a set of heuristic solutions from the population. The process is repeated, until the most optimal solution is found. Most GAs are based on ant-colony optimisation methods or particleswarm optimisation methods.

Artificial Neural Networks (ANNs) are computational systems that mimic the way humans learn. Networks are made up of nodes, and the learning takes place by updating the weights on such nodes during an iterative process that compares an initially produced network output to a desired target output. ANNs find application in a wide variety of areas. Recently Deep structures of ANNs have been proposed to solve more complex medical image processing tasks, including segmentation, and results obtained have outperformed many of the other classical segmentation techniques. A more detailed review of ANNs is provided in Chapter 4.

2.1.2 Ultrasound Segmentation of the Carotid Artery

With Ultrasound imaging becoming more and more popular, it is natural that a body of research has focused also on applying medical imaging segmentation techniques to this modality, to segment di↵erent organs. Notable review papers which broadly discuss this topic are covered in [119; 120]. This section focuses specifically on studies which concern themselves with the segmentation of the carotid artery with ultrasound imaging. Previous

literature shows certain patterns in the development of carotid segmentation techniques, allowing us to categorise them in their nature. The first division is that of addressing the segmentation problem on longitudinal or transverse sections of the carotid [119]. The majority of studies available address the former [84], since this type of segmentation then easily lends itself to evaluating the intima-media thickness (IMT) - a widely accepted clinical parameter, which is used in gauging degree of stenosis [121]. A comprehensive review paper by Molinari *et al.* [121] addresses some of the major works on the subject matter, whilst concurrently grouping them according to their similarity.

Segmentation in Longitudinal Images

A popular approach on the longitudinal carotid segmentation problem was through using edge-tracking or gradient-based techniques, as shown in a number of studies [122; 123; 124; 125; 126]. In the longitudinal section, the common carotid artery may be considered as a dark region surrounded by two bright line patters - the near wall and the far wall of the artery [121]. By considering the intensity profile, or the intensity gradient across a section cutting across the artery, the adventitial walls may be clearly identified, and the IMT estimate may be obtained as the distance between these two points [122; 123; 124; 125]. Faita *et al.* [126] improve on the method proposed by Ligouri [124] and Stein [125], by applying a first-order absolute moment edge operator on the intensity gradient profile, and by applying a pattern recognition approach, in order to mitigate the problem of superimposed noise experienced by the previous two techniques.

Another widely used segmentation approach is based on Active Contours, or Snakes. This entails having a set of vertices connected by line segments [121], which dynamically move to settle around the desired contour, under the action of defined forces. Molinari *et al.* [121] note in their review that at least 6 di↵erent contributions adopted some variation on this approach [127; 128; 129; 130; 131], with these including the use of multi-resolution analysis, novel damping forces or modifications on the external energy, amongst others. Snake models, however, have issues that e↵ect their performance. They require correct fine tuning of parameters for them to be correctly attracted to edges; they depend on the initialisation of the snake model, and they are also prone to leaking through edges which are not clearly defined [121].

Other segmentation approaches for longitudinal sections include the use of Dynamic Programming [132; 133], Nakagami modelling [134], the use of the Hough Transform [135], and the use of motion estimation and bayesian frameworks [136].

Segmentation in Transverse Images

A number of studies have also addressed the problem of segmenting carotid arteries in the transverse section, using either native 2D images or else from transverse slices extracted from 3DUS images. In 2009, Seabra *et al.* [137] proposed a semi-automatic technique for plaque segmentation in transverse images. In this method, the contour of the plaque was manually initialised in the first image, and then a 2D active contour algorithm was allowed to carry forward the segmentation in subsequent temporal images. Another study by Yang *et al.* [72] proposed to use active shape models to segment the MA and LI interfaces. Ukwatta *et al.* [84] proposed a novel semi-automated technique

based on a level-set method to segment the MA and LI interfaces. The operator was asked to provide anchor points as high-level domain knowledge, and this together with the incorporation of local and global image statistics with boundary separation-based constraints allowed accurate segmentation of the MA and LI interfaces. In 2007, Guerrero *et al.* [138] proposed to use a modified star-kalman algorithm to determine and track vessel contours. Other segmentation approaches for transverse sections included the use of deformable models [139; 140], modified Cohen Snakes [130] and a star algorithm improved by Kalman filtering [141].

Alternative methods have been proposed to segment the carotid structures or plaque morphology natively in three dimensions. Gill *et al.* [142] proposed a semi-automatic method, based on a dynamic balloon model in 2000. The model was represented by a triangular mesh, which is initialised within the carotid vessel by the user. The balloon is inflated to obtain the approximate LI interface, and refined using image edge-based forces. In 2010, Solovey *et al.* [143] also proposed an LI interface segmentation algorithm on native 3D images. They proposed a level set-based method, which incorporated a region term that minimises the probability distribution overlap and a weak geometric prior, which encourages convexity of the boundary in a level set framework. In 2015, Hossain *et al.* [63] presented a semiautomatic method for segmenting both MA and LI interfaces using a distance regularised level set algorithm, with a novel stopping criterion and a modified energy function.

One notes further that, particularly in the case of transverse segmentation, studies have aimed to segment either the LI interface alone [139; 140; 138; 143; 142; 144], or else both the LI and the MA interfaces [84; 73; 72; 63]. The

latter approach has increasingly gained interest due to new volumetric parameters, such as vessel wall volume (VWV) and total plaque volume (TPV), which have been proposed to characterise plaque burden [145]. Although these parameters have been used in a number of research studies, they have not yet gained widespread clinical acceptance due to certain challenges, amongst which are the tediousness and inter/intra-observer variability when manually delineating the LI and MA interfaces [63]. Studies have therefore sought to develop automatic or semi-automatic algorithms to segment these interfaces

Deep Networks in Medical Image Segmentation

Deep networks have recently garnered much interest, as they have driven forward the state-of-the-art in computer vision tasks, such as image classification, object detection and segmentation[146]. Such advancements have also been picked up by the medical imaging research community. The survey by Litjens *et al.* in 2017 [86] provides a comprehensive review of studies employing Deep networks for a variety of tasks and application areas within medical imaging. In their work, Litjens *et al.* assessed over 300 contributions since 2012, with approximately 20 of these being contributions within the ultrasound imaging field and in turn, 6 of these being related to cardiac / vascular applications. None of these, however, treat the subject of carotid segmentation. The study by Menchon-Lara *et al.* in [147] addresses longitudinal carotid ultrasound segmentation using a single layer perceptron network preceded by an autoencoder, for the purpose of intima-media thickness (IMT) estimation. Tajbakhsh *et al.* [148] also segment longitudinal

carotid ultrasound images using a convolutional neural network (CNN) to estimate the IMT.

In the author's own previous preliminary work in 2017 [91], the MA interface is segmented in transverse and longitudinal carotid images using deep convolutional neural networks. The various network configurations are evaluated to find the optimal network size and depth, as well as the optimal filter dimension. The author further propose a novel fusion of amplitude and phase congruency data as an input to the network, as the latter provides an intensity-invariant data source to the network.

In 2019, the studies by Zhou *et al.* [87; 88] seek to use Deep Learning to segment the MA and LI interfaces in transverse carotid ultrasound images. In [87], the authors use an unsupervised deep convolutional encoder-decoder structure to pre-train the parameters of a U-NET. The U-NET is then used to automatically segment the MA interface. In [88], Zhou *et al.* use a semiautomated approach for segmenting the MAB, whereby the user provides a set of anchor points to a dynamic CNN which then segments pixels along the norm line of the MAB contour. A modified U-NET is then used to automatically segment the LI contour. At the time of writing, the same group of Zhou *et al.* have also extended their work with CNNs to three dimensional voxel based data, in order to segment the carotid vessel wall volume natively in the volumetric space [89].

2.1.3 Choice of Segmentation Methodology

In this work, the choice of methodology was influenced by the prerequisite objectives of this project, namely that of having a fully automated

segmentation technique, which yields good performance and is insensitive to changes in amplitude from the user. The image segmentation techniques that have been considered in section 2.4.1, in particular those falling under thresholding, region-based, edge-based, or clustering-based methods – have a tendency of being susceptible to artefacts and noise, which come about from pixel intensity. Ultrasound imaging in particular is highly prone to pixel intensity artefacts, causing seemingly high (hyperechoic) or low (hypoechoic) pixel intensities due to shadowing, speckle, or indeed material change such as tissue calcification. Such artefacts would not form part of the organ of interest which requires segmentation, yet may cause these techniques to incorrectly delineate objects which are not there, or result in non-closed boundaries.

Popular contour-evolving methods, such as level-set and Active Contours, or graph cuts methods and genetic algorithms, are also heavily dependent on region-based or edge-based characteristics and thus susceptible to similar problems. They also often require initialisation from the end user to place seed points, thus precluding them as a fully automated choice. Additionally, graph cuts are only able to find a global optimum for a binary labelling problem with two classes, whereas the problem in this work involves labelling three classes. Atlas-guided methods o↵er the interesting potential of being merged with other methodologies, through the provision of priori knowledge concerning a model of the organ anatomy that needs to be segmented. However, if taken in isolation, they will most often require an image registration approach, which also commonly requires seeding with anchor points.

Neural networks o↵er the advantage of mimicking the human visual perception, which goes beyond simply considering a pixel's intensity and

presence or lack thereof an edge. They also inherently incorporate prior knowledge in the form of the training data, which is used to train the network model in a supervised manner, to produce a certain output, given a certain input. The networks can be trained end-to-end, and once a model is obtained, it can be applied in a fully automated manner to new sample data without requiring any further user intervention, initialisation or seeding. In the case of convolutional neural networks, the model is also spatially invariant. Neural networks also easily lend themselves to processing multi-stream data, thus permitting di↵erent perspectives and imaging modalities of the same object to be considered simultaneously. In this case, this allowed simultaneous consideration of both amplitude and phase information obtained from the ultrasound data, in order to factor in an amplitude invariant component (phase), to create a more robust model. Finally, the e↵ectiveness of neural networks has dramatically increased with the availability of improved computing power, allowing for more complex and deep networks, which are able to better model the data in the problem. The compelling evidence of performance of deep neural networks exceeding that of other segmentation strategies, is what ultimately guided this work towards choosing the latter for this problem.

2.2 Basic Neural Networks

Deep network architectures are based on the underlying principles of neural networks, and may thus be perceived as the generalised form of a linear or logistic regression problem. A detailed description of the theory behind neural networks is beyond the scope of this section, for which the reader is advised to consult appropriate literature such as LeCun *et al.* [149; 150; 151], Gu *et al.* [152]. However, the concepts behind such networks may be briefly explained

by considering initially a simplistic model of a feedforward perceptron neural network as shown in Figure 2.5. In such a network, the activation, *a*, of each neuron is formed through the linear combination of an input vector **x** and a set of learned coe cients, $W_{h,i}$ and $W_{h,0}$. This combination is then followed by the application of an element-wise non-linearity (.): Thus:

$$a_h = \sigma(\sum_{i=1}^{I} W_{h,i}x_i + W_{h,0}) \quad (2.1)$$

Such a network would also consist of one or more layers, *L*, of neurons stacked on each other, through which a signal is propagated. Thus, a *multi-layered* perceptron (MLP) is a variant of such a network, in which more than one feedforward layer is stacked on top of each other, causing the intermediate layers to be referred to as *hidden* layers. In contrast, the input and output layers at either end are observable layers. A *deep* network architecture is one which has many such hidden layers. With the feed-forward functions established, the network is trained such that it learns to produce the correct outputs when presented with a set of inputs. The training process is iterative, whereby the network is repeatedly presented with an example set of inputs, allowed to produce an output, and then this output is compared against the desired

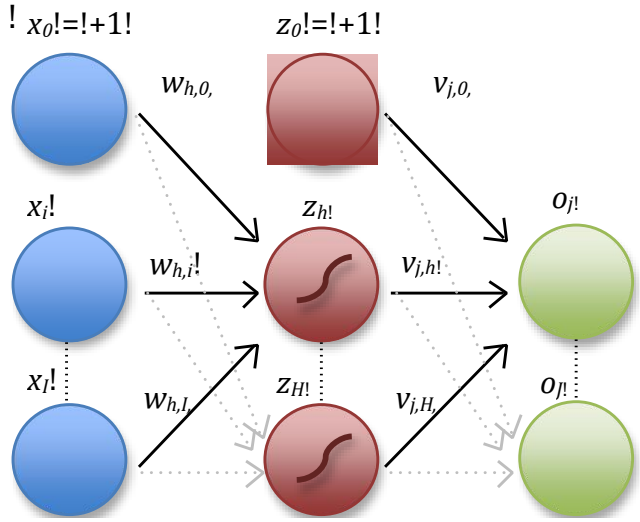

Figure 2.5: Simplistic Feedforward Neural Network with one hidden layer.

output. The di↵erence between the actual output and the desired output is used to adjust the weights in the network described previously, such that a better output is generated during the next iteration. After several iterations the network's actual output should converge towards the desired output, and this should ideally also carry on for new sets of inputs which the network would not have been exposed to before. This algorithm, which is used to train the network, is called the Backpropagation Algorithm.

2.3 Di erent Network Architectures

Deep network architectures were traditionally di cult to train e ciently. That was until in 2006, when research groups including that of Hinton *et al.* [153; 154], [155] showed that excellent pattern recognition performance could be obtained, when such deep structures would be greedily trained in an unsupervised manner, using a layer-by-layer approach, and then fine-tuned

with supervised training. In the following, two popular architectures are described which are trained in this manner, along with two additional popular alternatives which train the entire network in a direct end-to-end fashion using supervised training:

2.3.1 Auto-Encoders and Stacked Auto-Encoders

Autoencoders (AEs), [156; 157; 158] are constructed out of relatively simple structures, intended to reproduce the input **x** onto the output layer **x'** using only one hidden layer **h**. The weight coe cients producing the hidden layer and the output layers respectively are $\mathbf{W}_{x,h}$, $\mathbf{W}_{h,x0}$ and biases $b_{x,h}$, $b_{h,x0}$. In the scenario where the hidden layer had the same size as the input layer, and was constructed with no non-linearities, the AE would simply learn the identity function. The most important feature of such a network however, is the use of a non-linear activation function to compute the latent representation in **h**, that is smaller than **x**, thus projecting the input data onto a lower dimensional subspace.

$$h = \quad (\mathbf{W}_{x,h}\mathbf{x} + b_{x,h}) \tag{2.2}$$

In the reconstruction process, the matrix $\mathbf{W}_{h,x}$ is often taken as $\mathbf{W}^{T}_{x,h}$. A schematic diagram of the AE is shown in Figure 2.6a. An alternative, denoising AE, is that proposed by Vincent *et al.* [159], whereby the model is trained to reconstruct the input from a version which is corrupted by noise. *Stacked* autoencoders (SAEs) on the other hand are deep AEs, which are formed by stacking several auto-encoder layers on top of each other [86]. The individual layers are typically trained individually (*greedily*), after which the network is then fine-tuned to the problem at hand by training it in a supervised manner.

This entails providing it an input and a correspondingly labelled output to which the network must conform.

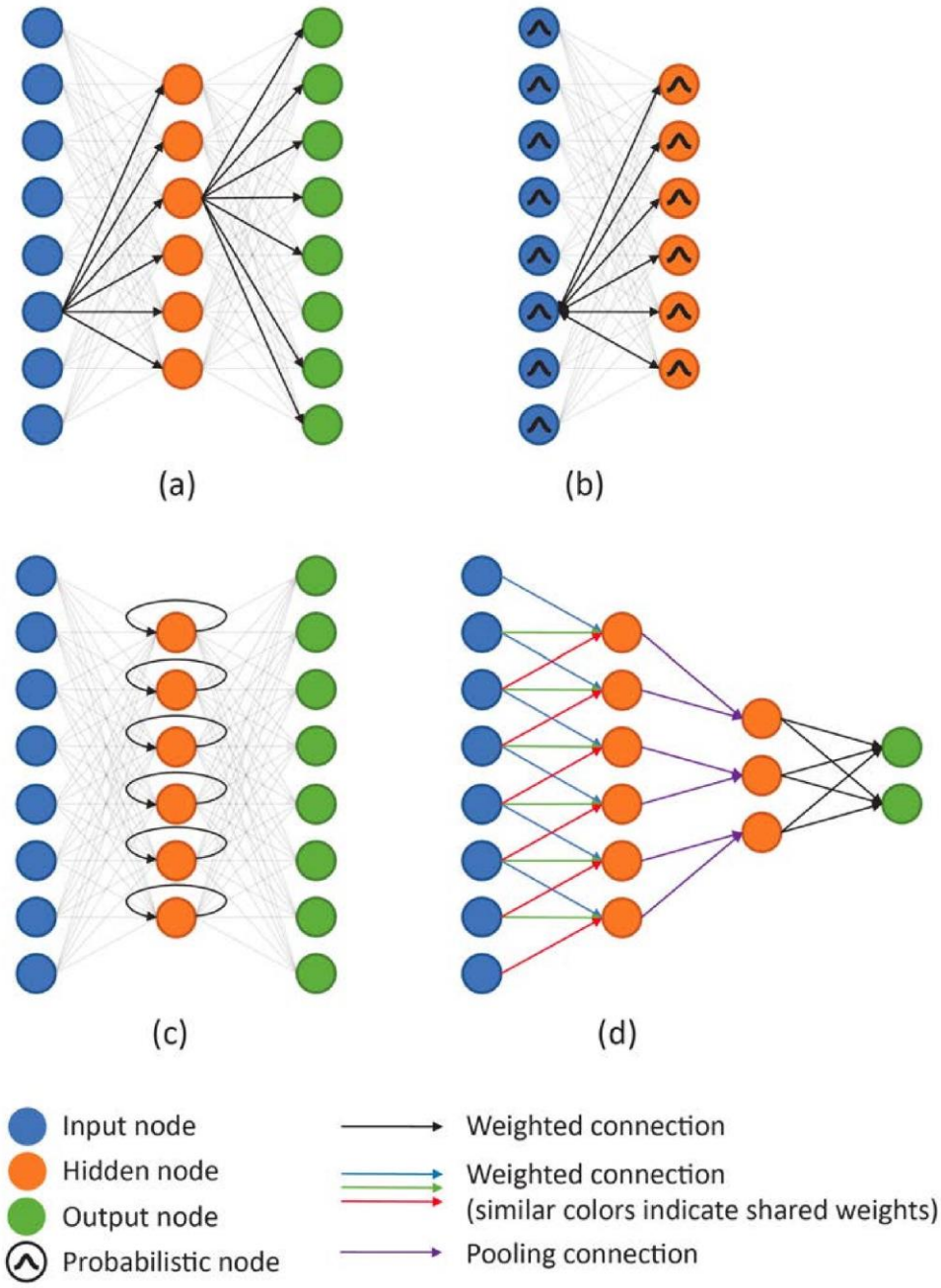

Figure 2.6: Various forms of deep architectures. a) Autoencoders; b) Restricted Boltzmann machines c) Recurrent networks; d) Convolutional Neural Networks. Reproduced under open-acess agreement from [86]

2.3.2 Restricted Boltzmann Machines & Deep Belief Networks

Restricted Boltzmann Machines (RBMs), proposed by Hinton in 2010 [160], [161], are a variant of Markov Random Fields (MRF), made up of a visible input layer **x** and a corresponding hidden layer **h** that contains the latent feature representation of **x**. In the RBM, given an input vector **x**, one may obtain the latent feature representation **h**, but also vice versa, since the connections between **x** and **h** are bidirectional. Therefore, the RBM is considered a generative model, since one may sample from it and generate additional data points coming from the distribution on which the same RBM is trained. RBMs are represented schematically in Figure 2.6b. In the case of Deep Belief Networks (DBNs), these are in practical terms, SAEs with the AE layers replaced by RBMs. Similarly to SAEs, the training of the individual layers is again done in an unsupervised manner. The final fine-tuning is then carried out by appending a linear classification layer to the final layer of the DBN and carrying out the required optimisation in a supervised manner.

2.3.3 Convolutional Neural Networks

Convolutional Neural Networks are built on the backbone of the powerful convolution operation, through a biologically inspired architecture similar to the visual cortex. In the visual cortex, we know that individual cells are sensitive to only a small sub-region from the entire visual field, referred to as

the receptive field [162]. These sub-regions are then tiled together to cover the entire visual field of interest, and the cells therefore act as local filters being applied over the input space. Convolutional Neural Networks (CNNs) similarly exploit such local spatial correlation, where it exists, by applying a local connectivity structure between neurons of adjacent layers. This gives rise to weight sharing that can be applied in a manner resulting in the convolution operation, and which concurrently reduces the parameters that need to be learned. It also renders the network invariant to translations in the input.

At each layer, the input is convolved with a set of filter kernels $W = \mathbf{W}_1, \mathbf{W}_2, \ldots, \mathbf{W}_K$, and also added to a set of biases $B = b_1, b_2, \ldots, b_K$, thus generating different feature maps $\mathbf{X}_K$. The said features maps are then processed with an element-wise non-linear transform (.), and this is repeated for every convolutional layer l:

$$\mathbf{X}_k^l = \sigma(\mathbf{W}_k^{l\ 1} \otimes \mathbf{X}^{l\ 1} + b_k^{l\ 1}) \tag{2.3}$$

The convolutional layers in CNNs are alternated with pooling layers which aggregate, using max or mean functions, the pixel values of a local neighbourhood of specified dimension. At the end of the network, a fully connected layer (without sharing) is usually added as a final classification layer. Unlike DBNs and SAEs, CNNs are trained in an end-to-end fashion in a completely supervised environment, as opposed to being trained layer-by-layer in an unsupervised manner. A simple schematic of a CNN is shown in Figure 2.6d.

2.3.4 Recurrent Neural Networks

Recurrent Neural Networks (RNNs), similarly to CNNs, exploit structure in the data - but this time using structure of a sequential nature. Instead of learning the probability of an output Y given an input vector $\mathbf{X}$, the network learns to predict the output based on a sequence of input vectors $\mathbf{x}_1,\mathbf{x}_2,....,\mathbf{x}_T$. The RNN therefore maintains a hidden, or latent state $\mathbf{h}$, at a particular time t, that is a non-linear mapping derived from the combination of its input $\mathbf{x}_t$ and the previous state $\mathbf{h}_{t\,1}$:

$$\mathbf{h}_t = \quad (\mathbf{W}\mathbf{x}_t + \mathbf{R}\mathbf{h}_{t\,1} + \mathbf{b}) \tag{2.4}$$

where coe cient matrices $\mathbf{W}$ and $\mathbf{R}$ are shared over time. If the problem being considered is one of classification, a fully connected layer is added towards the end, followed by a softmax function to map the sequence to a posterior over the classes. RNNs are considered deep structures, particularly in time, since the gradients need to be backpropagated from the output, back through previous time instants as well. They are therefore prone to su↵ering from the problems of fading or exploding gradients during the learning process. This problem is addressed with the use of special memory cells called Long Short Term Memory (LSTM) cells, which act as a di↵erentiable version of a computer memory chip [163]. Recurrent networks are shown schematically in Figure 2.6c.

2.4 Types of Deep CNN Architectures

With convolutional neural networks being by far the most ubiquitous, particularly in the medical imaging sector, this section focuses on describing di↵erent types of CNN architectures available and used in literature.

2.4.1 Classification Architectures

When it comes to classification, a popular benchmark dataset is that of the ImageNet Large Scale Visual Recognition Challenge (ILSVRC), which serves as a commonly available dataset against which researchers may test the mettle of their developed CNNs. Several (now famous) CNN classification architectures have been implemented and tested against this dataset, as shown in Figure 2.7

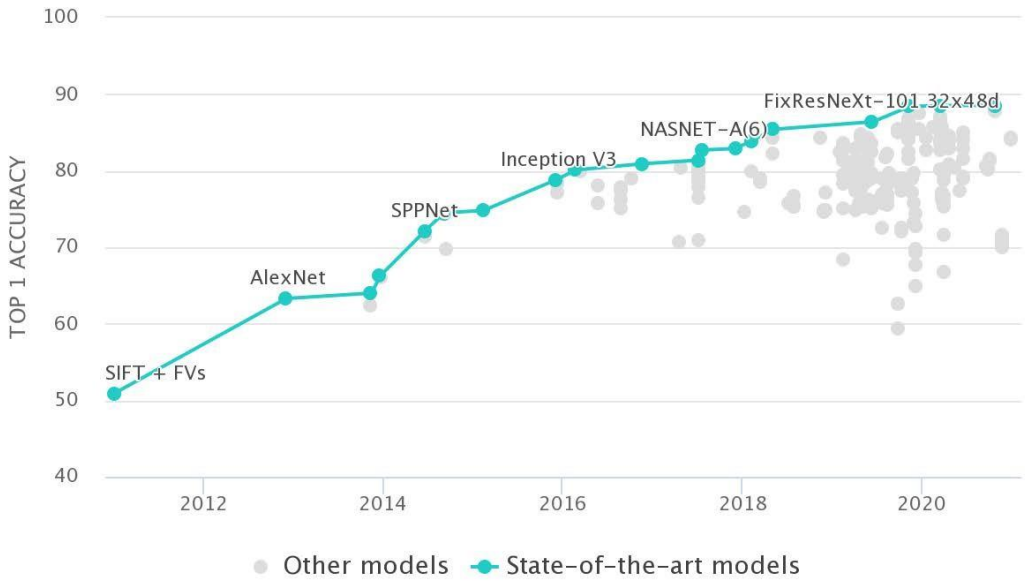

Figure 2.7: A chart showing performance of NNs on ImageNet dataset. Reproduced with permission from [164].

[164]. One of the more famous ones is that developed by Krizhevsky *et al.* in 2012 [165], who introduced the AlexNet and broke the previous records held till then. It is until today one of the best known CNN classification architectures which comprises feature maps of 96, 256, 384, 384, 256 kernels with pooling at the 1st, 2nd, and 5th layers.

More recently, deeper models with additionally complex structures seem to be more in preference - as they have shown to be able to represent classes in a better manner, while concurrently requiring a smaller memory footprint [166; 167]. The first of such networks, was the *VGG-19*, a 19-layer network developed by Simonyan and Zisserman in 2014 [168], which used fixed 3x3 sized kernels. In the same year, Szegedy *et al.*, [169] introduced a 22-layer network coined *GoogLeNet*, which made use of what authors termed 'inception blocks'. The latter may be thought of as a network-inside-a-network, where the input is branched onto multiple additional CNNs, which are then concatenated at the end of the block [170]. The Residual Network (*ResNet*) architecture proposed by He *et al.* in 2015, [171] won the ImageNet challenge in that same year, and many variations on this architecture were winners in subsequent years. A ResNet block is defined as:

$$y = x + F(x, W_R) \tag{2.5}$$

From equation 2.5, one notes that the network need only learn the residual $F(x, W_R)$. In this manner, the network is preconditioned to learn simple representations in each layer which are close to the identity function. As of 2020, the present winner of the ImageNet challenge is the ViT-H/14 Network proposed by Dosovitskiy *et al.* [172], which makes use of Transformer

Encoders with global self-attention, pre-trained on large datasets and then fine-tuned on the task at hand.

2.4.2 Multi-stream Architectures

The classic CNN architecture can inherently easily accommodate multiple sources of information in the form of separate input channels, presented at the input layer [86]. Fusion of di↵erent streams of data may however, in principle, happen at any point in the pipeline. The recent review paper by Zhou *et al.* [173] covers three principal broad categories of fusion strategies which are described as follows:

Input-level Fusion

In networks which employ input-level fusion, multiple channels of data, or alternatively multiple sources/modalities of data, are fused as a multi-channel

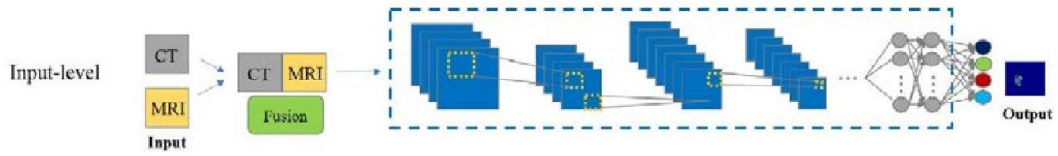

Figure 2.8: A typical input-level type fusion network. Reproduced under openaccess agreement from [173].

data matrix to the network. The network learns a fused feature representation prior to carrying out the task of detection, classification or segmentation. Indeed, many studies in the medical imaging field employ varying types of input-level fusion strategies [174; 175; 176; 177; 178; 179; 180; 181; 182]. A conceptual example is shown in Figure 2.8, where two modalities such as CT

and MRI are considered. At the input level, images from the two modalities are fused into one data matrix and fed into the CNN, which then takes rich feature information from both and exploits them at all layers in the network. This type of fusion at the input level is typically used for the following applications: multi-task applications; multi-view applications; multi-scale segmentations; and finally, Generative Adversarial Network (GAN) applications.

Layer-level Fusion

When fusing information at the layer level, a single-channel or multi-channel dataset is fed into separate, individual networks, which learn individual feature representations. These features being learned at di↵erent layers of the network, are then fused across the networks by concatenating a feature map at a particular layer, into the input of a corresponding layer on the other network. This concept is best explained through the diagram shown in Figure 2.9. The combined training outputs are then finally fed into the decision layer at the

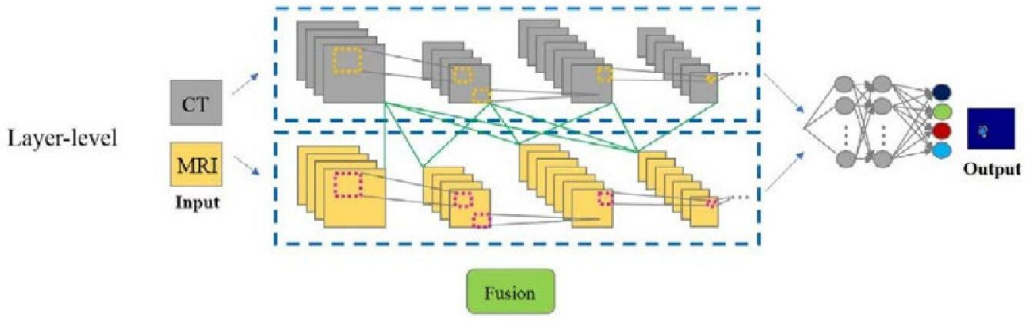

Figure 2.9: A typical layer-level type fusion network. Reproduced under openaccess agreement from [173].

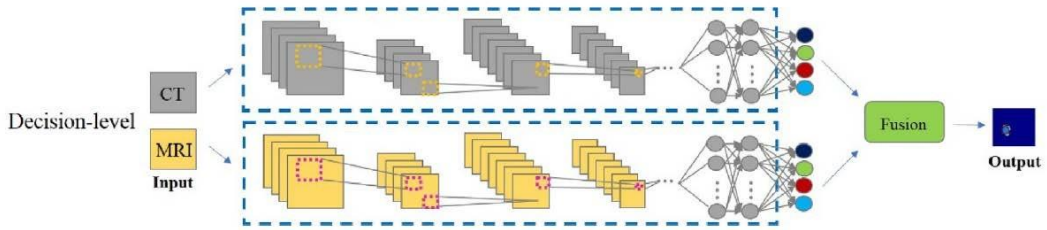

Figure 2.10: A typical decision-level type fusion network. Reproduced under open-access agreement from [173].

output side to obtain the final result. This type of layer-level fusion strategy e↵ectively integrates information across networks and the connections among various layers allows the capturing of complex relationships between input sources. Some studies which employ this layer-level fusion strategy include: [183; 184; 185; 186].

Decision-level Fusion

Lastly, in a decision-level fusion strategy, similarly to the layer-level fusion, separate individual networks are fed a single or multichannel data matrix as input. The individual networks train solely on this datasource to fully exploit the unique information available within that particular datasource. The outputs of the individual networks are then fused to get the final result. Different fusion techniques have been used in literature, but the most common approaches are those based on averaging of the confidence levels of the individual networks, or alternatively, taking a majority vote from the individual networks to determine the final result of a node [187; 188; 186; 189]. The diagram in Figure 2.10 shows this type of decision-level fusion. This type of fusion strategy is ideal for processing multiple *modalities* of inputs such as MRI, CT, Ultrasound etc, whereby the images have completely di↵erent

acquisition techniques and therefore quite di↵erent statistical properties. From a memory-usage perspective, decision-level fusion methods typically require more memory since more layers and parameters are needed to learn the models, given that the fusion is only happening at the end.

2.4.3 Segmentation Architectures

Image segmentation is a very commonly required task in medical applications. CNNs in their standard format are designed for classification problems, whereby an input is classified into one category or another. However, they are easily extended, such that they may categorise individual pixels in an image, thus producing a segmentation mask. Having said that, segmentation is a fundamentally di↵erent task to classification or object detection, in that the segmentation algorithm need not *necessarily* know or understand what the visual objects inside the image are. It would however be ideal, for the segmentation algorithm to do both, and thus segment known and unknown objects. There are many applications where such segmentation could complement classification or object detection, allowing the problem to be approached at the semantic level. For instance, in a problem of content-based image retrieval, a user might need to query a database to find *'all motorcycles available in the images in the database'*. Thus, given a new image, an ideal algorithm would label which pixels belong together semantically, rather than merely on a level of line intensities and contours. It is at present unclear as to how the human brain perceives images and finds the correct segmentations [190].

Traditional image segmentation algorithms have employed clustering methods, along with complimentary information related to edges and lines.

Well-known and successful methods have employed Markov models [191], hierarchical contour detection approaches [192] as well as region-growing strategies [193]. The recent advances in neural networks have however obtained better results, and are now considered to be the state-of-the-art, achieving the best benchmark performances on international datasets. These segmentation strategies may be largely grouped under the following categories, which are described as follows.

A detailed review is available in [190].

Region-based Semantic Segmentation

Techniques which are region-based typically follow a pipeline which applies segmentation after recognition. Initially therefore, free-form regions are first extracted and described from an image. This is then followed by region-based classification whereby the region-based predictions are transformed to pixel predictions, by assigning a pixel label to the highest scoring region that it finds itself within. An important study dedicated to region-based segmentation is that on Regions with CNN features (RCNN) by Girshick *et al.* [194]. This method uses selective search to first extract a large quantity of proposed objects. For each of these it computes features, and then classifies the regions using class-specific linear Support Vector Machines (SVMs). A technique called Second-order pooling is used to label the pixels [195]. RCNN can also be built upon other CNNs like AlexNet, GoogLeNet and ResNet [190].

RCNN also suffers from some disadvantages, which have spurred some additional research. Firstly is the problem that the features extracted are not designed for a segmentation task. To address this, Hariharan *et al.* [196] proposed an additional network which was specifically fine-tuned to segment

the relevant region foreground, since they argued that the RCNN was actually fine-tuned to only classify bounding boxes. Secondly, RCNN uses the activations from the final fully connected layer, which although containing more semantically meaningful information, lack the spatial information available in the intermediate layers. In [197], Hariharan *et al.* proposed to address this issue by adding Hyper-columns as pixel-descriptors, which include all the CNN activations in the network for a particular pixel. This is intuitively similar to appending both coarse and fine-layer information to allow for higher accuracy and spatial precision. Thirdly, RCNNs take time to generate segment-based proposals, and this greatly a↵ects performance. This is in contrast to more e cient alternatives of RCNN like [198; 199].

FCN-based Semantic Segmentation

Fully Convolutional Networks (FCNs) are essentially an extension to the classical CNN structure, which allow for the network to be trained end-to-end and to learn a mapping from pixels directly to pixels, without extracting any proposals related to regions [200; 201; 202]. In Figure 2.11, one observes a typical network structure for a regular CNN. One notices that on the output side is a fully connected MLP layer. This layer is the one which condenses the learned features of the CNN into a single classification output. For this reason also, there is a restriction on CNNs to accept and provide labels for inputs of a specific size only. In contrast, FCNs are constructed solely out

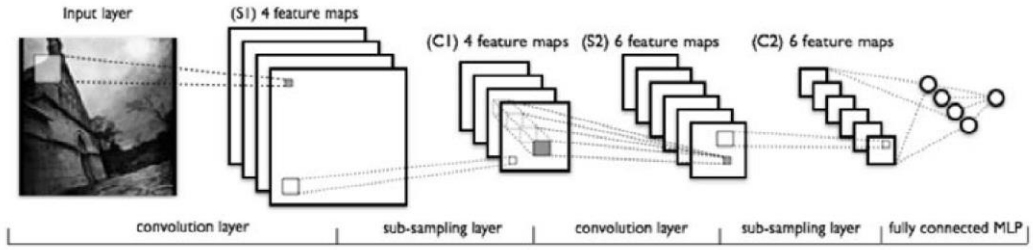

Figure 2.11: A full convolutional neural network constructed from convolution blocks, max pooling blocks and fully connected layers at the output. [203]

of convolutional layers and pooling layers, omitting the full-connected layer which is typically present at the output side of CNNs. They are thus termed *fully convolutional* networks for this reason. This gives them also the facility to make predictions on arbitrarily sized inputs, with the consequence however that the output size is no longer fixed, but also depends on the initial size of the input, given that it scales downwards and subsequently upwards from the input's initial dimension.

FCNs normally implement a per-pixel softmax function layer at the end of the output. This function has the e↵ect of making the value of the largest response for a pixel to be close to 1, while making all other output values at the same output nodes close to 0. The output of FCNs produce a pixel-dense output, capable of creating a segmentation mask. For this reason, they are commonly used for local segmentation tasks as opposed to global tasks like classification. They are also end-to-end trainable. An important issue that is encountered by FCNs is that, due to their contracting nature, which comes about by several layers of convolution and pooling, the output resolution at each layer decreases successively. This means that the output must necessarily have a low resolution, with quite fuzzy object boundaries. In order to address this

issue, Eigen *et al.* [201] proposed a multi-scale convolution network which implements multiple sub-networks having di↵erent resolution outputs, in order to iteratively refine the prediction. Another approach, proposed by Long *et al.* [200] was to apply a 'shift and stitch mechanism' along with deconvolution layers to produce an output of nearly the same resolution as the input. In 2015, the group of Wang *et al.* [204], proposed to append to the contracting encoding network, a similar but flipped expanding decoder section, comprised of deconvolution and up-sampling layers. A diagram showing this structure is shown in Figure 2.12. At around the same time, the group by Ronneberger *et al.* [205] also proposed a very similar structure which they called 'U-NET', having a contracting encoding path followed by an expanding decoder part, but with the addition of skip connections across encoder-decoder layers. The idea behind this was to transfer and concatenate spatial features from the high resolution maps to the expanding side. The U-NET would go on to become a popular implementation variant in FCN based studies, with several other studies proposing adaptations to it and extending it into 3D space [206; 207; 208].

Weakly Supervised Semantic Segmentation

Most of the works in semantic segmentation require a large number of training images with corresponding manually labelled image segmentation masks. Unfortunately, the manual annotation of such masks needs to be done by clinical experts often with limited time. This makes the process time-consuming, expensive and frustrating. Thus, a third type of semantic segmentation exists, which focuses on segmentation by making use of just

annotated bounding boxes or just image-level labels. This is referred to as weakly supervised semantic segmentation, due to the weakly labelled data.

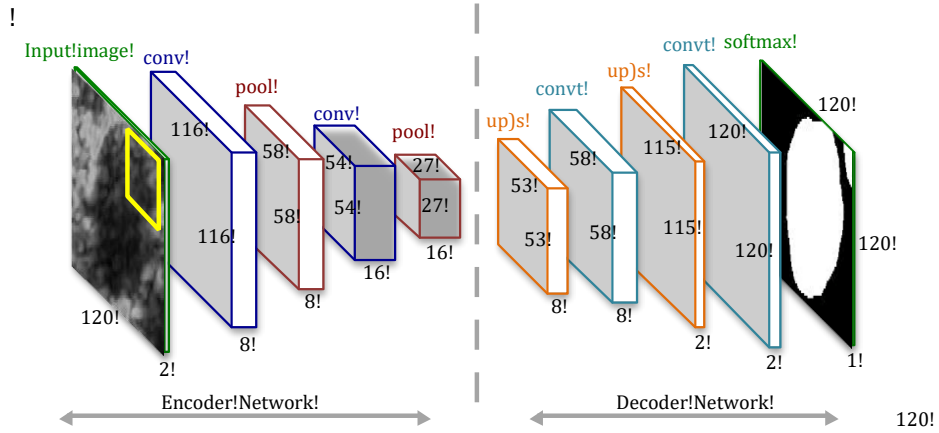

Figure 2.12: A graphical representation of the encoder-decoder structure of the Deep Neural Network.

The group of Dai *et al.* [209] proposed to train the network with bounding box annotations, and then improved iteratively the estimated masks for segmentation. Similarly, Papandreou *et al.* [210] trained semantic segmentation models with weakly annotated data, which included image-level or bounding box annotation. However, they proposed to use an Expectation-Maximisation (EM) method for training. They found that using image-level annotation alone yielded insu cient performance, whereas using bounding box annotation yielded performance which was competitive with pixel-level annotation. Ultimately, it proved beneficial to combine them together. The group of Khoreva *et al.* [211] treated the weak supervision problem as a denoising of input label noise, and proposed a recursive training model to reduce said noise.

The alternative to box annotations, that of using image-level annotations as a weak supervision, was explored by other research groups. The group of Pinheiro *et al.* [212] approached the segmentation problem within a multipleinstance learning framework. Similarly, Pathak *et al.* [213], cast each input image as a bag of pixel-level examples, and proposed a pixel-level loss to adapting to a multiple-instance learning framework. The same group later proposed a self-training framework [214], whereby a constrained CNN with a novel loss function was used to enforce consistency between the per-image annotation and the predicted segmentation mask. A general issue with the use of imagelevel annotations is that the object-localization is ignored. Several studies [215; 216; 217; 218] have sought to address this problem by exploiting the notion of objectness by incorporating it in the loss function or by pre-training a network as an additional external objectness module. An additional alternative proposed [216; 219] to improve performance is to use additional weakly supervised images such as web images, to train the CNNs.

2.4.4 Choice of Network Architecture

The choice of network architecture used for the problem of delineating the Media-Adventitia interface and Lumen-Intima interface in carotid ultrasound images is a decision which is influenced by the nature of the problem itself, but also by what is practically acceptable in-terms of commercially available ultrasound platforms. The delineation of these two vascular wall contours leads to clinical decisions concerning the patient's health, and thus segmentation accuracy is key. This therefore precludes methodologies built on regionbased semantic segmentation, which as described earlier tend to yield fuzzy segmentation masks with poor spatial precision. Indeed, such

methodologies find larger application in general multimedia applications, rather than in the medical sector for the task of object detection within an image database. In addition, commercial ultrasound platforms have, for a very long time, had a ready accessible 'windowing function', whereby the sonographer uses a trackball to select a region of interest prior to the ultrasound software applying further algorithms for processing. This manual selection of a square window around the region of interest is therefore ingrained in sonographers' use of these commercial machines, and it is an action performed quickly within seconds. The quick availability of a region of interest around the object (in this case the artery), therefore precludes the necessity of algorithms whose strength lies in object detection as opposed to pixel-by-pixel classification accuracy with high spatial precision.

Network architectures intended for weakly supervised supervision are obviously attractive, particularly from a training perspective. However, the ultimate goal of having two segmentation masks (one for the MAB and one for the LIB) makes it more di cult to simply apply a boxed window around two contours within each other. Additionally, clinicians' time is limited, and having access to laboriously, individually-labelled segmentation masks for each training image is often problematic. Still however, such techniques have yielded inferior performance to pixel-by-pixel labelled segmentation strategies, and thus where possible, if the additional investment of obtaining accurately labelled segmentation masks from experts is possible, as was in this case, then the tendency to lean towards pixel-by-pixel segmentation strategies will prevail. The author does acknowledge however, that further research into weakly supervised network architectures is always warranted, since it would

ultimately open the door to much larger datasets and therefore improved generalisability.

The obvious choice for this segmentation problem therefore becomes that of Fully Convolutional Networks. FCNs have been shown to be trainable in an end-to-end fashion, and yield a fully automated segmentation, with very good spatial accuracy, given a good training process. In order to have a good spatial resolution at the output segmentation mask, the obvious choice would be to use U-NET structures and their variants. Initially, at the time of research on this topic and eventual publication (as opposed to the time of writing of dissertation), the concept of U-NETs with skip layers was still in its infancy, and easily available development toolboxes implementing such U-NETs were not available to the author. Conversely however, what was available at the time was the ConvNet toolbox, which allowed the construction of U-NET-like structures comprising of encoder structures, followed by an appended decoder structure, but without the skip connections. In Chapter 4, the author therefore initially explores the use of such FCNs for the problem of delineating the MAB, being the first – to the author's knowledge – to apply such FCNs on carotid ultrasound data. A thorough evaluation on the final performance of the e↵ect of di↵erent network structural parameters, such as number of layers, depth of layers and filter size is carried out. In Chapter 5, a U-NET is subsequently used, along with a proposed modification to the cost function to improve performance.

Furthermore, a multi-stream architecture is adopted, by proposing the novel fusion of amplitude and phase congruency information which will be described in Chapter 3. The multi-stream architecture chosen is an input-level

fusion strategy, as this allows rich feature information from both channels to be exploited at all layers of the network while concurrently learning the intrinsic image features. At the time of research and publication on this particular topic, input-level fusion strategies were the available and popular approaches. The advent of layer-level fusion strategies and decision-layer fusion strategies started making their appearances from 2017 onwards according to the review by Zhou *et al.* [173]. At the time of publication, the input-level fusion strategy chosen yielded a good performance for the author's application, yet the potentially beneficial implications of layer-level or decision-level fusion strategies, to the latter's ability to learn more complex and complimentary information during the training process is acknowledged. Thus, evaluation of such alternative fusion strategies could form part of the scope of further work in this regard.

2.5 Conclusion

This chapter has provided an overview of the applicable, as well as emerging trends in ultrasound image processing techniques proposed in literature for obtaining novel indices, which measure plaque burden. Based on this, it has provided the justification for the choice of methodology in approaching the problem of quantifying the vessel wall volume from ultrasound images of the carotid. It has also provided an overview of deep neural network architectures used in medical imaging, and discussed justification for the proposed choice of network architecture used in this work. In the next chapter, the image acquisition protocol is firstly described, and then a novel, amplitude-invariant image representation based on phase information is described next.

3

Segmentation with RF data and Phase Information

3.1 Introduction

Similarly to other clinical applications, carotid ultrasound images are influenced by the presence of speckle noise and by an overall image quality that is dependant on machine settings [121]. As highlighted by Meiburger *et al.* in their recent review in 2018 [220], the segmentation of the carotid artery in ultrasound presents added complications of having high variability in vessel morphology, particularly when pathology is present, as well as the possible presence of backscattering in the lumen.

In Chapter 2 the popular techniques that have been applied to segment the carotid walls were described. The majority of segmentation methods are based on B-mode images and gray-scale data [221; 222], with the latter being constructed from the envelope of echo data. With the increasing availability of access to radio-frequency (RF) data from ultrasound machines however, a

marked interest in processing such RF data directly is noted [220]. What makes RF data attractive is the fact that it potentially contains more information than the processed envelope-based B-mode images. Recent studies have therefore opted to process this information natively. Boukerroui *et al.* propose one such algorithm, which uses 3D adaptive clustering on multi-parametric information, amongst which are some acoustic features derived from the RF data [221]. Dydenko et al. propose to extract envelope power, spectral based autoregressive parameters and velocity-based parameters from RF data. Boundary detection is then computed using a discontinuity adaptive smoothing filter [223]. Nillesen et al. propose an iterative, coarse-to-fine approach for calculation of maximum correlation coe cients (MCC) from RF data [224]. All these techniques obtain reasonable segmentation performance with RF-based features alone. However, improved performance, exceeding that even of envelope-based methods, is obtained when RF-based features are used in conjunction with other features [221; 224].

Phase information within ultrasound data has also been suggested as an alternative to extract information [119]. It is argued that it provides a better basis for segmentation, because phase should theoretically be invariant to intensity and magnitude [119]. Mulet-Prada *et al.* [225] and Belaid *et al.* [226] propose two such methodologies, using a local phase-based approach to extracting features from 2D images. Within the RF domain however, little analysis of such phase properties has been carried out, primarily because of the random nature that the RF phase signal exhibits [227]. Despotovic *et al.* however show, albeit perhaps only qualitatively, that although phase information in RF signals is mostly randomly distributed, there exists correlation between neighbouring phase samples [227]. In their study, an

entropy filter is applied to a phase di↵erence matrix, obtained from a premature baby's ultrasound scan. The authors analyse the entropy image and suggest that correlation exists between low entropy values and the presence of tissue structures.

3.1.1 Aims

The aim of this chapter is to propose an amplitude invariant data representation of carotid images, such that this may be used together with an image segmentation technique to delineate the MAB in a robust manner. Thus, a novel phase-based representation, computed directly on RF signals is proposed and tested. Firstly, a quantitative analysis is carried out on the performance achieved when computing an intra-frame phase di↵erence matrix from within the in-phase and quadrature demodulated RF signal lines. Subsequently, a Gabor Filter Bank is used to extract textural responses. In order to assess the quality of the data representation and compare against similar literature, a segmentation through k-means clustering and Active Contours is carried out.

This is then compared with the second alternative: the novel application of phase congruency maps directly on RF signals. This method is influenced by the work developed by Kovesi *et al.* in [228], whereby the phase congruency maps proposed by the latter are extracted from RF data directly as opposed to from envelope data. This allows one to exploit the shape of the frequency spectrum of RF data and results in a better signal-to-noise ratio. Similarly to the first approach, a segmentation through k-means clustering and Active Contours is carried out. A quantitative analysis of performance obtained is presented and the segmented data from both methods is compared to data

which is manually delineated by an expert, and also against a third technique proposed in literature.

Novelty Statement: At the time of publication (2015), the novelty of this chapter's work was the creation of an amplitude invariant, RF-phase-based representation of carotid ultrasound images.

3.2 Methodology

3.2.1 Acquisition Hardware

The data acquisition platform used for this work was a Sonix RP Diagnostic Ultrasound system, manufactured by Ultrasonix systems (later falling under BK Medical, Peabody, USA), shown in Figure 3.1 The system comprises a full scale diagnostic/clinical ultrasound system, which however comes packaged with an ultrasound research interface (URI) allowing researchers to acquire and store low level raw data in a variety of formats. The URI additionally allows the use of operational modes not classically available on a purely clinical system, as well as the retrieval and modification of low level parameters used to generate ultrasound images.

Figure 3.1: The Ultrasonix Sonix RP System used in the studies [17].

The hardware specifications of the system include 128 element transducers, 256 transmit channels and 32 receive channels, a 40MHz transmit and receive sampling clock, and a 10-bit analog to digital conversion. The system's URI includes a software development kit (SDK) available for MATLAB (Mathworks, USA), which allows for real-time acquisition of data in either raw RF formats or post-processed images.

The SDK also includes a set of MATLAB functions for implementing the image processing pipeline, required to convert raw RF images to traditional B-Mode images. These will be described in a later section. The ultrasound images were acquired in a two-dimensional (2D) format by using a 14 MHz L14-5

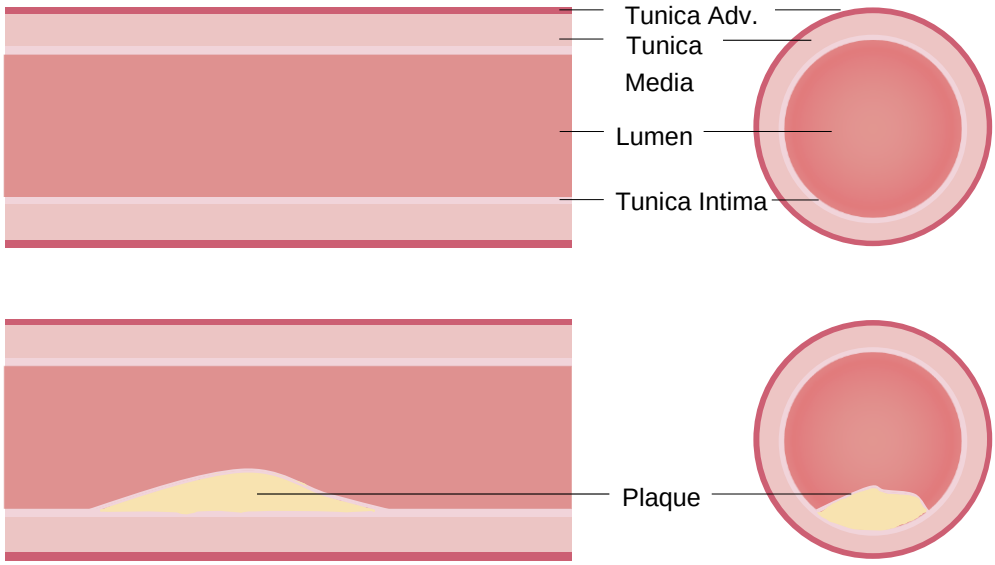

Figure 3.2: [Top] An example of a healthy carotid artery and [Bottom] an example of an artery with an accumulation of atherosclerotic plaque.

Linear Probe. The ultrasound scanner settings were set as follows: Frequency: 6.6MHz, Depth: 3.0cm, Sector width: 100%, Gain: 51%, Dynamic Range: 92dB, Persist Setting: 0, Map Setting: 9, Chroma Setting: 0; Power Setting: 0; and were kept constant across all subject acquisitions. Data was acquired and stored concurrently in both raw RF data formats and B-mode images.

3.2.2 Data acquisition protocols

A total of 50 transverse, two dimensional ultrasound images were obtained from each of 18 subjects, having carotid arteries which display varying degrees of stenosis. An example of stenosis caused by plaque is shown in Figure 3.2. In the case of the first five subjects, a total of 50 longitudinal, two dimensional ultrasound images were also acquired. Subjects had ages spanning between 25 to 80 years. Subjects provided signed consent to participate in the study, which was in turn approved by the University of Malta Research Ethics Committee. Table 3.1 shows the list of subjects with respective ages and degree of stenosis

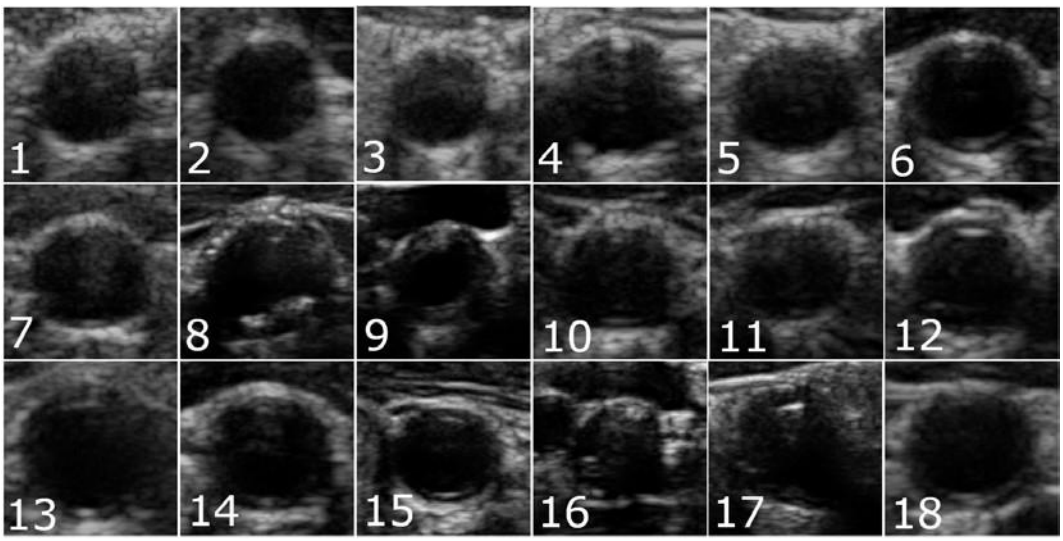

Figure 3.3: A single transverse B-mode image taken from each subject.

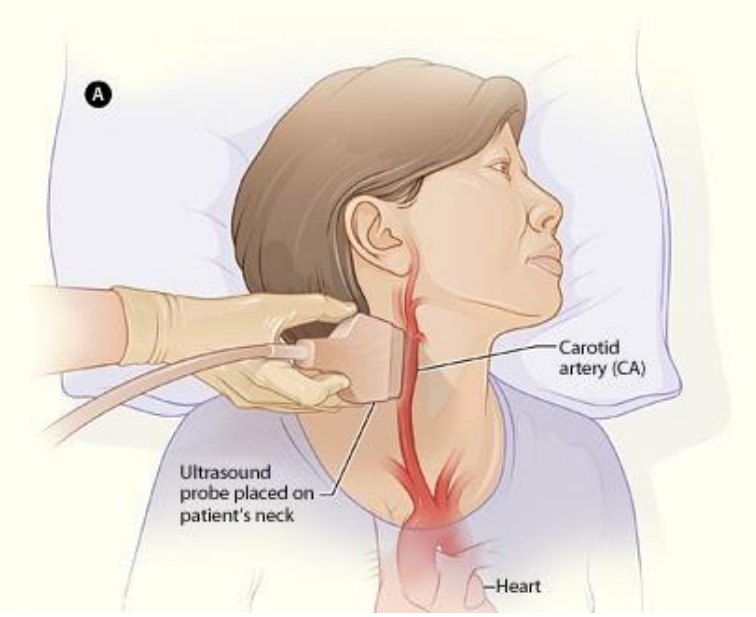

Figure 3.4: An example showing longitudinal ultrasound acquisition of the carotid artery while patient lies in supine position. Image reproduced under public domain license from [229].

found. Figure 3.3 shows an image taken from each subject. Subjects were asked to lie supine on a couch, and the probe was placed against the neck as shown in Figure 3.4, while an image sequence was acquired for transverse carotid sections. The probe was kept in the same spatial location, while the image sequence was acquired at a rate of 24 Hz, thus acquiring a clip of frames over a span of time.

All the acquired transverse and longitudinal ultrasound images were manually and independently traced with the assistance of two radiographers. The RF images that were acquired using the Ultrasonix machine were saved to an external USB disk and transferred to a processing notebook (Apple Macbook Pro, running Intel Core i7 processor and NVIDIA GeForce GT650M GPU). Using

the Ultrasonix SDK available for MATLAB, a Hilbert transform was Table 3.1: Subject ages and the degree of stenosis for each patient.

% Subject Number	**Age**	**Stenosis**
Subject 1	30	< 30%
Subject 2	29	< 30%
Subject 3	31	< 30%
Subject 4	28	< 30%
Subject 5	50	< 30%
Subject 6	70	< 39%
Subject 7	71	< 30%
Subject 8	65	47%
Subject 9	69	59%
Subject 10	68	< 30%
Subject 11	72	< 30%
Subject 12	62	39%
Subject 13	61	48%
Subject 14	73	36%
Subject 15	77	46%
Subject 16	78	63%
Subject 17	80	59%
Subject 18	79	39%

used to demodulate the amplitude information from the RF sinusoids. This amplitude information was then used to compute the B-mode images and phase congruency maps. For the creation of B-mode images, the amplitude data was passed through a logarithmic function to adjust for dynamic range, and then decimated by a factor of four. The resulting B-mode image was scan converted to obtain correct image geometry. A median speckle reduction filter was then implemented as per methodology described in [230], and applied over all the B-mode images to reduce the e↵ect of speckle noise.

In order to trace the MAB and LIB and label the regions in the images, a script was prepared which uses the MATLAB function *impoly* to allow a user to manually delineate the border of the MAB and LIB by placing 100 points on the contour. The script was presented to two radiographers to delineate the contours on the images presented to them.The contours returned by the *impoly* function were used to create an image mask, with '0's outside the contour and '1's inside the contour. The process was repeated for both MAB and LIB, and another script used to then relabel the three zones as '0' (outside MAB), '1's (inside MAB but outside LIB) and '2's (inside LIB). An example of the delineating process and the labelled result is shown in Figure 3.5.

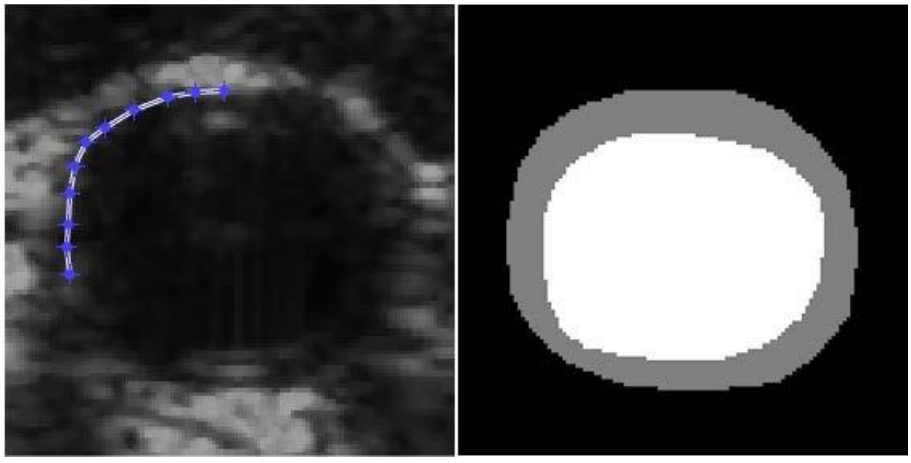

Figure 3.5: [LEFT] An example of a B-mode image with delineation of MAB interface. [RIGHT] An example of a labelled image showing three di↵erent regions: Background (Black), Vessel wall between MAB and LIB (Grey), and Lumen (White).

For the first 5 subjects, the radiographers labelled all the images. For the remaining subjects, the radiographers labelled key images at specific intervals, whereas the author labelled the remaining images under their supervision. Both radiographers and author manually traced the image sets twice, with a period of 2 weeks in between sessions, and an average trace was used as ground truth. For all the subjects, both the MAB and LIB of the carotid artery were delineated in the transverse section. Additionally, for the first 5 subjects, the radiographers traced the MAB in the longitudinal sections as well. The latter dataset was used in a subset of experiments, which will be described in the next chapter.

3.2.3 Segmentation with Phase Di erence Matrices

In the work by Despotovic *et al.* in [227], the authors show that there exists a correlation between neighbouring phase-signal samples taken from RF data, that can be used for tissue characterisation when examining pre-term brain tissue. In this work these concepts are examined when applied to vascular tissue of the carotid artery. The data acquisition took place as described in section 3.2.1 and 3.2.2, and the SONIX RP platform produced an RF data matrix having dimensions of 1568 x 256, that is 256 individual RF lines (referred to as A-lines), each with 1568 samples. An example of a single A-line may be seen in Figure 3.6, part A. In relation to the B-mode image shown in part B, this RF line is taken from the middle, that is, it is the 127th RF line demarcated with

the dotted yellow line. The three large peaks correspond to three large reflections, which take place at the interfaces of the arterial walls, denoted by i1, i2 and i3. Figure 3.6 part C shows the complete RF image.

Demodulation and Image Construction

Part of the process of converting RF-data to B-mode images involves amplitude demodulation. Two very popular methods of doing this are through either a Hilbert Transform or through Inphase-Quadrature (IQ) Demodulation. The latter is a good approach to reduce the amount of data in the RF signal without losing important information, since it applies complex base-band modulation with bandwidth reduction. It also allows for a better Signal-to-Noise (SNR) ratio, since IQ demodulation keeps the sampling rate high and that, in turn, suppresses the quantisation noise during analog-to-digital conversion. The process of IQ-demodulation consists of 4 steps, which are schematically shown in Figure 3.7.

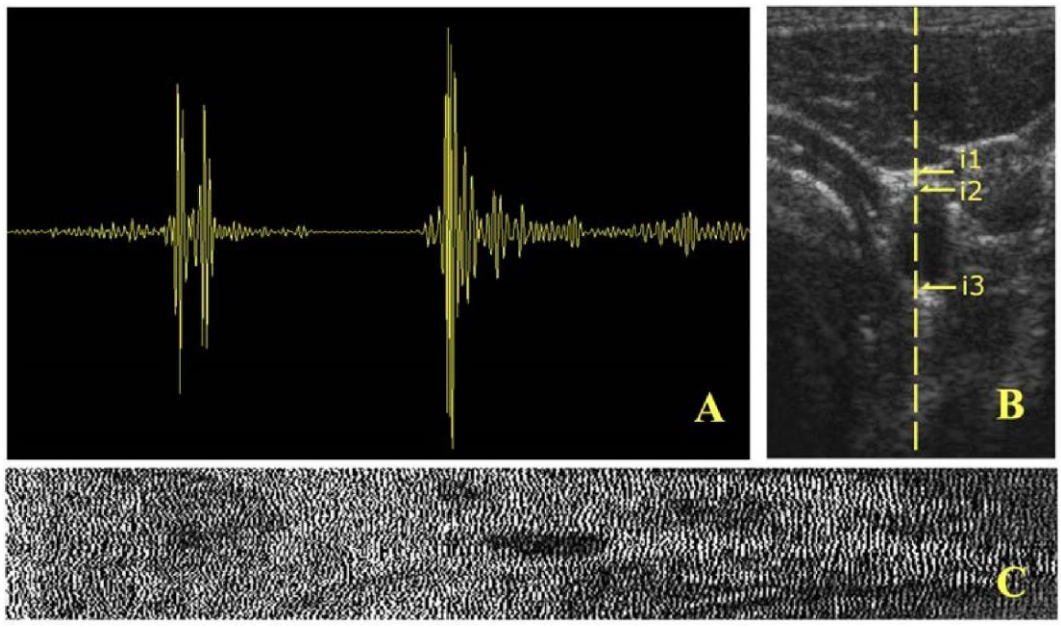

Figure 3.6: [A] A single RF signal line sampled from the ultrasound transducer [B] A typical B-mode image; [C] A corresponding unprocessed RF data image which has not undergone signal processing (shown in landscape).

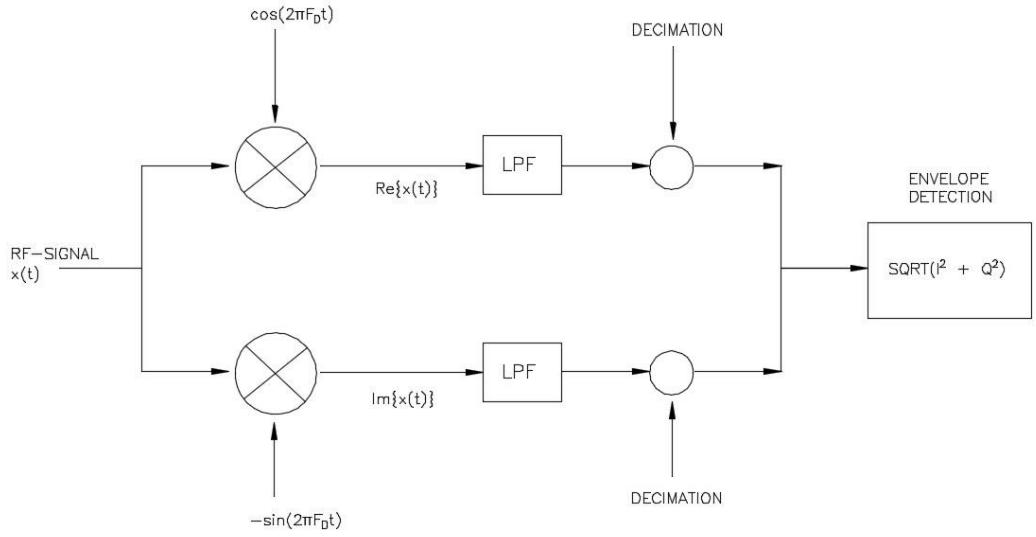

Figure 3.7: IQ demodulation process, showing downmixing, low pass filtering, decimation and envelope detection.

1. Down-mixing
2. Low pass filtering of both signals
3. Decimation
4. Envelope Detection

The IQ demodulation process produces the envelope of the RF signal, which is typically considered to contain most information, and is represented by the following equation:

$$Envelope = \sqrt{I_f(n)^2 + Q_f(n)^2} \quad (3.1)$$

where $I_f(n)$ and $Q_f(n)$ are the low pass filtered and decimated outputs of the in-phase and quadrature signals given by:

$$I(n) = x_{RF}(n) \cdot \cos(2\pi f_d nT) \tag{3.2}$$
$$Q(n) = x_{RF}(n) \cdot \sin(2\pi f_d nT) \tag{3.3}$$

In this work, the process of IQ demodulation was carried out in MATLAB by adapting code which was already available as part of a different function forming part of the 'SONIX DATA TOOLS' Ultrasonix Software Development Kit. The RF data was imported as a data matrix as described in Section 3.3.1. The complex sinusoidal signal was implemented in two parts by means of preinitialised, windowed sine and cosine table, for computing the in-phase and quadrature signals respectively. The window chosen was a hamming window with a length of 20 samples and an overlap of 5 samples. Each A-line from the RF matrix was processed separately, and multiplied in windowed segments with the corresponding windows of sine and cosine tables.

The windowing technique is used in ultrasound IQ demodulation to cater for spectral leakage arising from discontinuities in the signal, and also to permit compensation as required for attenuation of the signal in relation to imaging depth. The latter is in fact required in dynamic IQ demodulation, whereby the demodulation frequency f_d is equated to the mean central frequency which varies along signal depth (or in other words further down along the length of the A-line) [231]. In this case, however, a simple case is assumed and f_d is taken to be constant at 6.6 MHz.

After down-mixing, a low pass filter was implemented in MATLAB, having a cut-off frequency f_{cut} of 5 MHz and applied to both in-phase and quadrature components. The remaining processes of decimation and envelope detection

were not implemented at this stage, choosing instead to retain the signal in the in-phase and quadrature representation for subsequent computation of the phase signal, which is described in the next section.

Phase Di erence Matrix Computation

In the work by Despotovic *et al.* [227], the authors examine the distribution of phase di erence by working out the entropy from local neighborhoods. This information is obtained once the amplitude and phase information have been determined from the preceding IQ demodulation. Despotovic *et al.* observe that the phase information taken purely in isolation appears as a randomly and uniformly distributed variable, which makes it di cult to analyse. To this end, they compute phase di erence information in both axial and lateral dimensions as per equation 3.4 below, as they observe these to be locally correlated.

$$c_1 c_2^* = A_1 A_2 e^{j(\varphi_1 \quad \varphi_2)} = A e^{j(\Delta\varphi)} \quad (3.4)$$

In this equation, $c_1 = A_1 e^{j(\prime_1)}$ and $c_2 = A_2 e^{j(\prime_2)}$ are representing two di erent samples from the same A-line (taken axially) or from two di erent A-lines (taken laterally), and where A_1, A_2 are the amplitudes, $\prime_1, \prime_2$ are the phases, and c_2^* is the complex conjugate of c_2.

An alternative method of obtaining the phase di erence values is as follows. If we consider that the in-phase and quadrature components are shifted 90° apart, then a graph may be plotted showing the same in-phase and quadrature components as shown in Figure 3.8. From this graph, it should be readily apparent that the calculation of the phase may be obtained through simple trigonometric relationships of the filtered in-phase component $I_f(n)$ and the filtered quadrature components $Q_f(n)$ in four di erent quadrants. The phase

may be calculated as per equation 3.5 below, where (n) represents the phase and may take on values of between and .

$$(n) = \begin{cases} \arctan\dfrac{Q_f(n)}{I_f(n)} & \text{if } I_f(n) > 0 \text{ if } I_f(n) < 0, Q_f(n) \quad 0 \\ \arctan\dfrac{Q_f(n)}{I_f(n)} + \pi & \text{if } I_f(n) < 0, Q_f(n) < 0 \\ +\dfrac{\pi}{2} & \text{if } Q_f(n) > 0, I_f(n) = 0 \text{ if } Q_f(n) < 0, I_f(n) = 0 \\ \dfrac{\pi}{2} & \text{>>:undefined if } I_f(n) = 0, Q_f(n) = 0 \end{cases} \tag{3.5}$$

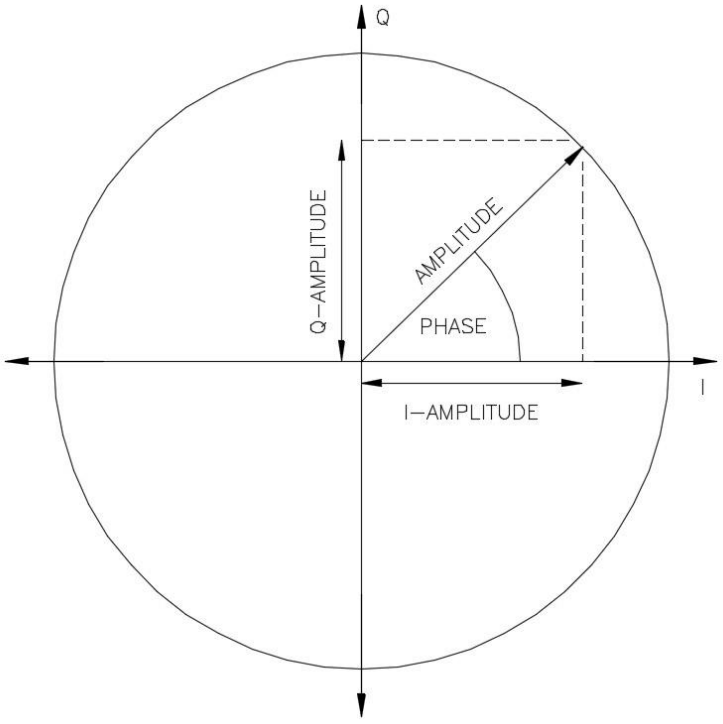

Figure 3.8: Graph showing a plot of in-phase and quadrature components which are shifted 90° apart.

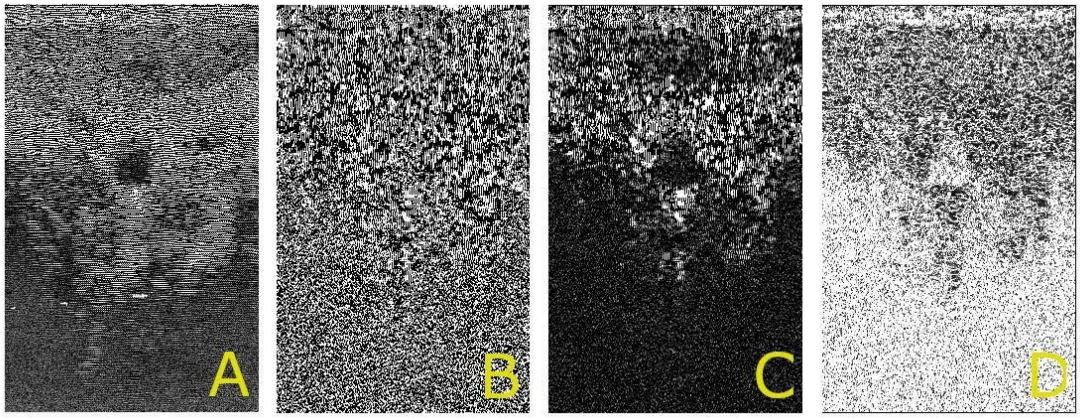

Figure 3.9: A] Original RF image; [B] In-phase component image; [C] Phase Image; [D] Phase Di↵erence Image.

The computation of phase di↵erence, (n), may then be calculated for adjacent samples in a single scan line, or from neighbouring samples across di↵erent scan lines, by simply subtracting the value of one sample from the other one as described in equation 3.6:

$$(n) = \quad (n+1) \qquad (n) \tag{3.6}$$

The phase di↵erence signals calculated are then combined together in columnwise or row-wise fashion to form a phase di↵erence image matrix. The computation of Phase Di↵erence Matrix (PDM) was implemented in MATLAB using a simple custom-built function which implements the calculation shown in equation 3.5 on the filtered in-phase and quadrature matrices to create a phase matrix. The phase matrix was scanned through using a nested loop to compute axial and lateral phase di↵erence matrices. Figure 3.9 shows four di↵erent images, starting with the original RF image in part A; followed by the

phase signal image in part B. Also shown is a representation of the in-phase component image in part C for reference purposes. In part D, one observes the axial PDM. The PDM calculated in the lateral direction showed less structure than that calculated axially. The axial PDM in part D on the other hand clearly shows structural correlation with the original RF image, whereby the central circular artery, shown in darker contrast, is present. For this purpose, the author opted to proceed with the subsequent stages of segmentation using Gabor Texture filters and K-Means clustering, by using the axial PDM as basis.

Texture Feature Extraction with Gabor Filters

Extracting texture features is based on the use of Gabor Filters, and the underlying model of texture perception in the human visual system, described by Malik and Perona in their work [232] and adapted by Jain *et al.* in [233]. In [232], the authors described the model as consisting of three stages:

1. Convolving an image with a bank of even-symmetric linear filters. This is then followed by half-wave rectification to provide a set of responses.

2. Inhibition among the neural response profiles, which are localised in space, such that these result in suppression of weak responses in favour of strong ones at the same or nearby locations.

3. Detection of texture-boundary, by using wide odd-symmetric mechanisms.

The first stage is implemented by using a bank of two-dimensional Gabor filters. The Gabor function is constructed in two dimensions by having a sinusoidal plane wave, having some frequency and some orientation, and

modulating this with a two-dimensional Gaussian function. This may be represented in the spatial domain by using the following equation:

$$g_{\lambda\theta\psi\sigma\gamma} = exp(\quad \frac{x'^2 + \gamma^2 y'^2}{2\sigma^2})cos(2\pi\frac{x'}{\lambda} + \psi) \tag{3.7}$$

where $x^0 = xcos(\checkmark) + ysin(\checkmark)$ $y^0 = ycos(\checkmark) +$

$xsin(\checkmark)$

(x, y) represent the spatial coordinates in 2D space, ✓refers to the orientation of the normal to the parallel lines of a Gabor function, expressed in degrees, refers to the wavelength of the cosine factor, refers to the phase o↵set expressed in degrees, refers to the spatial aspect ratio, and finally, refers to the standard deviation of the Gaussian, and determines the linear size of the receptive field.

The spatial frequency of the cosine factor is expressed as $f = 1/$, and this will be assigned a range of values as will be seen later. The ratio $_s/$ determines the spatial frequency bandwidth, and therefore the number of parallel excitatory and inhibitory zones which may be observed. The ratio is expressed as:

$$\frac{\sigma_s}{\lambda} = \frac{1}{\pi}\sqrt{\frac{ln2}{2}}\frac{2^b + 1}{2^b \quad 1} \tag{3.8}$$

The second stage of feature extraction, based on the adaptation by Jain *et al.* uses a non-linear sigmoidal function to saturate the output of the filters. The function is expressed as follows:

$$tanh(\alpha t) = \frac{1 \quad \exp^{\ 2\alpha t}}{1 + \exp^{\ 2\alpha t}} \tag{3.9}$$

This saturates the sinusoidal modulations in the filtered images to square modulations, and the results may therefore be interpreted by a blob detector. Additionally a Gaussian smoothing function is applied to compute the average

absolute deviation (AAD) for each filtered image. This is given by the following function:

$$g(x,y) = \exp\left(\frac{x^2+y^2}{2\sigma^2}\right) \quad (3.10)$$

The algorithms for texture feature extraction with Gabor filters were implemented by modifying standard functions available within the MATLAB CVPR toolbox, developed by Seo *et al.* [234]. For the choice of filter orientations used with the Gabor function, di↵erent orientations were used, separated by equal spacing each time. Thus, for instance, taking a total of six orientations as an example, the values of ✓were set as follows:

✓: 0°,30°;60°;90°;120°;150°

Furthermore, the spatial frequency $f = 1/$ was set to the following values as recommended in [235]:

$$0.25 \qquad 2^{i\,0.5}/N_c \; f \; 0.25 + 2^{i\,0.5}/N_c \quad (3.11)$$

where $i = 1,2,\ldots,log_2(N_c/\ _{split})$, N_c is the width of the images in columns (in this case, 256), and $_{split}$ is a constant, set manually to determine how many frequencies (or wavelengths) to consider. This would yield a bank of filters of size F_N (no. of orientations x no. of frequencies). During the experiments, the number of filters in the bank were altered by changing the number of orientations and spatial frequencies, to observe their e↵ect on performance. The phase o↵set was set to 0, while the spatial aspect ratio of the Gaussian function was set to 1. The ratio / , which determines the spatial frequency as defined in equation 3.9, was set with a half-response spatial frequency bandwidth b in octaves set to 1. Lastly, the constant , of the non-linear

sigmoidal function was set to 0.25, whereas of the Gaussian smoothing function, defined in 3.8 and 3.11, was set to = 3 $_s$, with $_s$ being the scale parameter given earlier in 3.9.

Texture Boundary Detection using K-Means Clustering

The final stage of the algorithm is to group the pixels into a number of clusters representing di↵erent regions. A standard naive k-means was used to group the pixels into two main clusters: tissue and non-tissue.

The standard algorithm was first proposed in 1957 by Stuart Lloyd of Bells Labs, and is implemented as a three step process. Given a dataset of observations, i = 1,2,3...n, having coordinates x_i,y_i, the algorithm initially assigns K random centroids, which may or may not be taken from the already existing observations. Once the centroids are set (say, C1 and C2 for a scenario with two clusters and K = 2), all the observations are assigned to one of the clusters based on the distance between them and the di↵erent cluster centroids. An observation is assigned to the cluster with whom it has the smallest distance. The distance for a particular observation i to centroids C1 and C2 is calculated based on a Euclidean metric which is defined as follows:

$$d_{i,C1} = p(x_i \quad C1_x)_2 + (y_i \quad C1_y)_2$$
$$d_{i,C2} = p(x_i \quad C2_x)_2 + (y_i \quad C2_y)_2 \tag{3.12}$$

Following the initial round of clustering, a new updated centroid is computed for each particular cluster, by taking the mean of all x coordinates and y coordinates of observations in that cluster:

$$C^x = \frac{\sum_{i=1}^{n} x_i}{n}$$
$$C^y = \frac{\sum_{i=1}^{n} y_i}{n} \quad (3.13)$$

With a new cluster centroid available, the algorithm repeats the process of calculating the distance between each observation and the new cluster centroids, and eventually updating again the centroids of the new clusters. When the cluster centroids remain unchanged following two successive iterations, the algorithm stops and the clusters are considered as final.

The algorithm for K-means clustering was implemented by adapting the standard K-means algorithm available within the MATLAB CVPR toolbox developed by Seo *et al.* [234]. In this case, the algorithm was required to find two main clusters: tissue and non-tissue, and therefore the number of clusters *K* was set to 2. This strategy was taken on the basis that the lumen inside the carotid artery filled with blood would contain di↵erent phase content to the adjacent tissue, and therefore it would be represented as having a di↵erent texture by the texture extraction algorithm.

The responses of the PDMs convolved with each of the filter banks were reshaped into one-dimensional vectors. The one dimensional vectors were then arranged into a matrix, such that each observation was described by F_N values in a hyper-dimensional space. Two observations were chosen at random as initial centroids in the said hyper-dimensional space, and a Euclidian distance metric was implemented to measure the distance of all other observations to these initial centroids. After clustering, the process was repeated as described in the preceding section, until no further updates to the centroids was achieved.

Figure 3.10 [Left], shows an example output.

Contour Extraction with Active Contours

After obtaining a clustered output from the K-Means algorithm, the result is in a binary format, allowing for the extraction the contour of interest. MATLAB includes built-in functions such as *bwboundaries* which allows for easy extraction of the contour along the edge of the closed space, in the middle of the clustered result. However, since the clustering result is not perfect, this method would produce a contour exhibiting localised concavities which does not correctly reflect the vascular structure. The author therefore opted instead to use an Active Contour algorithm, proposed initially by Williams and

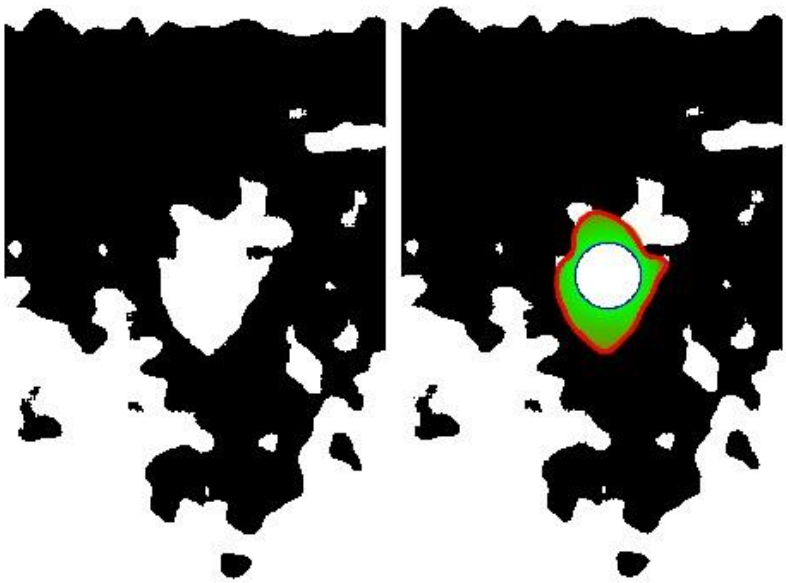

Figure 3.10: Left] Result from K-Means Clustering process; [Right] Contour Extraction with Active Contours.

Shah [236], to obtain a contour from the clustered result, since this includes a membrane term in its energy functions, allowing one to limit the contour from

evolving into such localised concavities. Additionally, this approach also allows one to employ a like-with-like performance comparison with other segmentation techniques in literature, such as that by Stoitsis *et al.* in [237].

An active contour is a type of deformable model which starts with a parametric contour represented by the equation $v(s) = [x(s),y(s)]$, where x,y represent the spatial coordinates of an image. The active contour (or snake) deforms and adapts itself by a dynamic process that minimises a global energy function, $E_{snake}(v)$, which is defined as follows [121]:

$$E_{snake}(v) = E_{int}(v) + E_{ext}(v) \tag{3.14}$$

where $E_{int}(v)$ represents an internal energy term and $E_{ext}(v)$ represents an external driving energy. The internal energy is normally formulated as a function of some physical constraints [121] related to elasticity and rigidity of the snake:

$$E_{int}(v) = \int_0^1 \alpha(s)|v'(s)|^2 + \beta(s)|v''(s)|^2 ds \tag{3.15}$$

where $\alpha(s)$ represents snake elasticity, which causes it to behave like a membrane, and $\beta(s)$ represents snake rigidity, which makes the snake behave like a thin plate. This combined internal energy limits the snake from bending and twisting excessively, such that it does not deform into localised concavities. The external energy $E_{ext}(v)$ drives the evolution of the contour on the basis of relevant image features such as edges, lines, terminations and corners [121]. The total external energy is therefore expressed as a weighted combination of three image functions as follows:

$$E_{ext}(v) = w_{line}E_{line} + w_{edge}E_{edge} + w_{term}E_{term} \tag{3.16}$$

where E_{line} is the image intensity, and therefore the sign of w_{line} makes the snake attracted to either dark lines or bright lines. The second term E_{edge} is defined as follows:

$$E_{edge} = \quad |\mathrm{r}I(x,y)|^2 \tag{3.17}$$

and makes the snake attracted to large image gradients. The third image function E_{term} represents the curvature of the level lines and thus makes the snake attracted to terminations and corners. The snake reaches equilibrium and stops evolving when the forces in equation 3.16 become balanced[121].

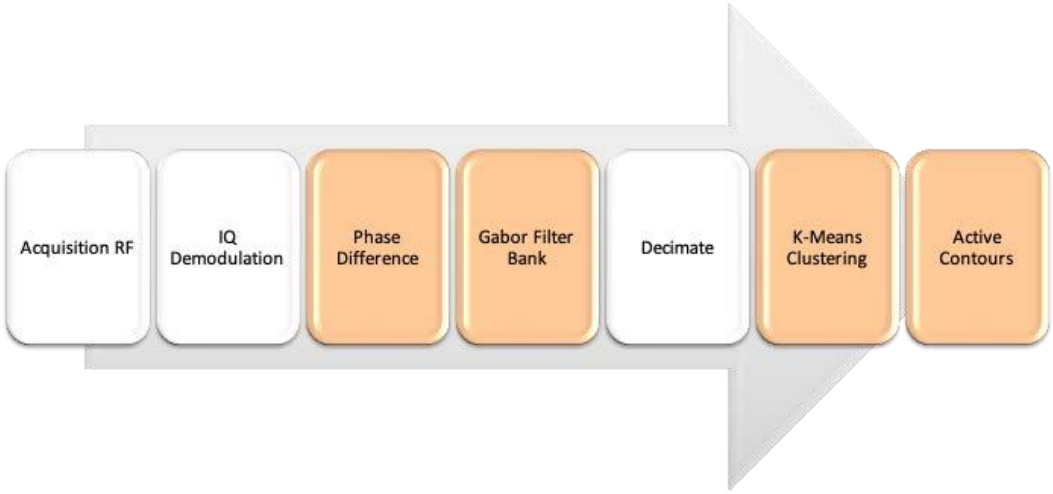

Figure 3.11: A block diagram showing the pipeline of segmentation using Phase Di↵erence Matrices.

In this work, the snakes algorithm was implemented using the MATLAB function developed by Kroon *et al.* in [238]. The terms (*s*) and (*s*) were set, after empirical trials, to 0.5 and 50 respectively. The coe cients W_{line},W_{edge}, and W_{term} were set to 0.01, 3 and 1 respectively. This allowed the snake to deform optimally towards the edges of the artery, clustered by the K-means algorithm,

without deforming into the local concavities. After the clustered image was obtained with K-means, the snakes algorithm required a seed point to initialise the starting node. This was manually initialised by providing the x and y coordinates of the centroid inside the artery.

An overall representation of the pipeline for the segmentation using Phase Di↵erence matrices is given below in Figure 3.11. In the next section, the pipeline for segmenting using Phase Congruency Maps is described, and later, both methods and their performances are compared.

3.2.4 Segmentation with Phase Congruency Maps

Another phase-based methodology of extracting features related to edges and lines is that based on Phase Congruency. In 1987, Morrone *et al.* [239] proposed a model of feature perception called the local energy model. The said model postulates that features may be perceived within an image, at the points where the Fourier components are maximally in phase. Many types of features result in points having high phase congruency, including lines, step edges and roof edges. It is shown therefore, that this model accurately explains the processes behind feature perception in the human eye.

As an example, a square wave and its corresponding Fourier series is considered, a diagram of which is shown in Figure 3.12. At the point of an edge transition or 'step', all the Fourier components of the constituent sine waves are exactly in phase at angles of 0° and 180°. Hence at these points, the phase congruency is said to reach a maximum value. More specifically, one notes that the phase is 0° at positive edges, and 180° at negative edges. At all other points in the square wave, the phase congruency will be low [228].

Edge Detection using Phase Congruency

In their work, Morrone et al. define phase congruency in terms of the Fourier series expansion of a signal for a location *x*:

$$PC(x) = \max_{\bar{\phi}(x)\in[0,2\pi]} \frac{\sum_n A_n \cos(\phi_n(x) \quad \bar{\phi}(x))}{\sum_n A_n} \tag{3.18}$$

where A_n is the amplitude of the *n*th Fourier component of a one-dimensional signal $I(x) = \sum_n A_n cos(\ _n(x))$, and $_n(x)$ is the local phase of the n^{th} Fourier component at position *x*. The value of $\bar{}(x)$, over which the equation is maximised, is the amplitude-weighted mean local phase angle of all Fourier com-

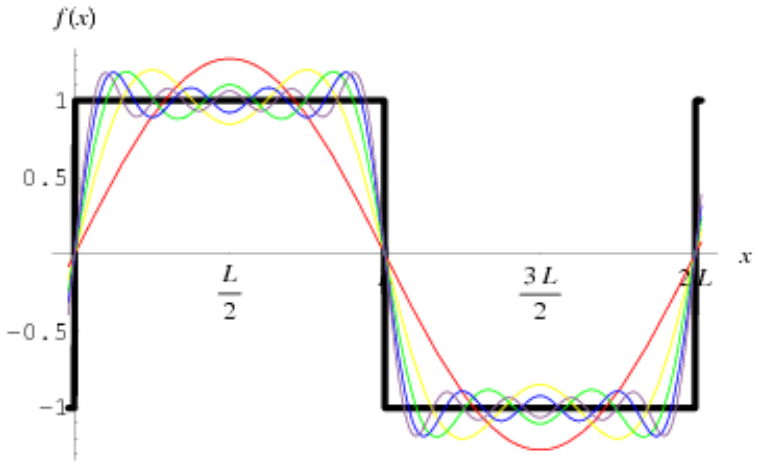

Figure 3.12: Construction of a square wave from its Fourier Series components. Reproduced with permission from [240].

ponents at the point being considered. Kovesi *et al.* [228] proposed a more convenient method of computing phase congruency via the local energy model, by convolving the signal with a filter bank of quadrature logarithmic Gabor filters. These allow for an arbitrarily large bandwidth filter to be constructed, while still maintaining a zero DC component in the even-symmetric filter. The log Gabor function has the following transfer function on the linear frequency scale [228]:

$$G(\omega) = \exp \quad \frac{(\log(\omega/\omega_0))^2}{2(\log(\kappa/\omega_0))^2} \tag{3.19}$$

where $!_0$ is the filter's centre frequency, and the term $?/!_0$ ensures a constant shape-bandwidth ratio by keeping it constant over varying $!_0$. The reader is referred to [228] for a more detailed treatise on the subject. Now for the analysis of two dimensional data, one must apply the one dimensional analysis over a number of separate orientations, and then combine the result to obtain a single measure of edge significance [225]. Kovesi *et al.* [228] suggested to construct a series of orientable 2D filters by spreading a Log-Gabor function into 2D. Thus, considering the one dimensional Log-Gabor filters defined earlier with geometrically increasing centre frequencies and bandwidths, one now masks these with an angular Gaussian tuned to a particular orientation $_0$ as follows [225]:

$$G(\omega_r, \phi) = \exp \left(\frac{(log(\omega_r/\omega_{r0}))^2}{2(log(\kappa/\omega_0))^2} + \frac{(\phi \quad \phi_0)^2}{2\sigma_\phi^2} \right)^{!} \tag{3.20}$$

where defines the standard deviation of the Gaussian spreading function in the angular direction. From these filters, Kovesi *et al.* [241] proposed to use the maximum moments of phase congruency as an indication of feature

significance. The maximum moment of phase congruency is obtained by computing the Phase Congruency Covariance Matrix for each point in the image as follows:

$$G = \begin{bmatrix} \sum PC_x^2 & \sum PC_x PC_y \\ \sum PC_x PC_y & \sum PC_y^2 \end{bmatrix} \tag{3.21}$$

where PC_x and PC_y are the x and y components of the Phase Congruency $PC(x)$, for each orientation. The maximum moment may be obtained by taking the maximum singular value of the covariance matrix G. This provides a Phase Congruency Map, an example of which may be noted in Figure 3.13b.

In this work, the phase congruency algorithm was implemented in MATLAB by adapting the two-dimensional algorithm developed by Kovesi and made available on his online repository at [242]. The algorithm convolves the input image with a set of orientable Log-Gabor filters with di↵erent wavelet scales, which are themselves masked with an angular Gaussian spreading function. In the example shown in Figure 3.13b, the Phase Congruency Maps (PCMs) are

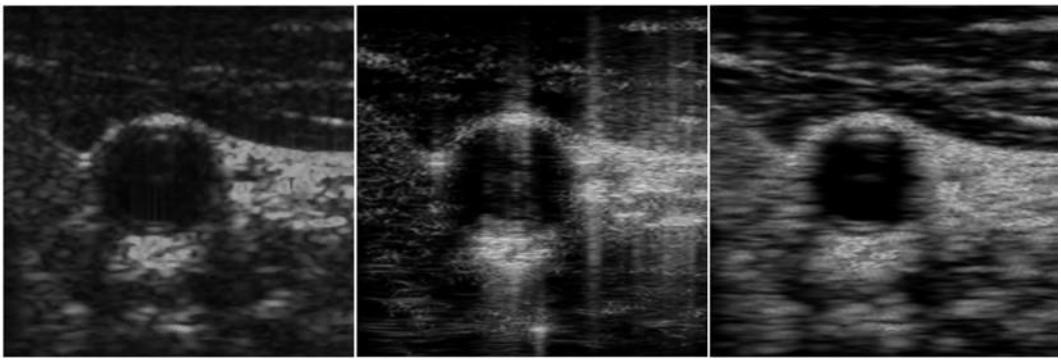

Figure 3.13: (a) A transverse B-Mode image of a carotid artery, (b) Corresponding Phase Congruency Map computed on envelope data, showing

vertical artifacts, (c) Phase Congruency Map computed on RF data, which is artifact free and amplitude invariant

computed from the 2D envelope signal, obtained using a Hilbert Transform on the RF data. Indeed, with the exception of the additional decimation and scan-conversion steps, this would be the equivalent of computing the Phase Congruency Maps on the B-mode images. The application of the Hilbert Transform in MATLAB however only provides a numerical approximation to the true Hilbert Transform, and as such elevates the local energy $E(x)$ of the signal as described by Boche *et al.* in [243]. This may be observed in Figure 3.14, where the root mean square (RMS) value of the RF line is 4284, whereas the RMS value of the corresponding envelope is 6058 (both shown in red lines).

This elevation in signal energy was seen to be reflected in the signal energy of the Gabor filter responses, once these were convolved with the signal, and also with the noise content of the signal. Now, the noise threshold proposed by Kovesi does take into account the energy of the signal. However, some leakage still occurs, resulting in the artefacts seen in Figure 3.13b. A possible solution to overcome such artefacts, would be to scale the noise threshold upward by a

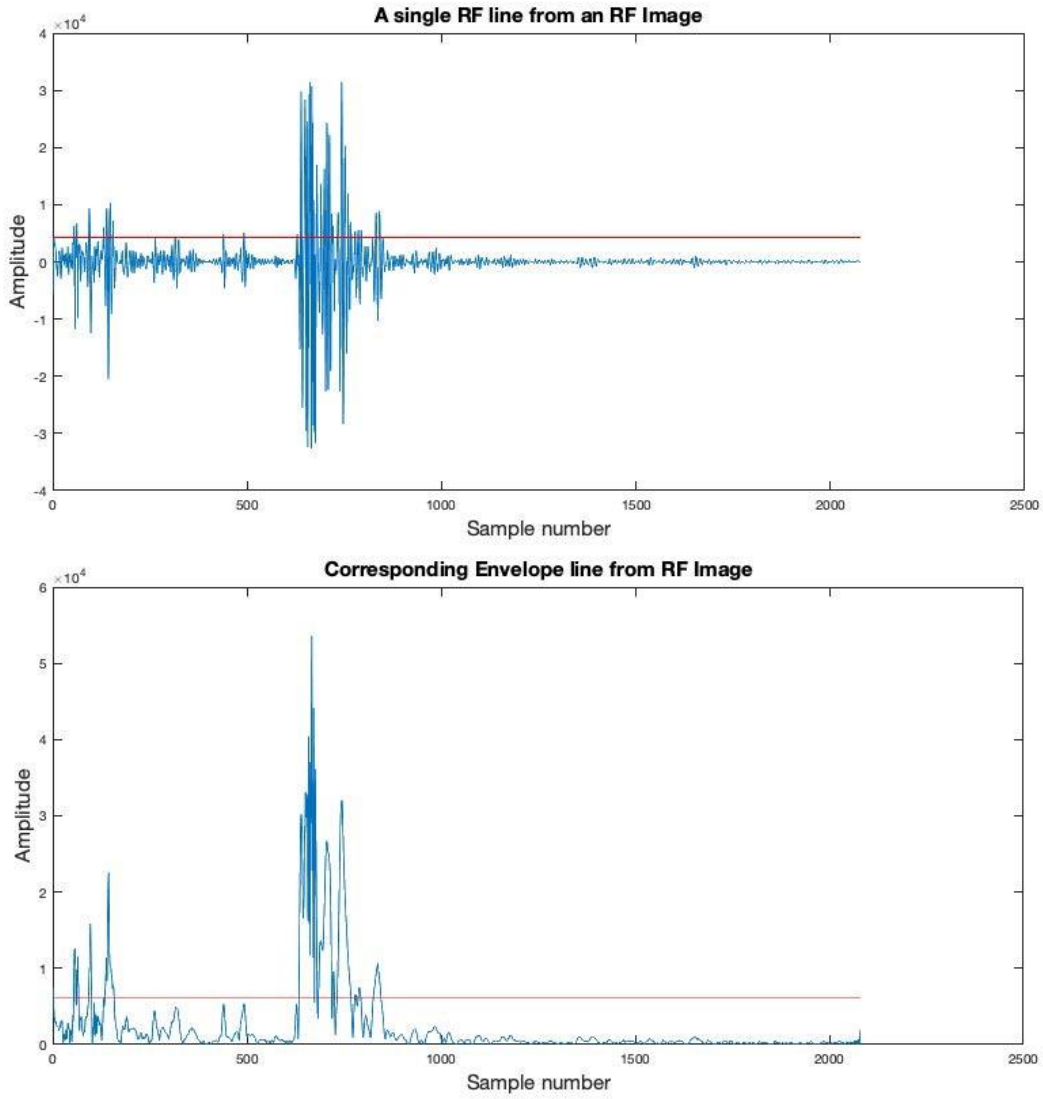

Figure 3.14: A sample RF line with RMSV: 4284 (top) and its corresponding Hilbert Transform with RMSV: 6058 (bottom) with elevated energy.

noise coe cient *k*. However, this would also have the undesired side e↵ect of omitting important edge feature information. Another solution is to increase the size of the wavelength of the smallest wavelet. If we assume that the noise spectrum is flat, then the Gabor filters will gather energy from the noise as a

function of their bandwidth, which in turn is proportional to the centre frequency. This means that the smallest wavelet has the largest bandwidth and accrues the most energy from noise. Conversely, increasing the size of the smallest wavelet also decreases the amount of noise that it gathers. The downside to this is that having the smallest wavelet with a large wavelength would mean that it interrogates less fine spatial detail.

A third possibility which allowed for keeping the wavelength small, but to also avoid increasing the noise threshold, was to apply the PCM directly on the raw RF data without extraction of the envelope, since it was empirically observed that this yielded lower local energy values and an improved signal-to-noise ratio. If we consider the frequency spectrum of a single 1D envelope signal, shown in Figure 3.15a, one observes, that the amplitude demodulating process shifts the frequency spectrum of the RF signal (Figure 3.15b) downwards to the left. The log Gabor filter applied subsequently, has a band pass filtering characteristic (shown overlapped in red), which does not a↵ect the content of the unshifted spectrum of the RF signal. Therefore, the principally desirable frequency content is allowed through. Conversely, when the envelope signal data is used, the subsequent log Gabor filter heavily attenuates the shifted principal low frequency content of this signal as shown in Figure 3.15a (with the filter shown overlapping in red). This allows only the noisier high frequency content to go through unattenuated, and this may present as potential artefacts seen in Figure 3.13b. This is confirmed by taking the inverse Fourier Transform on the filtered signal, after the frequency spectra of both envelope signal and RF signal were multiplied with the log Gabor filter in the frequency domain. In the time domain, the filtered RF signal had spurious peaks of much lower amplitude than the envelope signal. Additionally, the

resulting PCMs were void of artefacts as shown in Figure 3.13c, and these were thus the ones chosen for further processing.

The Log-Gabor filters were set to have a total of 2 di↵erent wavelet scales, with the smaller one having a wavelength of 2 and the larger one having a wavelength of 4.2. A total of 12 di↵erent filter orientations were used, such that the filters interrogated edges in various orientations with a fine degree of

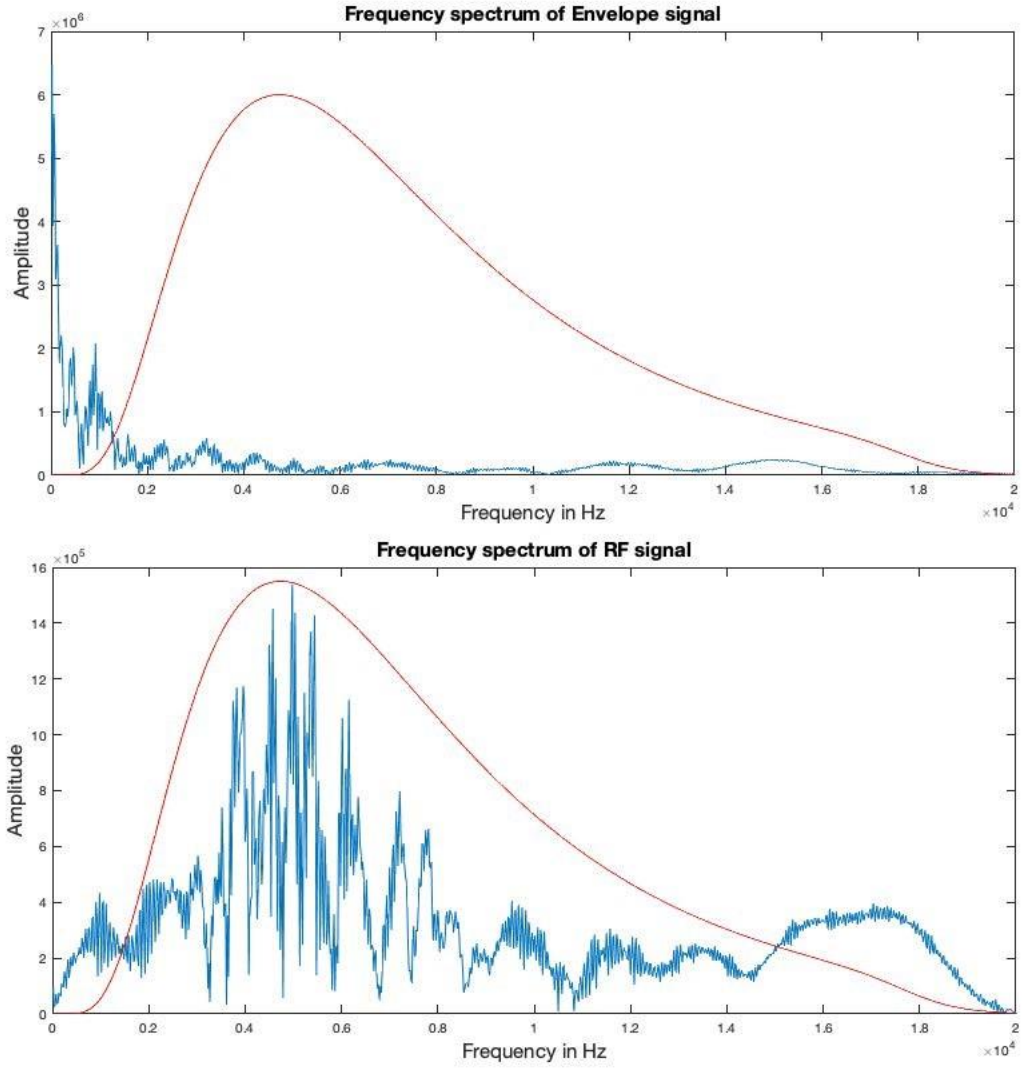

Figure 3.15: [A] Top Graph: Frequency spectrum of envelope signal. [B] Bottom Graph: Frequency spectrum of RF line. Gabor filter characteristic superimposed in red in both. Amplitude of filter modified for visual purposes.

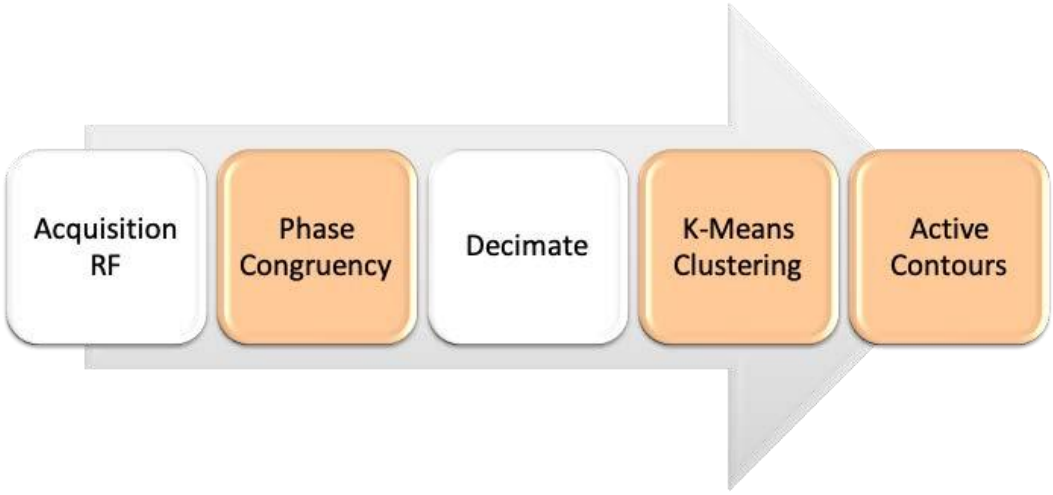

Figure 3.16: A block diagram showing the pipeline of segmentation using Phase Congruency Maps.

intervals. The experiments section shall describe how the number of wavelet scales and orientations are altered to determine change in performance in each case. The following additional parameters were set: *//f*, which refers to ratio of the standard deviation of the Gaussian, describing the log Gabor filter's transfer function in the frequency domain, to the filter centre frequency; *✓/* , which refers to ratio of angular interval between filter orientations and the standard deviation of the angular Gaussian function; and *k*, the previously described coe cient which refers to the number of standard deviations of the noise energy, beyond the mean at which the noise threshold point is set. These were respectively set to 0.55, 1.2 and 2.0 following empirical trials.

The algorithm returned a matrix array *M* which contained the PCM, that is, the maximum moment of phase congruency covariance which is an indicator of edge strength. The output was then decimated by a factor of 4, to bring the dimensions of the PCM in line with those of the B-mode images having undergone decimation. After obtaining a PCM as shown in Figure 3.13, this was passed through a K-means algorithm to obtain a binary, clustered image, and the contour extraction was implemented using the active contour algorithm, in an identical manner to that described for the PDMs in section 3.2.3. A block diagram showing the PCM segmentation pipeline is shown in Figure 3.16.

3.2.5 Experiment Construction

In this section, a series of experiments are described to compare performance of the same segmentation technique when using Phase Di↵erence Matrices (PDMs) or when using Phase Congruency Maps (PCMs). The experiments explore the e↵ects, or lack thereof, of varying key hyper-parameters when producing either PDMs or PCMs. The performance metrics used for comparison are also described. The results and their implications are then discussed. The experiments carried out were designed to compare the segmentation performance of the PDM and PCM based segmentation techniques between themselves and between a third reference technique reported in literature. Additionally, key hyper-parameters which control how the PDM and PCM based techniques behave, were altered, and the manner in which their segmentation performance changed was reported.

The overall pipeline of the PDM and PCM based segmentation techniques are similar. Phase Di↵erence Matrices or Phase Congruency Maps, were extracted from the input data, to generate an alternative ”representation” of the image.

The PDMs or PCMs were then fed into a *k-means* clustering algorithm, to cluster pixels into 2 relevant categories: tissue and non-tissue. The contour of interest from the binary image was then extracted using the Active Contour algorithm described previously. The third technique chosen for comparison was also based on the use of Active Contours, initialised however using a Hough

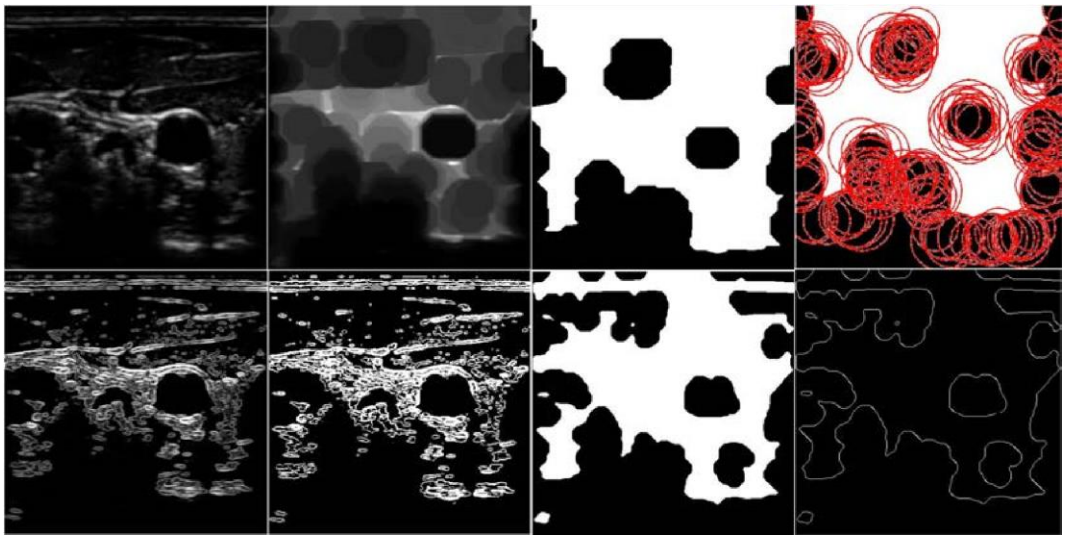

Figure 3.17: Top row; left to right: 1) Filtered B-mode image, 2) Morpholigical closing, 3) Image thresholding and 4) Hough Transform application. Bottom row, left to right: 1) Image gradients, 2) Image thresholding, 3) Morpholigical closing and 4) Sobel Edge detection.

Transform on the B-mode image, as proposed by Stoitsis *et al.* in [237].

In order to be able to compare performances across all techniques on the available dataset, the author replicated the algorithm of Stoitsis *et al.* as follows. Firstly, some image pre-processing was applied to the images, before using the Hough Transform to identify a number of candidate circles which

best fit the vessel spaces within the image. Thus, the ultrasound images were first normalised such that the average gray-scale median value of pixels in blood ranged between 0 and 5, whereas those of pixels in the adventitia ranged between 180 and 190. The images then had their area automatically reduced to a rectangular region of interest, in order to minimise the possibility of detecting unwanted structures. Details of this process are available in [237]. The regions of interest were then low pass filtered with a symmetric Gaussian lowpass filter to remove high-frequency noise. Morphological closing was then applied to close small 'channels' and 'openings' in the image. Edge detection was then carried out by applying a global thresholding procedure based on the histogram width of the image. The threshold was set at the 15% mark in the histogram. Edge points were then detected using a Sobel gradient operator [237]. The output of the Sobel detector produced binary images having 1s at the edges and 0s elsewhere. The Hough Transform was then applied on the binary images produced from the Sobel Detector, producing the said number of candidate circles. The centroid of largest / most dominant candidate was then used as a seed for an Active Contour algorithm, which deformed the the initial circular contour into the vessel's boundary, based on a gradient vector flow field. An example of the di↵erent pipeline processes is shown in Figure 3.17.

In the work by Stoitsis *et al.*, the authors use the Active Contours as a strategy to extract the contour of the carotid artery from the image, which would have ultimately been binarised thanks to the preceding pre-processing steps, involving the filtering, morphological closing and edge detection. When comparing this against the pipelines using PDM and PCM based segmentation, the final contour extraction step in this work was also carried out using an

Active Contours algorithm, in order to keep this element of contour extraction fixed across all techniques being compared. The seed point for the Active Contour was however manually initialised by choosing the necessary *x,y* coordinates. Therefore, the extraction of PDM and PCM data was intended as an alternative feature representation of the ultrasound images, a process which is analogous to the image pre-processing steps involving Gaussian Filtering, Morphological closing and edge detection used by Stoitsis *et al.*. The scope of the experiments is to compare the performance of this alternative representation / feature extraction process which is being compared.

In addition to the above, during the experiments key hyper-parameters which control how the PDM and PCM feature extraction behave were altered, and the change in segmentation performance in each case was recorded. In the case of Phase Di↵erence Matrices, after computing the phase di↵erence values in the axial direction from phase matrices, a bank of Gabor filters having set orientations and frequency scales were constructed and applied to the data for texture feature extraction. The number of filters in the bank were altered by first changing the number of filter orientations used, while keeping the choice of scale fixed at the default value. The filter orientations were equally spaced, and thus the spacing changed according to the number of orientations tested each time. Additionally, the performance changes when altering the number of frequency scales used was also tested, while keeping the choice of orientations fixed at the default value. The rationale behind testing these alterations was to determine the number of orientations and scales su cient to interrogate structural edges in the data at various angles and density. Too few would presumably miss out on important structural edge information, but having too many may not necessarily yield consistently increasing gains. Similarly, since

in the case of Phase Congruency Maps a bank of orientable Gabor Filters with di↵erent wavelet scales was also constructed, a similar principle during experiment construction was applied, and the performances obtained when varying the number of wavelet scales and number of orientations was tested, each time keeping the other variable fixed at the default value.

In a third set of experiments, the segmentation performance when using PCMs with di↵erent gain settings on the source RF data was tested. This was compared to a segmentation technique which used K-means clustering and Active contours directly on brightness-mode (envelope) data as opposed to PCMs. The scope of this set of experiments was to determine whether the objective of obtaining an amplitude-invariant image representation was achieved.

3.2.6 Performance Metrics

A number of evaluation metrics were used in order to quantify the performance of PDM and PCM based techniques in comparison to that using the Hough Transform initialised Active Contours. Each metric functions on the basis of comparison of pixel regions between the segmented images and a ground truth established manually by an expert. Thus, all the acquired transverse ultrasound images were manually traced with the assistance of two radiographers as described in section 3.2.2. The radiographers traced the MAB contours in the images, and this contour therefore established the demarcation point between the two important regions in the image (’tissue’ and ’non-tissue’ at this stage). The following metrics were then used for comparison.

DICE Similarity. The similarity between the segmented result and the ground truth is computed using the Dice Index which is calculated as follows:

$$\epsilon = \frac{2(Area(C_M) \cap Area(C_A))}{Area(C_M) + Area(C_A)} \tag{3.22}$$

where C_M refers to the manually produced contour and C_A refers to the contour produced by the algorithm. The Dice Coe cient of Similarity is used to gauge the degree of overlap between the areas subtended by the two boundaries C_M and C_A.

Sensitivity. The Sensitivity term, taken within the context of a clinical scenario, is defined as the ability of a method or test to correctly classify a sample as being diseased [244], or in other words, the probability of a test being positive when the disease is indeed present. It is defined mathematically as:

$$Sensitivity = \frac{TP}{TP + FN} \tag{3.23}$$

whereby *TP* refers to the number of True Positives or the number of pixels correctly identified as positive and *FN* refers to the number of False Negatives, or the number of pixels incorrectly identified as negative. A method which displays high sensitivity is considered reliable when it produces a negative result, since it rarely misdiagnoses instances where the disease is present.

Specificity. The Specificity term, is defined as the ability of a method or test to correctly classify a sample as being free from disease [244], or in other words, the probability of a test being negative when the disease is indeed absent. It is defined mathematically as:

$$Specificity = \frac{TN}{TN + FP} \tag{3.24}$$

whereby *TN* refers to the number of True Negatives or the number of pixels correctly identified as negative and *FP* refers to the number of False Positives, or the number of pixels which are incorrectly identified as positive. A method which displays high specificity may be considered to reliably exclude the presence of disease when this is in fact absent.

Accuracy. The Accuracy term is an overall measure of the correct predictions made by a method, be it both positive or negative predictions, expressed as a ratio of the total number of predictions made. It is mathematically defined as:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{3.25}$$

Accuracy is an intuitive and simple to report performance metric, but should be taken into consideration together with sensitivity and specificity, particularly if the number of false positives and false negatives is not symmetric in the dataset.

3.3 Results and Discussions

The first set of results are presented in the set of four graphs in Figure 3.18. These graphs show how, when using PDM features, the four performance metrics being measured vary with a di↵erent number of Gabor filter orientations ✓. The number of scales was kept constant at the default value of 12 in each case. The results show that an increase in the number of Gabor filter orientations does not yield an improved performance, since the best results for DICE coe cient, sensitivity and accuracy are obtained with 4 filter orientations. In the case of specificity, the variation in the number of filter orientations remains relatively constant at 0.998±0.001. In order to understand why this is so, the responses from convolving the PDMs with Gabor filters of various orientations

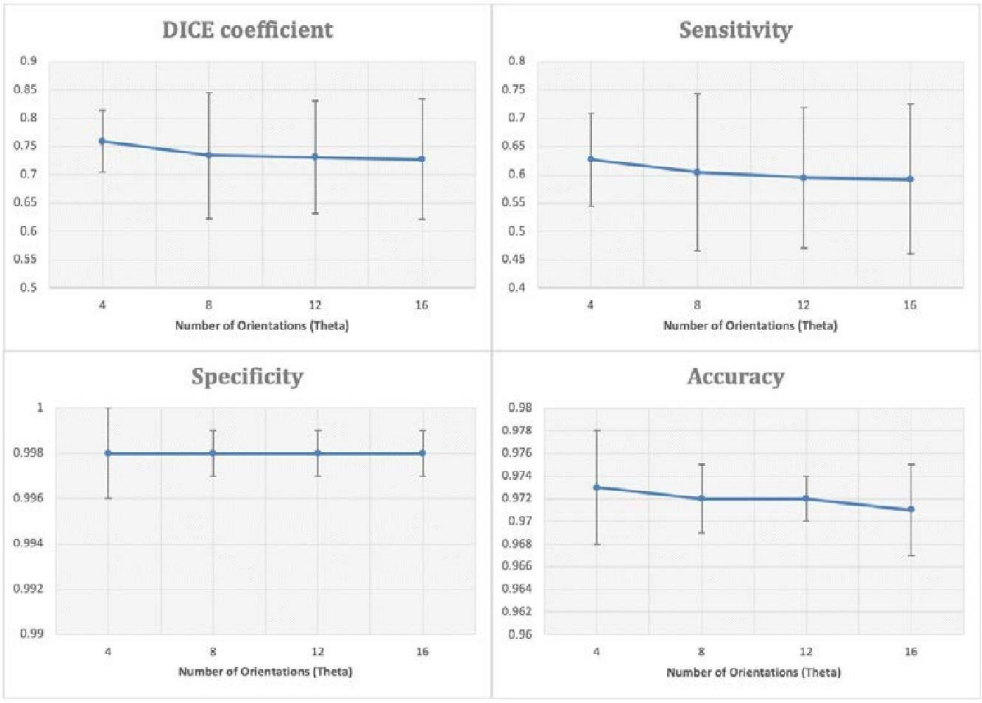

Figure 3.18: PDM results for varying number of orientations ✓. = 12

were plotted. An example is shown in Figure 3.19, which shows the responses to six di↵erent orientations, including amongst them the horizontal (0°) and vertical (90°) angles. From these, one notes that the vertical and horizontal responses are the ones which contain the sharpest structural features, whereas the diagonal responses include blurred and less distinguishable features. This leads to the understanding that the diagonal responses, at least at several different wavelength scales, may actually provide confounding influence to the k-means algorithm, causing it to mis-cluster pixels, and therefore in turn leading to a degraded performance.

In Figure 3.20, one observes a similar set of graphs this time showing the per-

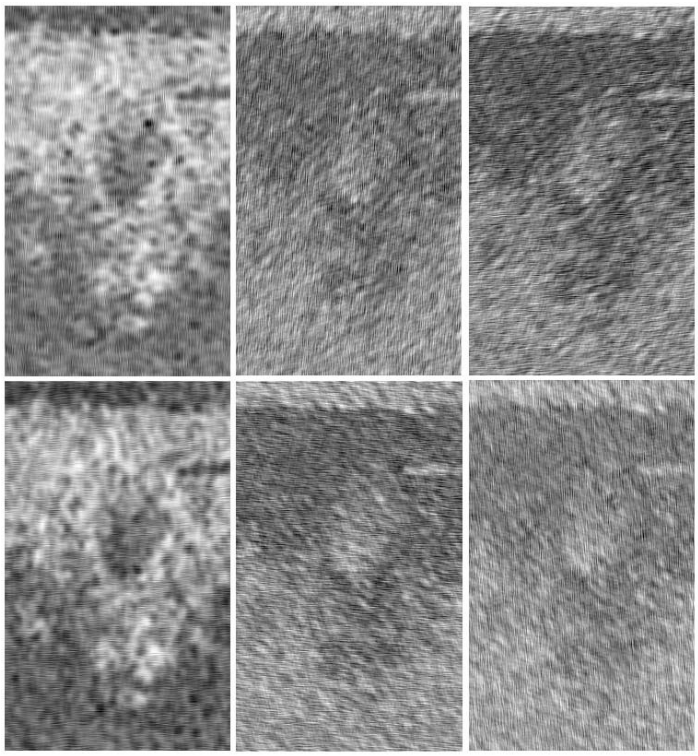

Figure 3.19: PDMs convolved with six di↵erent Gabor Filters having angles as follows: 1st column: 0°; 90°; 2nd column: 60°; 150°; 3rd column: 30°; 120° formance variation for di↵erent number of scales $_n$, while keeping ✓constant at the default value of 8 orientations. Each label on the horizontal axis indicates the number of wavelengths considered in each case. Here a general trend of performance improvement corresponding to a greater number of filter scales used may be observed. These results may imply that a greater number of filter scales impart more useful information to the k-means algorithm when it comes to clustering pixels into tissue or non-tissue, than does the number of filter orientations used. The range of values of chosen is determined by equation 3.11, and also entails a wider range of wavelengths considered.

The addition of di↵erent wavelengths, particularly the smaller ones, allow the filters to pick up finer structural detail. This is shown in Figure 3.21. In this figure, the four responses to the left pertain to a wavelength of 5.27, whereas

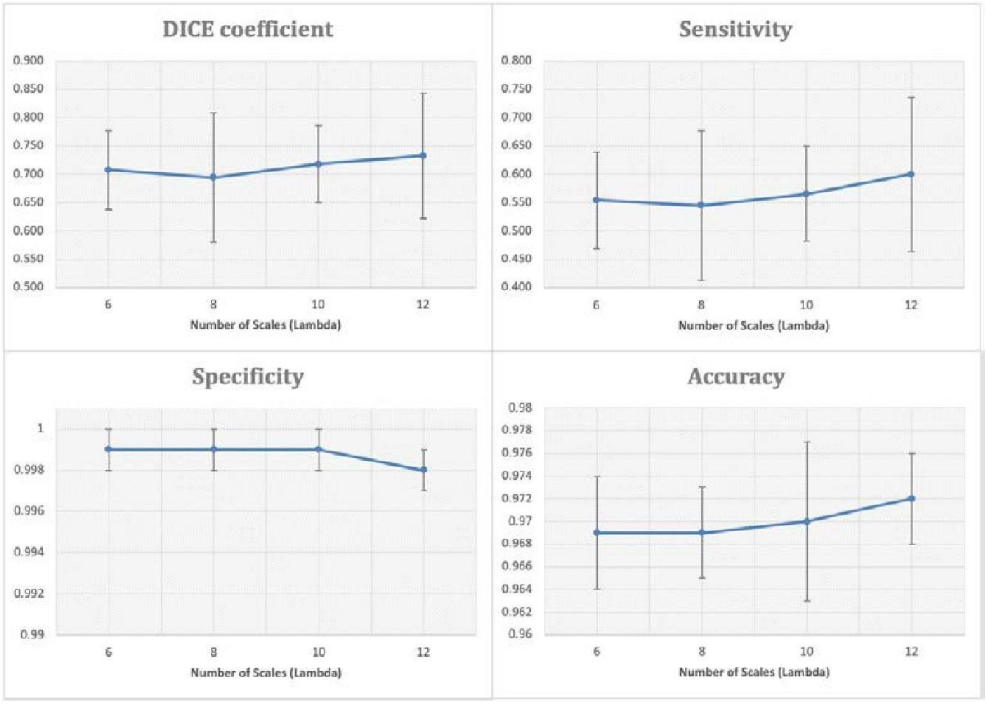

Figure 3.20: PDM results for varying number of scales $_n$. ✓= 8

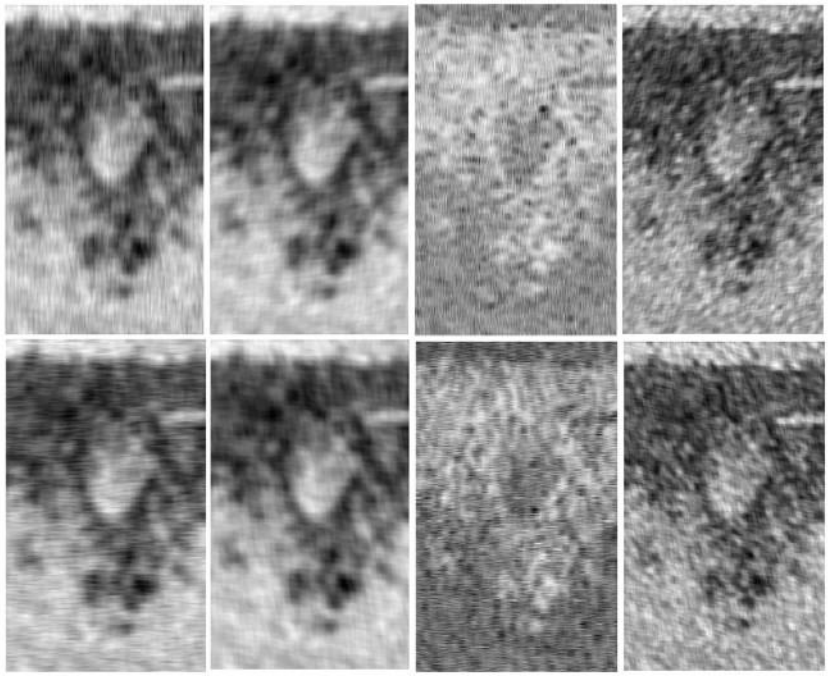

Figure 3.21: Filter responses with di↵erent scales. [Left Bank] Filter responses with scale of 5.27. [Right Bank] Filter responses with scale of 2.68.

the four responses on the right pertain the a wavelength of 2.68. The responses from the smaller wavelength show greater structural detail, at the cost of being noisier. The responses from the larger wavelengths lack the same detail, but their blurry nature gives broader context, to prevent the clustering algorithm from mis-clustering noisy pixels. It is also true however that increasing the number of wavelengths indefinitely, ultimately does not keep yielding improvements, with the performance levelling out in the end.

In Figure 3.22 and 3.23, graphs showing results obtained for PCM features are presented. As before, Figure 3.22 shows the performance on di↵erent metrics when the Gabor filter orientations used to compute the Phase Congruency

Maps are altered between 4 and 16 orientations. Here one observes a general trend of improvement in the DICE coe cient, specificity and accuracy as

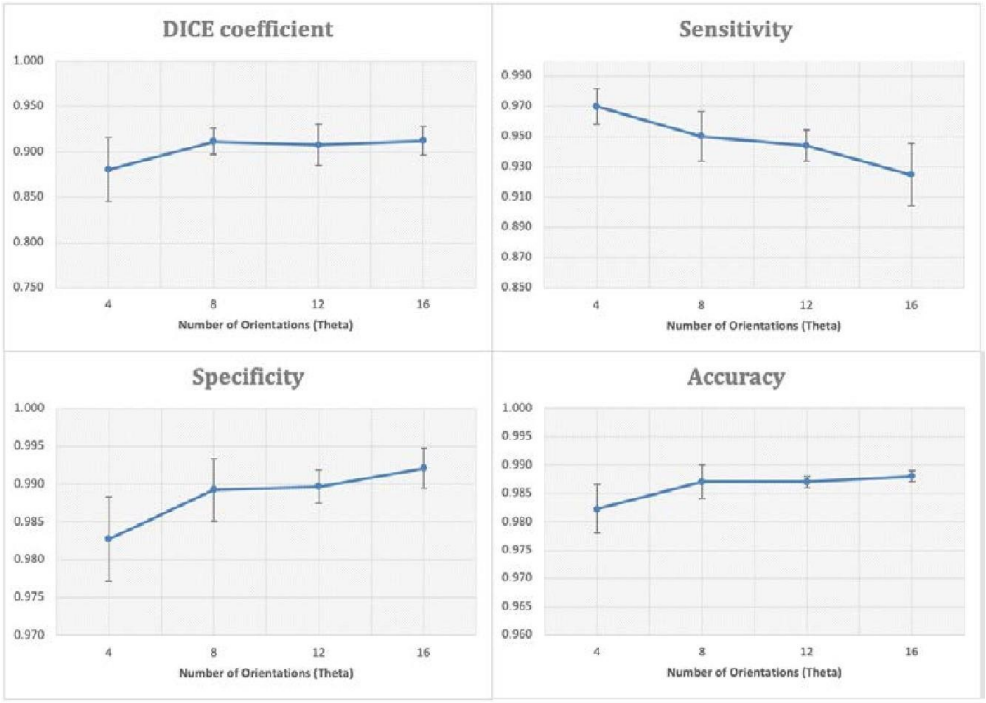

Figure 3.22: PCM results for varying number of orientations ✓. = 8

the number of filter orientations increase - but concurrently a trend of performance decrease in sensitivity. An analysis of the numerical data revealed a progressively smaller number of true positive and false positive pixels and an increasing number of false negatives, which indicates that the automatically delineated contour was decreasing in size while at the same time aligning itself better with the manual contour. This appears to indicate that with just 4 orientations, the boundary information provided to the k-means algorithm and thus to the Active Contour, were 'looser', and the automatic contour obtained was larger than the manual contour, having a large degree of overlap but also a large number of false positives.

In Figure 3.23, while keeping the number of orientations at their default value of 12 and varying the number of filter scales, one observe that the specificity and accuracy retain mostly similar values, and that the DICE coe cient and sensitivity register a downward trend in performance as the scale increases. An analysis of the pixel areas revealed a trend of decreasing area under the automatically delineated contour, and similarly a decreasing area of overlap. If one look at PCM images such as 3.13c, it may be observed that these are less noisy than their PDM counterparts, and are in fact fairly similar to B-mode images. With a larger number of filter scales, and thus the inclusion of finer wavelengths, the results appear to indicate that the Gabor filters picked up and accentuated non-relevant artefacts in the lumen, and thus caused the k-means algorithm to produce a more restrictive boundary to the Active Contours. For the results shown in Figures 3.20, 3.21, 3.22 and 3.23, a Holm-Bonferroni corrected paired T-test for DICE between scenarios with di↵erent number of orientations/scales showed that the di↵erence between results was not statistically significant ($p > 0.05$). While this may possibly be due to the lack of samples, it also indicates that the change in number of scales or orientation might have little overall e↵ect on the segmentation performance of the techniques.

In Figure 3.24, a comparison across techniques is presented by choosing the best performing combinations of hyperparameters for the PDM and PCM techniques, and pitting these against the technique proposed by Stoitsis *et al.*. From the results shown in this figure, it may be firstly noted that the performance obtained using the HT initialised Active Contours method is in agreement with that stated in the literature by Stoitsis *et al.* in [237], whereby the authors achieve a sensitivity of 0.850±0.040, a specificity of 0.990±0.003

and an accuracy of 0.980 ± 0.020. This therefore provides a good algorithmic baseline for comparison. Then, despite the poorer DICE coe cient and sensitivity of the PDM method, a DICE coe cient of 0.759 ± 0.055 quantitatively

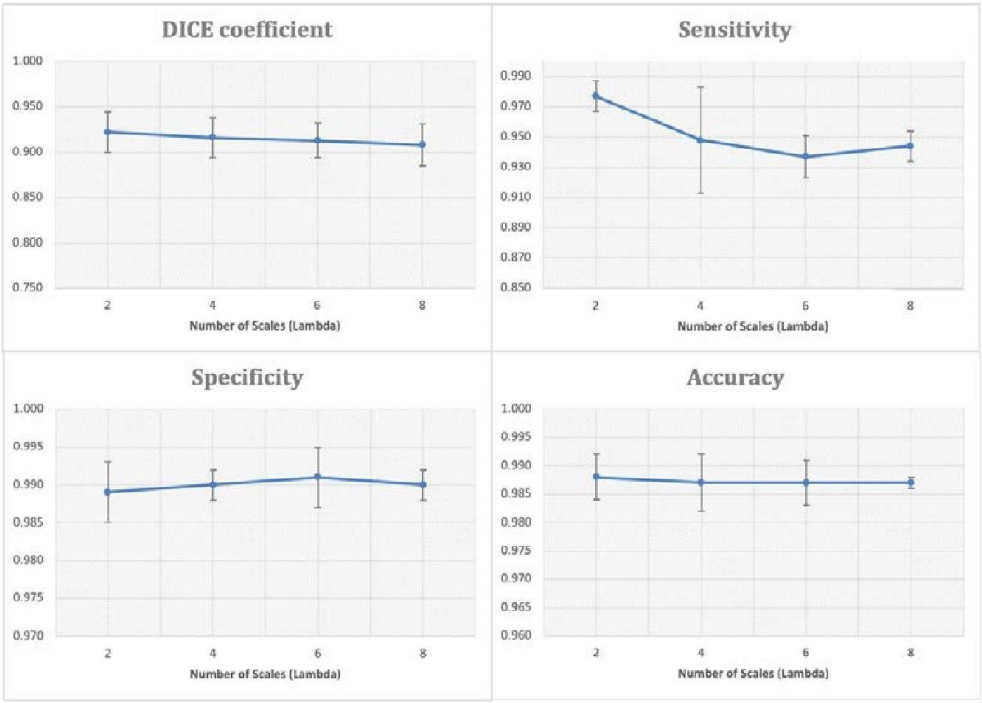

Figure 3.23: PCM results for varying number of scales . ✓= 12

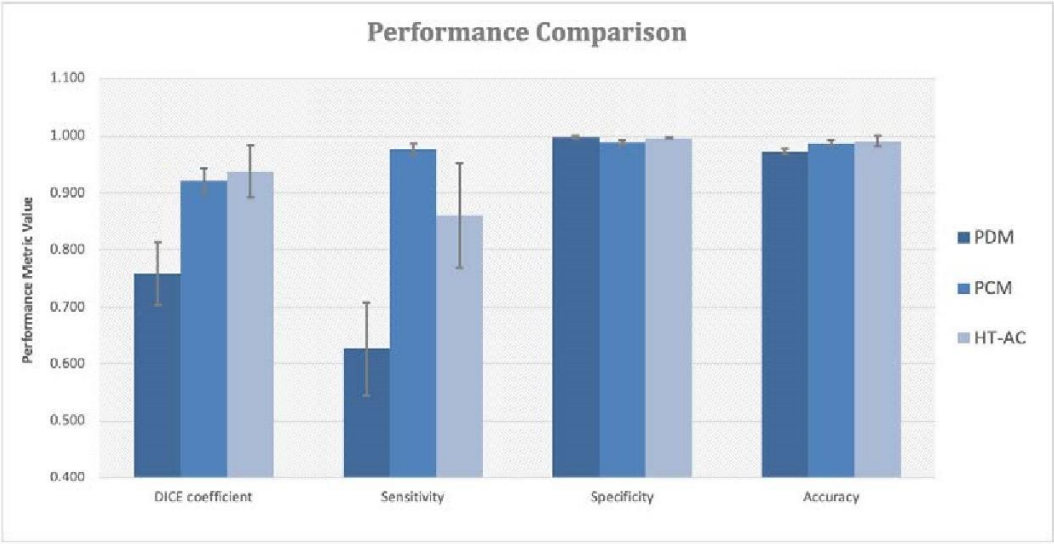

Figure 3.24: Comparison results across PDM, PCM and HT-AC techniques

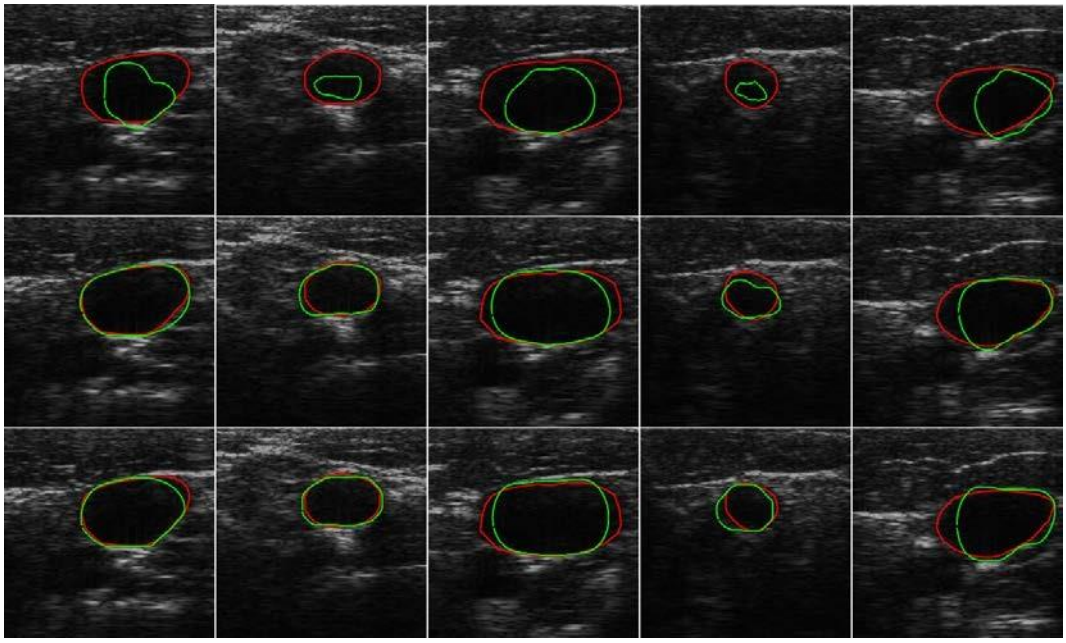

Figure 3.25: Qualitative comparison of segmentation results across PDM (top row), PCM (middle row) and HT-AC (bottom row) techniques, respectively for all 5 subjects

demonstrates that this method does extract an anatomically relevant result, and that phase information contains important structural information within the image. This successfully corroborates the claim made by Despotovic *et al.* in [227], in stating that phase di↵erence information extracted from RF data does indeed correlate to the presence or lack of tissue, and thus elicits structure which is contained within the image. The comparison of performance between the proposed PDM method and the Hough Transform-initialised Active Contour of Stoitsis *et al.* is published in the author's work in [90].

In the second proposed method based on PCM features, the performance is very much improved, with an average DICE coe cient reaching 0.922±0.022, a specificity and accuracy which are very similar to that of Stoitsis *et al.*

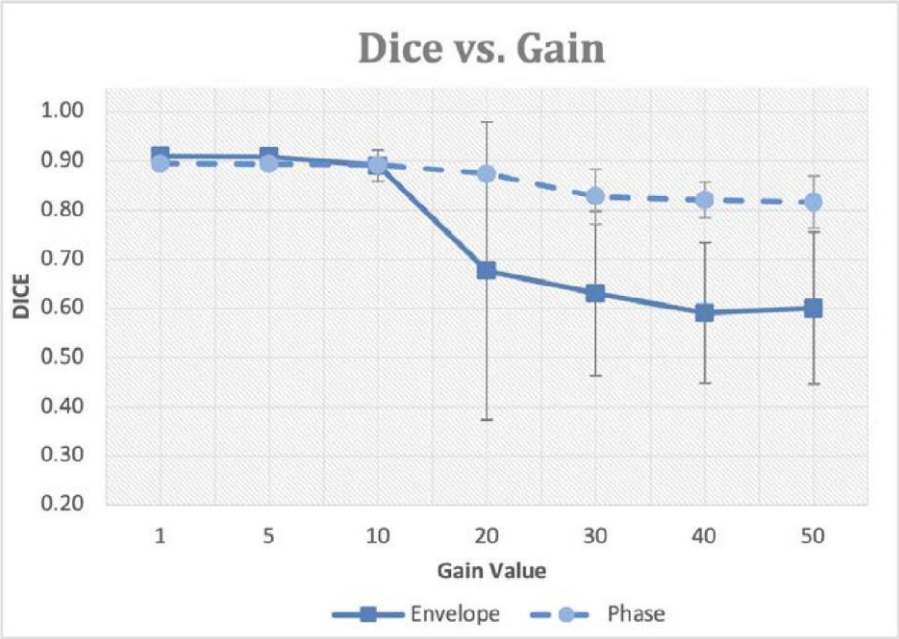

Figure 3.26: Comparison of segmentation results for Envelope and Phase data when progressively amplifying the source RF signals.

at 0989 ± 0.004 and 0.988 ± 0.004 respectively, and with a sensitivity performance which exceeds that of the other techniques at 0.977 ± 0.010. A Holm-Bonferroni corrected paired sample t-Test shows that there is a statistical di↵erence ($p < 0.01$) between the DICE values of PCM and PDM. However, the same corrected paired sample t-Test failed to show a statistically significant di↵erence ($p > 0.05$) between the DICE values of the PCM and HT-AC techniques. From a qualitative perspective, in Figure 3.25 one observes that although the PDM based technique obtains a DICE coe cient of 0.75, the

delineated result does not make anatomical sense. The PCM based technique however, with a DICE coe cient of 0.922 achieves significantly close results to those of the Hough-transform initialised Active Contours.

The benefits of Phase Congruency Maps are that, since they are based on phase information, this makes them invariant to changes in signal magnitude. Thus, unlike techniques based on intensity / envelope information, a segmentation technique based on phase information should theoretically be less a↵ected by an incorrect gain setting used by the clinician. The next set of experiments examines this hypothesis by varying the gain through a range of values, causing a progressively worse clipping of the RF signal, while concurrently plotting the DICE coe cient in each case to evaluate segmentation performance. The experiment was performed using PCM features, but also by using k-means on B-mode images derived from the amplified RF data. This was done to compare the e↵ect of gain across envelope and phase data. In Figure 3.26, the averaged experimental results from across 5 patients are plotted. The RF data was amplified by factors ranging from unity gain to x50. The performance obtained from envelope based segmentation was initially observed to be stable, while the RF data was still not experiencing clipping, with DICE coe cients of approximately 0.91 ± 0.01. Beyond a gain factor of x10, considerable clipping in the RF data started taking place, and the k-means algorithm followed by Active Contours produced progressively worse results, with the DICE coe cient decreasing down to 0.59 ± 0.14. As shown in Figure 3.28 [Top row], the clipping of the RF data caused heavy artefacts in the clustering of the k-means algorithm. In turn, the Active Contours produced smaller contours.

The situation is not reflected in the results for PCM features. While the initial performance was marginally worse than that of the envelope data (average

DICE of 0.89 as opposed to 0.91, gain factors of x10 and beyond decreased the performance to 0.82 ± 0.05. Holm-Bonferroni corrected paired statistical t-Tests showed statistically significant di↵erences between B-Mode and PCM techniques ($p < 0.05$) for gain values 30 and greater, but were not statistically di↵erent ($p > 0.05$) for gain values of 20 and below. In Figure 3.27 [Bottom

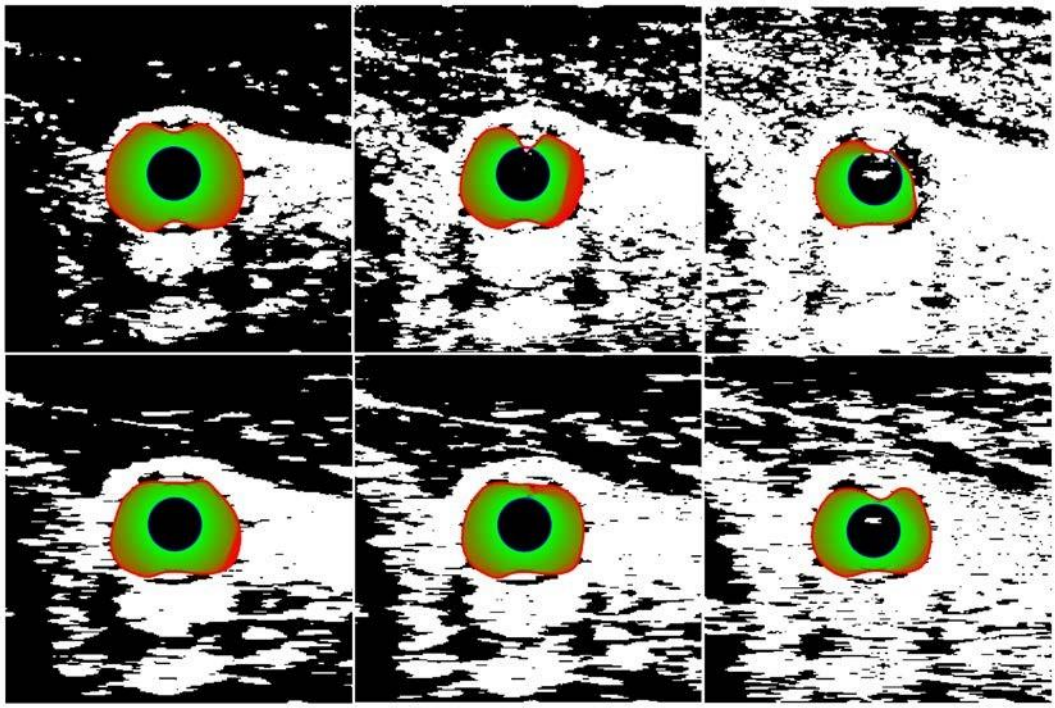

Figure 3.27: [Top row] K-means followed by Active Contours on envelope data; [Bottom row] K-means followed by Active Contours on PCM data; Gain settings of 5, 20 and 30 respectively in both cases .

row], one observes that while clipping still introduced artefacts in the PCM feature map, the e↵ect was much reduced, and the lumen of the artery was better preserved. The k-means and Active Contour algorithms extracted contours with better DICE coe cients than their envelope counterparts. This shows that when phase features are used in isolation, it is possible to achieve

delineations which correlate with the tissue structures in the image, and obtain segmentation performance which closely follows that of envelope-based methods. The phase information on its own generally produces poorer results than envelope data under normal circumstances, but has a greater resilience to amplitude variation.

3.4 Conclusion

In this Chapter, a novel application of Phase Di↵erence Matrices and Phase Congruency Maps to RF-based ultrasound images of the carotid has been proposed. It was shown that phase information contains important structural information regarding the carotid wall tissue. Initially, with PDMs, it was shown, at least qualitatively, that a general distinction between tissue and non-tissue exists in the phase information. With Phase Congruency Maps, it was proven that this phase information in isolation is enough to segment the Media-Adventitia interface of the carotid artery, with a DICE performance that is su ciently acceptable and close to that of amplitude based methods. In the next chapter, a novel fusion of both amplitude and phase information is proposed, using them together with deep convolutional networks to delineate the same Media-Adventitia interface in the carotid artery.

4

Deep Convolutional Networks for Automated Segmentation

4.1 Introduction

Deep Neural Networks have recently been employed with good success on various image segmentation tasks, object detection and image classification. In Chapter 2 one notes how they have also been picked up by the medical imaging research community for application to various clinical problems. In this chapter, the author proposes a novel segmentation technique which delineates the MAB in a fully automated manner, by being the first to apply a Deep Convolutional Network (DCN) to transverse Carotid Ultrasound data, and by it being applied on a novel fusion of amplitude and phase information at input level.

4.1.1 Aims

The aim of this chapter is to describe a set of experiments to examine the performance of the proposed network and fusion strategy, by comparing

against each other: a) the use of phase information alone as a datasource, b) the use of amplitude information alone as a datasource, and c) the proposed novel fusion of both amplitude and phase information. Furthermore - given that the deep FCN structure used is based on the encoder-decoder model, the structure of the network, including number of layers and the size of filters used, naturally has a bearing on the segmentation performance of the network. Thus, a set of experiments are further described to test, evaluate and report the performance of various network configurations, depths and filter sizes, in order to identify the best performing one. Finally, experiments are described to evaluate training time, as well as the e↵ect of input-level gain/amplification on the performance of the network. A discussion of results follows each experimental section described.

Novelty Statement: At the time of publication (2017), the novelty of this chapter's work was that it was the first Deep network based segmentation of transverse carotid ultrasound images, which included fusion of amplitude and phase information.

4.2 Methodology

4.2.1 Preparation of Data

The input data was acquired as described in section 3.2.2. Two separate datasets were acquired for these experiments, one with images of the carotid taken in transverse section and the other one with images taken in the longitudinal section. The images were manually cropped to a region of interest having size of 120 x 120 pixels centred on the carotid artery. The cropping was applied manually, since as described previously, most ultrasound platform

GUIs allow for a very simple windowing function, which would similarly allow the user to crop to a region of interest. Additionally, since for training deep FCNs, a large number of images is desired, an artificial augmentation of each dataset was implemented. Initially, the number of original images available included 50 2D slices taken from each of 5 patients, for a total of 250 images. Copies of the original images and associated output labels were created and rotated by 90, 180, and 270 degrees respectively. These were then appended to the original input and output datasets, creating a total of 1,000 input samples.

Apart from the input made up of B-mode images, a second dataset based on PCMs was created using the technique described in section 3.3.2. The PCMs were appended to the B-mode images - thus implementing an input-level fusion strategy by creating a two-channel dataset matrix of size: 120 x 120 x 2 x 1000 sample images, containing both amplitude and phase information. The output matrix of labels was similarly of size 120 x 120, and labelled such that pixels falling within lumen of the carotid were labelled as '1's and the rest as '0's. From all the images available, a random selection of 70% were kept as training data, 20% as validation data, and 10% as testing data.

4.2.2 Deep FCN Construction

The FCNs were constructed on MATLAB, using a deep network toolbox called MatConvNet [245], which contained pre-built functions to allow creation of deep convolutional networks of various layers and sizes. A number of networks were constructed with various configurations of depth, layers and filter dimensions, in order to search for a structure which yielded best results. This is described in a later section on experiment setup. All the networks were

however constructed as described in section 4.4.3, whereby the first half of the network was set up as an encoder with progressively decreasing size, and the second half of the network was set up as a decoder with progressively increasing size, until the output reached the same dimensions of the input.

The convolutional layers were set up with filter kernels having di↵erent square dimensions, but also with varying depths. Convolutional layers were followed by a RELU activation function. No padding was applied and the stride values during convolution were set to 1. The pooling layers were set up with a 2 x 2 averaging function in each case. The stride value during pooling was set to 2, such that with the averaging window of 2 x 2, the output of the pooling function was half the size of the input. At the middle intersection of the network, where the encoder and decoder meet, no pooling function was applied, and instead the convolutional layer fed directly into a convolutional transpose layer. In the decoder section of the networks, convolutional transpose layers were also constructed with filter kernels of varying dimensions and depth. No cropping was applied to the convolutional transpose layers, and each such layer was followed with an up-sampling function using a nearest neighbour method. The up-sampling functions scaled the output dimensions by a factor of 2.

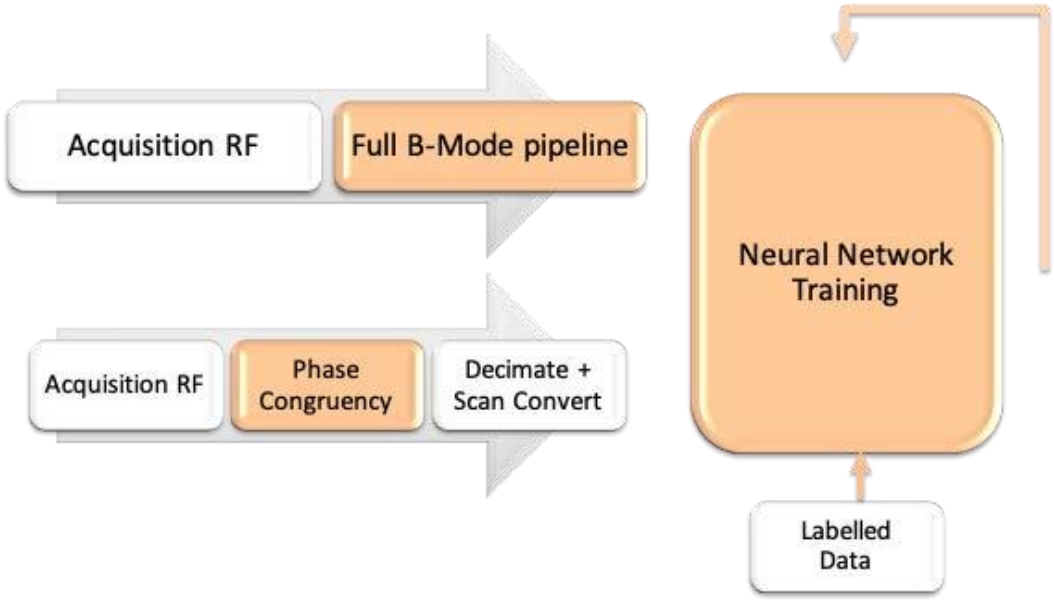

Figure 4.1: A block diagram showing the pipeline of training using Deep Neural Networks.

Output classification

At the output end of the decoder network, the number of output maps were reduced to two, and fed into a softmax classifier providing logistic regression for a two class problem. The softmax classifier maximises the the maximum value of outputs in one of the output maps, such that node pixels falling within the lumen region are close to '1', while node pixels outside the lumen are close to '0'. The e↵ect is inverted in the second output map, whereby the node pixels outside the lumen are close to '1' and the node pixels inside the lumen are made close to '0'.

Training environment

The FCN was built and trained on an Intel Core i7 machine, running additionally with a Geforce GT 650M GPU. The latter allowed for a CUDA enabled

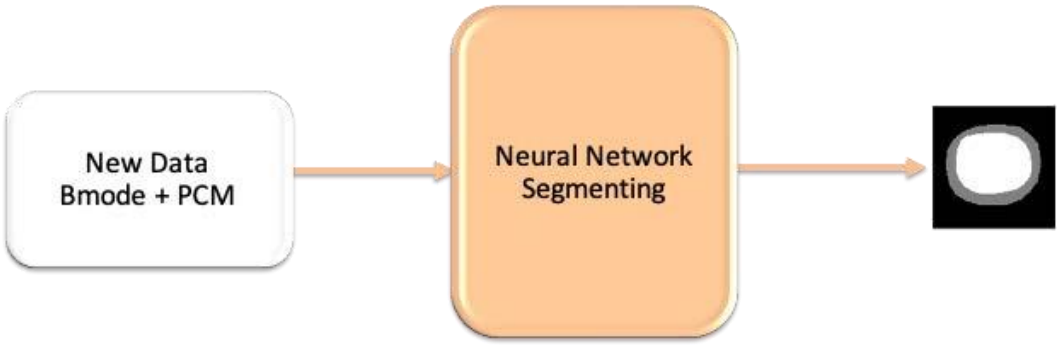

Figure 4.2: A block diagram showing the pipeline of testing using Deep Neural Networks.

parallel processing environment which improves training time. Training was allowed to run until convergence was achieved (typically at > 20 epochs), but where practicable in terms of training time, and for an equitably comparable environment, it was allowed to run for the same number of epochs each time a variable was altered. The optimisation algorithm used was Batched Stochastic Gradient Descent which trained on batches of 5 ultrasound images at a time. Figure 4.1 show a block diagram of the process, whereby the data acquired was used to create both B-mode data and PCM data, and the fusion of the latter was used as input training data to the network. Figure 4.2 shows the process of segmentation on new data once the network was trained.

The following additional hyper-parameters were set as follows: 1) the learning rate, , which is a coe cient multiplied by the gradient, was set to 1e-2. This

determines how aggressively the weights are updated between each iteration. Thus, a very low learning rate will not miss local minima, but might make training slow. A large learning rate may speed up training but could miss local minima and cause the network not to converge to the optimal value. The learning rate was also made smaller by a factor of 10 every 5 epochs. 2) The value of weight decay was set to 1e-4. This is a small coe cient which performs weight regularisation in order to avoid overfitting. It is multiplied by the sum of weights squared, and is intended to prevent the loss function from becoming too huge, while at the same time penalising the same loss function with the sum of weights for it not to become too complex. Lastly, 3) the value of momentum was set to 0.90. This hyper-parameter adds to each iteration's gradient calculation (and direction), a weighted average of past gradients from previous iterations. This obtains more stable directions of descent, particularly in ill-conditioned problems. From an experimental point of view, a detailed grid search to find the best performing combination of hyper-parameters was not considered feasible. Thus the values of these hyper-parameters were set after empirical trials, by altering them carefully one at a time and observing their e↵ect on training. Initially the values were set to small conservative values, which favoured stable performance over speed. These were then increased slowly up to the point where the network was converging faster, yet still stable. Finally, values were chosen, which in combination with the others, yielded the best results.

4.2.3 Experiment Construction

In this section, a series of experiments are described to compare the performance of di↵erent structures and configurations of deep convolutional

neural networks, each set to carry out semantic segmentation on a pixel-by-pixel basis. The purpose is to determine, from the wide search space of di↵erent network structures available, which would perform optimally for the task of segmenting the MAB boundary in transverse and longitudinal ultrasound images. The performance of the network is also compared to scenarios when it uses amplitude information alone, phase information alone, and when the input is a fusion of amplitude and phase information provided in the form of a 2 channel input matrix. Finally, the e↵ect of changes in gain in the source data are examined with respect to the performance of the network when using fused input data. The performance metrics used for comparison in all cases are described, and the results and their implications are later discussed.

Finding the optimal structure and configuration of the FCN for this segmentation task requires a quasi-grid-search approach to testing the available parameter space, coupled with some heuristics to aid the process. This was required given that the parameter search space is vast and di cult to test exhaustively in a practical manner. The choice of network used to commence experiments was the shallowest possible for an encoder-decoder type structure. Therefore for the choice of number of layer 'stacks', a minimum of 2 layers was used, whereby this signifies at least one convolution and pooling layer on the encoder side, and one up-sampling and convolutional transpose layer on the decoder side. The square sizes of filter kernels were also kept small, starting with kernels of size 3 x 3 pixels on the encoder side and 4 x 4 pixels on the decoder side. The number of output maps produced by the kernels (which shall be referred to as 'layer depth' within a stack) was kept at 8 initially. This provided what was considered to be a minimally shallow

network which was stable and converged. Structures smaller or simpler than what is described proved non-functional as they did not converge to a meaningful result. In some instances, the smallest combinations available for more than one parameter also did not yield stable results, and hence one of the parameters would have to be increased. With this particular shallow structure, the initial choice of hyper-parameters such as learning rate, weight decay and momentum, were set to their default values recommended with the MatConvNet toolbox. The values were then varied slowly one by one, observing the e↵ect (or lack-thereof) on the convergence, speed and performance. This process was repeated again later when the optimal network structure and depth was determined.

Several experiments were subsequently carried out to test the e↵ect of the di↵erent structural FCN parameters on the overall segmentation performance. These were: a) the number of layer 'stacks', that is: the number of [convolution + subsampling] or [transpose-convolution + upsampling] stacks used between the input and final segmentation mask; b) the number of input / output maps utilised within each layer stack; and c) the filter dimensions used within convolutional / convolutional transpose layers. Given that a full grid-search of parameter combinations yields a large and expensive parameter space to test, a heuristic approach was used whereby a single parameter was varied, while keeping the others constant. Once the best performing value for a particular parameter was identified, it was fixed, and the next parameter varied. The parameters were varied in the same sequence as described.

After identifying a network structure and configuration which yielded the optimal performance, this structure was retained and used for additional

experiments that compared the network performance when using di↵erent types of input data. Further experiments were thus carried out whereby the network was trained and tested solely on amplitude (B-mode) data first. The network was then trained and tested solely on phase information by using phase-congruency maps described earlier in Section 3.3.2. Finally the network was trained and tested on a dataset which fused both amplitude and phase information. A 2-channel input matrix was constructed using 120 x 120 B-mode images as the amplitude channel and the 120 x 120 phase congruency maps as the phase channel. The phase channel initially had values ranging from -1 to 1, since this was the range of outputs yielded from the phase congruency extraction algorithm. In order to bring the magnitude of values on par with that of the amplitude data, and thus avoid a 'magnitude imbalance', this was normalised to a range of between 0 and 255.

As described in Section 3.2.2, for these experiments, a subset of the full dataset had been collected up to this point, with the same data acquisition protocol described previously. Experiments described in later chapters included a much larger dataset, and therefore resulted in di↵erent performances. Therefore, at this point, a total of 250 transverse and 250 longitudinal images were obtained from 5 subjects (50 images each) with normal (non-stenotic) carotids and with ages spanning 25 - 40 years. Images were manually and independently traced by two radiographers. Each radiographer traced the image sets twice, with a period of 2 weeks in between sessions. The MAB was traced in both transverse and longitudinal sections. During each experiment, a hold-out test set that the training process never touches was retained for testing. All experiments described were repeated and reported separately for both transverse carotid images and longitudinal carotid images.

4.2.4 Performance Metrics

Two main performance metrics were used in order to quantify the performance of the network while varying structural parameters and while testing di↵erent inputs and fused inputs. The first metric was the DICE coe cient of similarity described previously, which is used to gauge the degree of overlap between the areas subtended by the two boundaries C_M and C_A, where C_M refers to the manually produced contour and C_A refers to the contour produced by the network. The second performance metric used was the Modified Hausdor↵ Distance, which is described as follows:

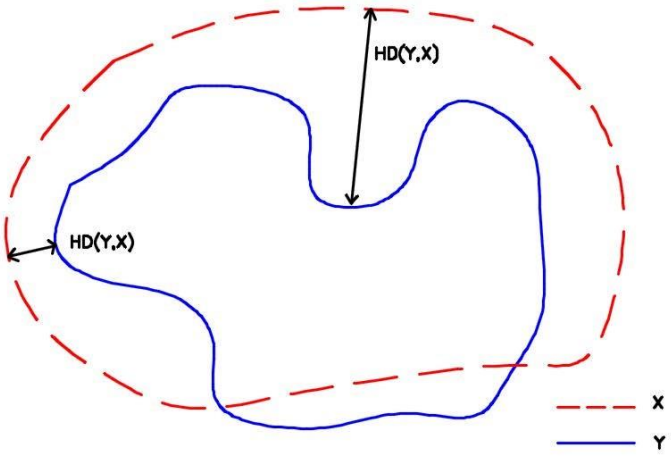

Figure 4.3: A graphical representation of the Hausdor↵ Distance between two point sets X and Y.

Modified Hausdor Distance. The Hausdor↵ Distance (HD) is a measure of distance between two point sets as shown in Figure 4.3. It measures the largest distance present between a point in one set of points (say, a contour) to the

closest available point in another set. The Modified Hausdor↵ distance (MHD) [246] on the other hand finds the mean distance between two boundaries and is computed as follows:

$$MHD = max(d(A,B),d(B,A)), \tag{4.1}$$

where

$$d(A,B) = \frac{1}{N_a}\sum_{a\in A} d(a,B \tag{4.2}$$

)

Since the HD is a measure of the largest mismatch between two boundaries, a single point can cause a high HD value. The MHD mitigates this e↵ect.

4.3 Results and Discussion

For the first set of experiments which examine structure and configuration of the network, both DICE coe cient of similarity and the Modified Hausdor↵ Distance were used as metrics of performance. The first experiment carried out was that of varying the number of stacked layers in the encoder and decoder structure. Figure 4.4 shows two graphs, whereby the DICE value is shown on the top graph and the MHD is shown on the bottom graph, in both cases on the y-axis. The number of stacked layers on the x-axis in both cases represents the total number of stacked *pairs* of layers, i.e. the number of (convolution + pooling) layers and the number of (up-sampling and convolutional transpose) layers. The number of stacks on the encoder side is always equal to the number of stacks on the decoder side. This means for instance, that a total number of 4 layers would mean 2 pairs on the encoder side and 2 pairs on the decoder side. The graphs also show superimposed results for both transverse carotid images and longitudinal carotid images. In these experiments, the filter sizes had to be set to a minimum of 5 x 5 on the encoder side and 6 x 6 on the decoder side to produce meaningful results.

From the results, one may observe that, given that the other parameters such as output map depths and filter sizes had not been optimised yet, the performance on both DICE coe cients and MHD peaks at a total number of 4 layers in the network. For the DICE coe cient, the performance peaks at 0.758±0.041 for the transverse images and at 0.921±0.024 for the longitudinal images. A, Holm-Bonferroni corrected paired T-test for DICE between di↵erent number of layers showed that the di↵erence between results was statistically significant ($p < 0.05$). Conversely, with the MHD showing best performance when the numbers are small, one similarly notes that with 4 layer stacks the performance with transverse images is that of 1.262 ± 0.293 mm and that of longitudinal images is 0.272 ± 0.071 mm. These first set of results must be treated with caution, since the other parameters of the network such as output map depth and filter size were still restricted. Thus, while the results provide an indication of the level of depth in terms of layer stacks the network requires to start producing meaningful results, they are not to be considered definitive for the task at hand.

In Figure 4.5, one observes the results from a second set of experiments carried out, this time for the instances when the number of output maps at each of the layer stacks were progressively varied. Again the vertical axes in these graphs show the DICE (top) and MHD (bottom) performance respectively. Since the number of output maps at each layer were di↵erent, and since it is di cult to visualise all the combinations of output maps being used on the x-axis, a 'summarised' representation is shown, whereby the graph only shows the number of output maps present at the 'middle set' of layers in the network. Thus, on the x-axis, a value of '8/4' signifies that on the middle set of layers of a 4 layer network, the number of output maps being produced by the

(convolution + pooling) layer pair is 8, and that these are fed into the next (up-sampling and convolutional transpose) layer pair.

Additionally, on the basis of the previous results concerning the best performing number of layer stacks, the experiments are started o↵ with a network having a minimum of 4 total layer stacks (2 on the encoder and 2 on the decoder), and worked upwards to explore also whether the performance changes when the output maps or number of stacks are increased. As evident from Figure 4.5, with a minimum of 4 stacks, increasing the number of output maps at each stack to a maximum of 64 at the centre stack, improves the perfor-

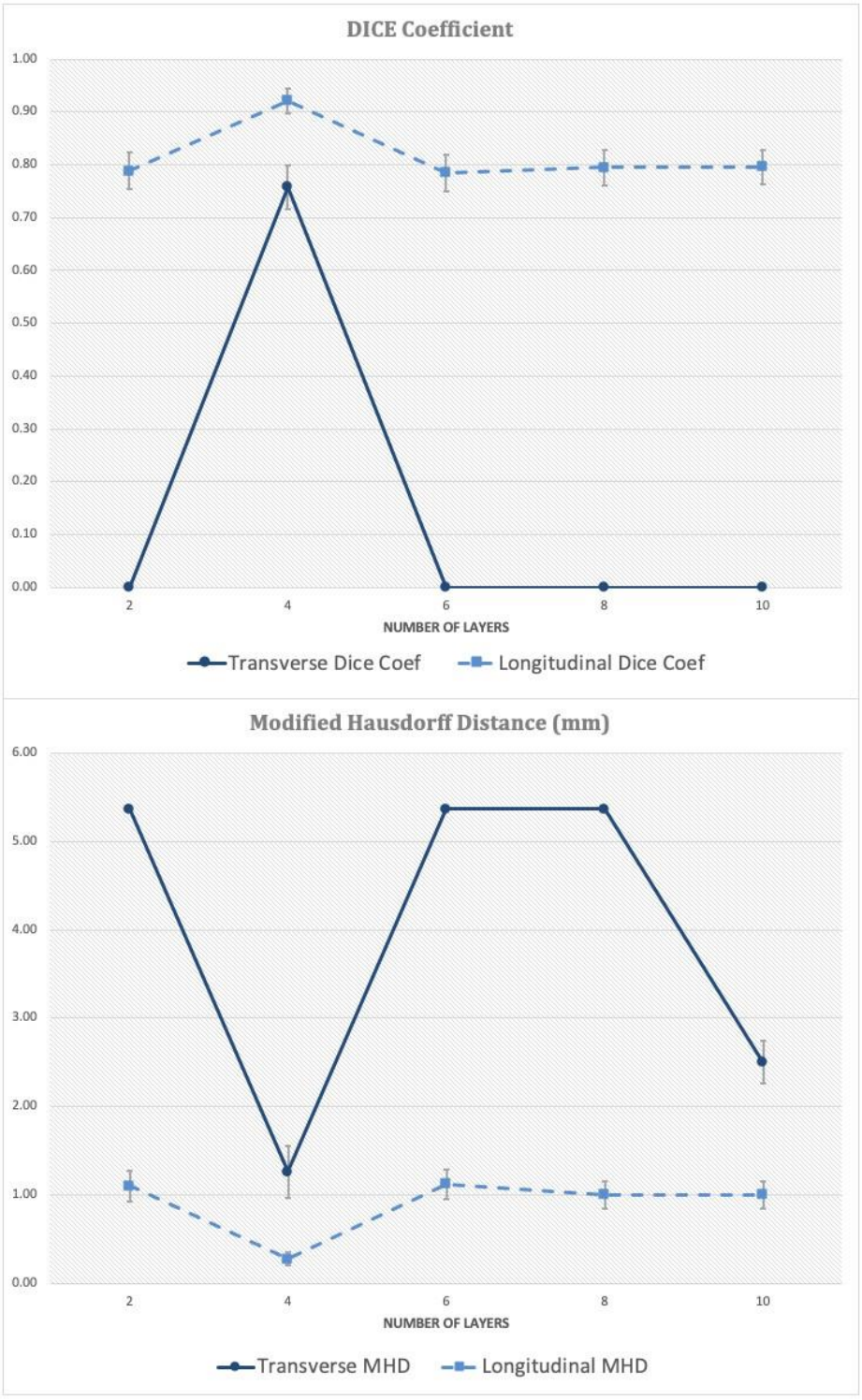
DICE Coefficient
1.00
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
2
4
6
8
10
NUMBER OF LAYERS
Transverse Dice Coef
Longitudinal Dice Coef
Modified Hausdorff Distance (mm)
6.00
5.00
4.00
3.00
2.00
1.00
0.00
2
4
6
8
10
NUMBER OF LAYERS
Transverse MHD
Longitudinal MHD

Figure 4.4: Results showing performance of networks of various stacked layers. mance consistently before trailing o↵ when increasing to 128 output maps. A further performance improvement was observed when the number of stacks are increased to 6 in total with maximum of 128 output maps at the central stack. With this configuration, DICE coe cients for transverse and longitudinal images are 0.968±0.006 and 0.987±0.007 respectively, while MHD performances are 0.104±0.037 mm and 0.056±0.044 mm respectively. Increasing the size of layer stacks to 8 with a maximum of 64 or 128 maps was found to to take too long to converge for a similar performance. A similar outcome was observed for increasing the maximum number of output maps to 256 at the central stack. Corrected paired T-tests failed to show statistically signifiant di↵erence in performance ($p > 0.05$) for the longitudinal images. However, a statistically significant di↵erence ($p < 0.05$) was present for transverse images, between using 128 output maps with 6 layers and the previous scenarios.

In Figure 4.6, one notes the performance results obtained when varying the size of the filter kernels on both encoder and decoder segments. For brevity, the graph shows on the x-axis the square filter size at each layer stack using a single digit. Hence for instance, the label '3, 2, 2, 2, 3, 4' shows 6 single digits, one for each layer of the network, whereby the first layer had filter dimensions of 3x3, the second layer had filter dimensions of 2x2, the third also had filter dimensions of 2x2, etc. Again, on the basis of the best performances obtained from the previous experiments, the number of network stacks was fixed to 6, and the number of output maps to 128, while varying only the filter dimensions as shown on the x-axis.

As observed in Figure 4.6, one notices a consistent increase in performance with increase in filter size, with gains levelling out at filter dimensions of 9, 9, 7, 7, 10, 10. With these filter dimensions, the network obtained DICE co-

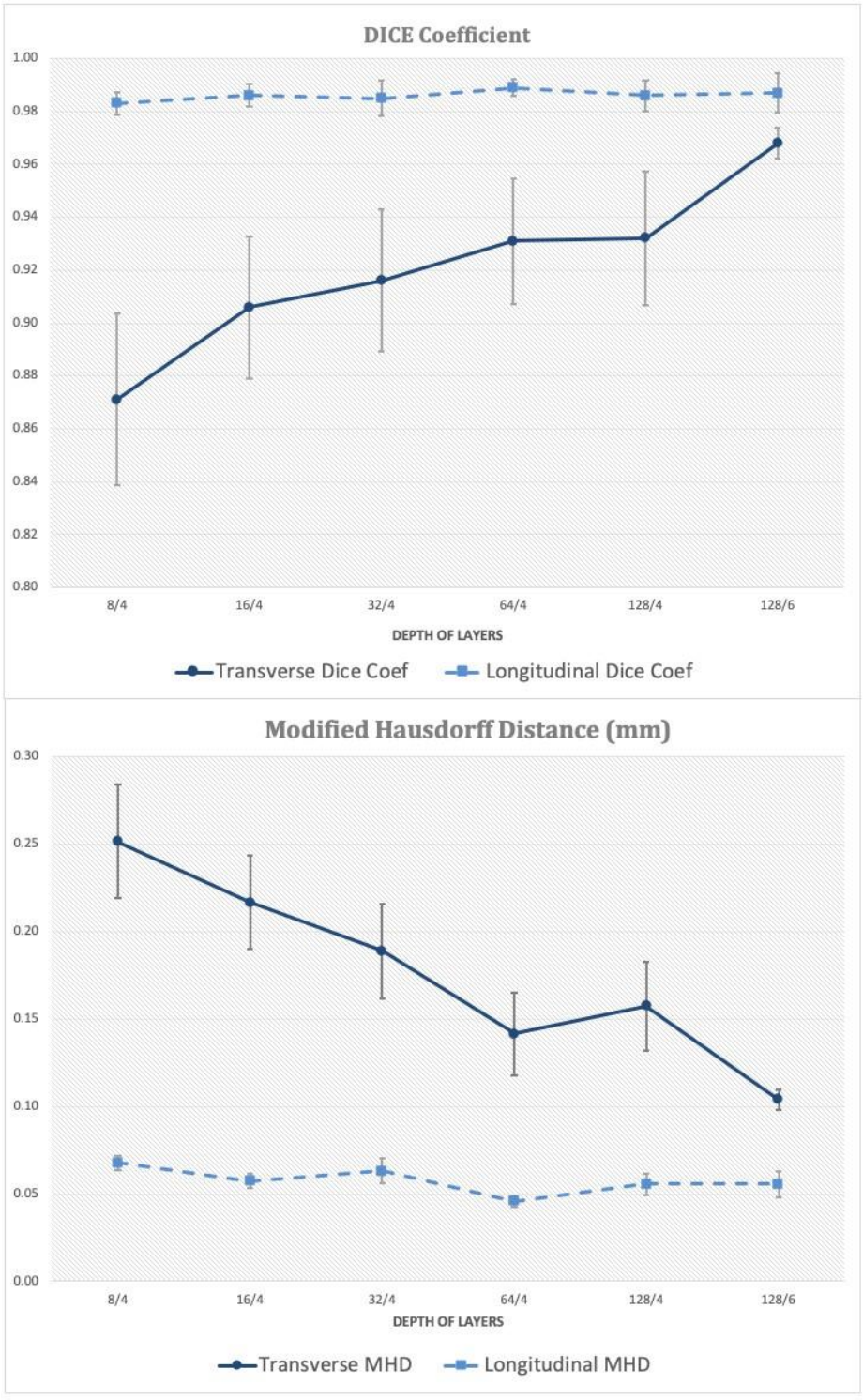

Figure 4.5: Results showing performance with various output map depths.

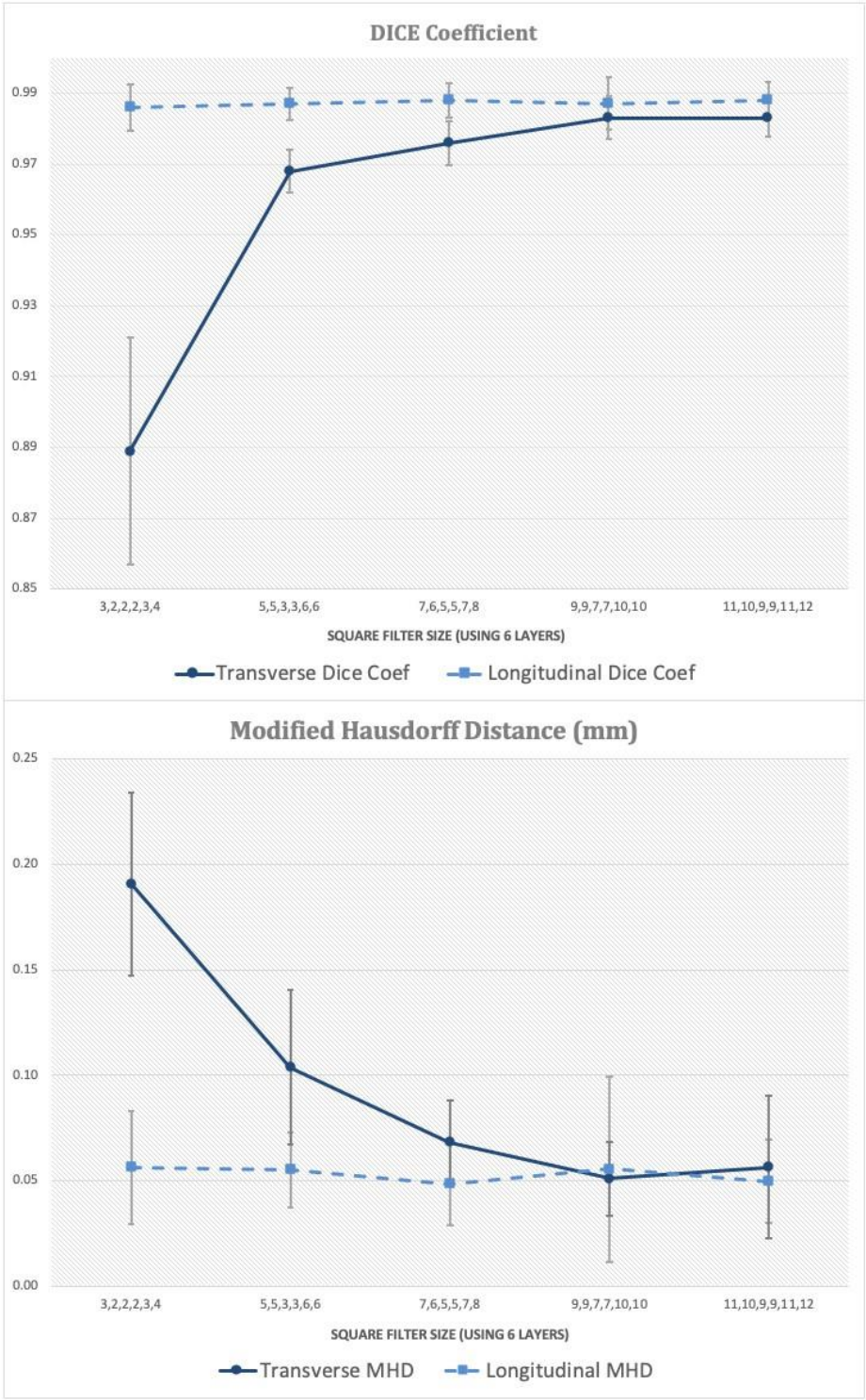
DICE Coefficient
0.99
0.97
0.95
0.93
0.91
0.89
0.87
0.85
3,2,2,2,3,4
5,5,3,3,6,6
7,6,5,5,7,8
9,9,7,7,10,10
11,10,9,9,11,12
SQUARE FILTER SIZE (USING 6 LAYERS)
Transverse Dice Coef
Longitudinal Dice Coef
Modified Hausdorff Distance (mm)
0.25
0.20
0.15
0.10
0.05
0.00
3,2,2,2,3,4
5,5,3,3,6,6
7,6,5,5,7,8
9,9,7,7,10,10
11,10,9,9,11,12
SQUARE FILTER SIZE (USING 6 LAYERS)
Transverse MHD
Longitudinal MHD

Figure 4.6: Results showing performance with various filter kernel sizes. e cients of 0.983 ± 0.005 and 0.988 ± 0.005 for transverse and longitudinal images respectively. It also obtained MHD values of 0.051 ± 0.018 mm and 0.056±0.044 mm respectively. With the next set of filter dimensions of 11, 10, 9, 9, 11, 12, very minor improvements of just 0.001 were observed in the DICE coe cient of longitudinal images, and a decrease in performance of 0.006 mm in the MHD of transverse images. Thus considering the increase in computational time incurred, the previous filter dimensions were retained for the next set of experiments. For longitudinal images, corrected paired T-tests failed to show statistically significant di↵erence ($p < 0.05$). However for transverse images, using the larger two sets of filter dimensions showed statistically significant di↵erences ($p < 0.05$) in comparison to the smallest two sets of filter dimensions.

The study by Srivastava *et al.* in [247] explores the question of whether very deep or very complex network configurations would benefit a particular application beyond a certain level of complexity. Indeed they acknowledge that deep networks can better represent certain function classes more e ciently than shallow ones, but that at the same time, training becomes more di cult as depth and complexity increases. Additionally, they show with the MNIST dataset example that the early layers are important for learning and reducing error, but that additional layers beyond a certain threshold seem to have close to no e↵ect on the final performance.

The observations one derives from these experiments are in keeping with the observations of Srivastava *et al.*, whereby it is shown that typically, the initial increments in network layer stacks, output maps or filter dimensions result in

significant gains in performance. However, these performance gains start to level out beyond certain network complexities, and indeed in some cases also start to degrade. Increasing the depth and complexity of deep networks is thought to allow the network to better learn non-linearities in the data and to better model the functions being learned. In the case of encoder-decoder structures which must ultimately produce a labelled output of equal dimensions to the input, one must however also consider the detrimental e↵ect of too many sub-sampling layers at the encoder side. Each subsequent pooling layer reduces and compromises the resolution in the matrices at the inner layers. Thus, it is also quite possible that if the central stacks reach very small dimensions (say 4x4 for instance), this would be too small for the decoder to recover meaningful structure during the decoding step.

The charts in Figure 4.7 display the DICE coe cients obtained with the optimal network structure derived from the previous experiments, while on this occasion testing it on di↵erent sources of input data. Results are shown for using amplitude data, phase congruency maps, and finally a fusion of amplitudephase congruency data as a two channel input. One may note that while amplitude data on its own provides better results than phase information, the latter yields reasonable DICE coe cients which are in excess of 0.95. Phase information has the advantage of being amplitude invariant, and thus theoretically provides consistent results regardless of user selected amplitude settings. The fusion of amplitude and phase information in turn yields the best results across both transverse and longitudinal images, with DICE coe cients reaching 0.981±0.006 and 0.989±0.005 respectively. A Holm-Bonferroni corrected paired T-test showed statistically significant di↵erences ($p < 0.05$) between using fused inputs and either envelope or phase alone for

transverse images. The same was not the case for longitudinal images. A comparison was also made against the performance of other techniques from recent literature, described in [84; 72] and [73]. In these studies, only the performance on transverse im-

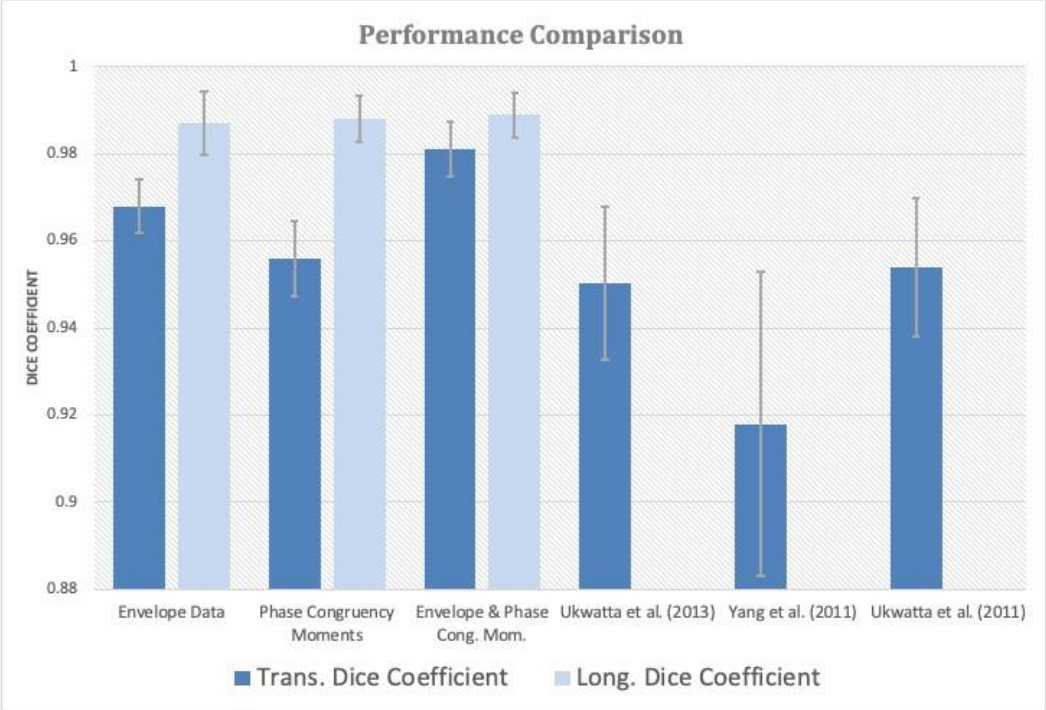

Figure 4.7: Performance of envelope, phase congruency, and fused data on dice coe cients.

ages was reported for the MAB segmentation, and therefore only this metric is included in the graph. One notes that deep FCNs applied to amplitude information alone yields comparable performance to that obtained by Ukwatta *et al.* in [84], but that its combination with phase information yields superior results. Figure 4.8 shows examples of the FCN generated contours (green) in

comparison to the average manually segmented contours from the expert (red).

In the final set of experiments, the performance of the FCN network is compared with the three di↵erent input types, amplitude, phase, and fused input, while varying the level of gain being applied on the original RF dataset, prior to computing the envelope and PCM data. This means that the RF data was

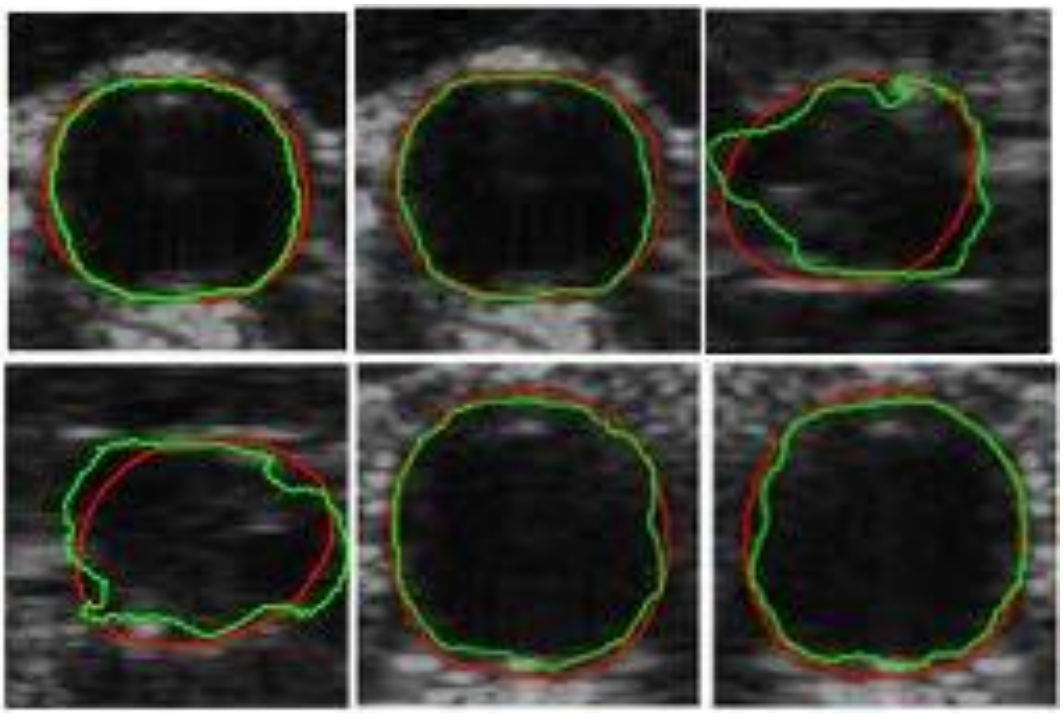

Figure 4.8: Examples of segmented Transverse images. Manual annotation is in red, while DCN segmentation is in green.

subject to clipping at the ceiling and floor level values of 65,535 in a similar fashion to that in section 3.4.3. The gain was also similarly varied from a value

of 1 (unity gain) to a value of 50. The results are displayed in Figure 4.9. Initially, one may observe that with gain values of 1 across all three instances, the performance of the FCN network is similar to that reported earlier in Figure 4.7. As the gain is increased progressively, the performance of the envelope-based FCN starts to drop, until it levels out for DICE values in the region of 0.86 ± 0.006. The reason why the performance levels out is thought to be due to the fact that the data is completely saturated at this point, as shown in Figure 4.10. The network has no useful input data to process, other than a fully saturated matrix with values of 255. The network thus produces a response which is the result of convolving its various filters with a uniformly equal input matrix. This is in keeping with the fact that the last three gain levels produce nearly identical segmentation performances. The final performance of the FCN network on envelope data might lead one to also think that networks are *inherently* more resilient to changes in gain, in comparison to the k-means and Active Contours methods reported in Chapter 3. The latter suffered a DICE degradation of over 20% at a gain of 20, whereas the FCN dropped down by only 10%. However, it is also true that the network's output at gain levels of 20 and upwards start becoming nonsensical, since it is the network's 'blind' response to a uniform input. The relatively high DICE performance of 88% at that point is merely a coincidental byproduct of the network's response having some degree of overlap with the ground truth. The performances at these gain levels must be treated with caution.

The result for the envelope + PCM fused input data also decreased with increased gain. A possible reason for this is that the saturated envelope data would still have some detrimental effect on the network's ability to correctly interpret the input information, as the latter is trained to consider both data

sources and not to ignore the envelope in favour of the phase data. It is however noted that this network displays an improved performance of approximately 2% in DICE over the envelope-only FCN – an improvement which is attributed to the presence of phase information. Lastly, the performance of the PCM-only FCN retained an approximately level performance throughout. Its initial performance was inferior to both that of the envelope and fused data, in keeping with previous results. However, changes in gain did not significantly degrade the PCM image, such that the FCN was able to retain a significantly improved performance over the other input types towards the end. Additionally, Holm-Bonferroni corrected T-tests showed statistically significant di↵erences between corresponding performances across the three methods ($p < 0.05$),

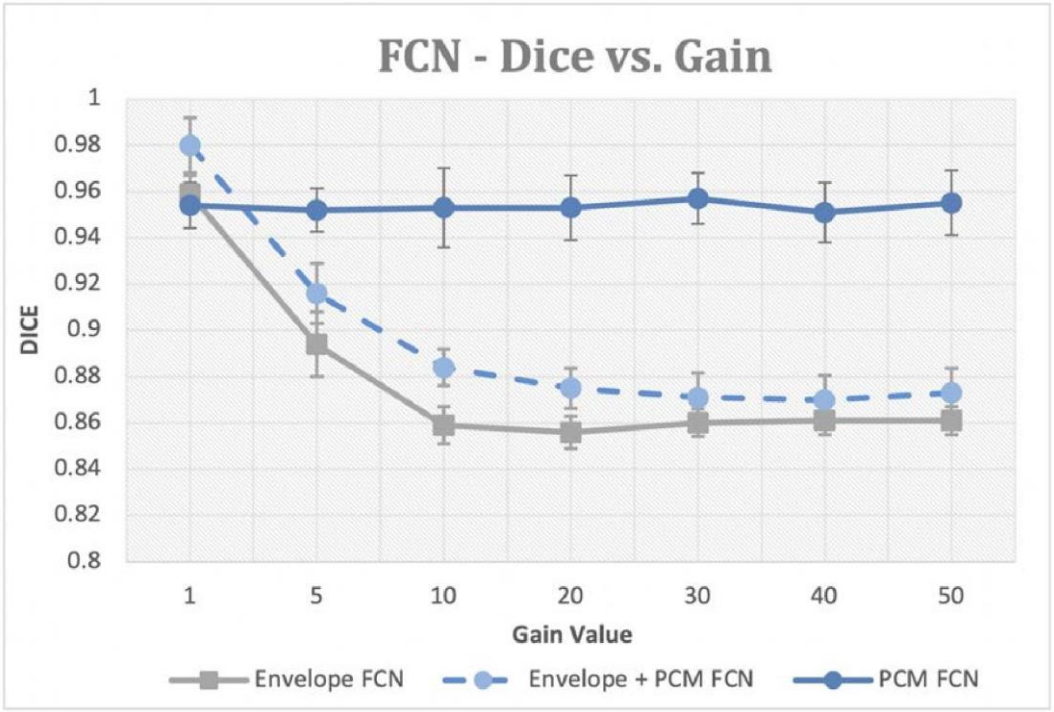

Figure 4.9: Performance of envelope, phase congruency, and fused data when gain of input data is varied from 1 to 50.

Figure 4.10: [Left] B-mode image with gain of 5; [Middle] Gain of 20; [Right] Gain of 50.

except for individual envelope only and phase only inputs at a gain of 1.

4.4 Conclusion

This Chapter has proposed the novel use of a Deep FCN, configured in an encoder-decoder structure, as a tool for semantic segmentation of the MediaAdventitia Boundary in transverse and longitudinal carotid images. It has explored the e↵ects of variation of parameters governing the size and structure of the network on segmentation performance. It has also demonstrated that the FCN achieves performance which meets the (then) state-of-the-art in MAB delineation when using amplitude data, and also that it exceeds the state-of-the-art when a novel input-level fusion of amplitude and phase information is used. Finally, it has demonstrated that phase information provides an amplitude invariant source of data which allows the FCN to behave consistently, regardless of variations in user gain settings. A recognised limitation of the study was the size of the dataset available to train the

network. The next chapter will study the e↵ects of using an improved network architecture based on U-NETS, together with the proposed use of a novel, modified objective cost function which applies geometric constraints to the problem. Additionally the semantic segmentation is carried out using amplitude and phase information to segment directly both the Media-Adventitia boundary and the Lumen-Intima boundary at one go, thus achieving a direct, fully automated segmentation method for the vessel wall.

5

Geometrically Constrained Deep Convolutional Networks

5.1 Introduction

In Chapter 4, we have proposed the novel application of a Deep FCN with an encoder-decoder (U-shape) structure to carotid ultrasound images, and explored the various sizes and configurations which yield the best performance when segmenting the Media-Adventitia Boundary. We have also proposed a novel, input-level fusion strategy of a combination of amplitude and phase information, and showed that this yields improved performance over that of using amplitude or phase alone.

Now, in order for complete measurement of vessel geometry to take place be it for later three dimensional reconstruction of the artery or for further assessment of plaque burden using metrics such as Total Plaque Volume (TPV) or Vessel Wall Volume (VWV) – *both* wall interfaces need to be identified: the Media-Adventitia boundary *and* the Lumen-Intima boundary [84]. Both these

interfaces need to be delineated in a robust and reproducible manner, and manual methods have been shown to be tedious, labour intensive [84], and prone to variability [139].

5.1.1 Aims

The aims of this chapter are to propose a novel, deep convolutional U-NET, including skip connections, as a fully automated segmentation tool, this time applying it to segment both the Media-Adventitia boundary and the LumenIntima boundary. The novelty lies in the use of a newly proposed, geometrically constrained objective function, which integrates a priori information as part of the network's Stochastic Gradient Descent optimisation. This allows the optimisation to be tuned on the basis of the anatomical structure of the carotid.The segmentation performance of the new network on transverse sections of the carotid is examined, but tests are now carried out on a larger base dataset, augmented to 97,200 images. The results are compared to other studies by reporting the DICE coe cient of similarity, modified Hausdor↵ Distance, sensitivity and specificity.

Novelty Statement: At the time of publication (2019), the novelty of this chapter's work was the creation of a geometrically constrained objective function for a deep U-Net used to segment carotid arteries.

5.2 Methodology

5.2.1 U-NET Networks

U-NETs have been proposed by Ronneberger *et al.* in [205]. They derive their name from the 'U-shaped' structure that they implement, which is very similar

to that of encoder-decoder based, deep FCNs used by Wang *et al.* in [204] at around the same period. In similar fashion, the main idea behind U-NETs is to implement a contracting network, which captures the semantic or contextual information from the image, followed by a symmetric expanding network to recover spatial information and increase the resolution of the output back to full scale.

However, in order to improve the localisation performance during classification, high resolution features from the contracting path are combined with the upsampling path during expansion. These are referred to as *'skip connections'*, and their role ensures that features that are learnt during the downsampling process are also reused during the upsampling process, thus combining precise localisation information with contextual information from the contracting path. A successive convolution layer is applied right after the concatenation from the skip connection, in order to assemble a combined output from both inputs. This structure is illustrated in Figure 5.1, whereby Ronneberger *et al.* [205] proposed a 4-stack contracting path followed by a 4-stack expanding path, with a further 2 layers of convolutions or upwards convolutions at each stack. On the contracting side, the convolutions used 3x3 filter kernels with RELU activation, followed by 2x2 max pooling layers. On the expanding side, the upwards convolutions used 2x2 filter kernels, whereas the convolutions used for concatenating the skip connections also used 3x3 filter kernels.

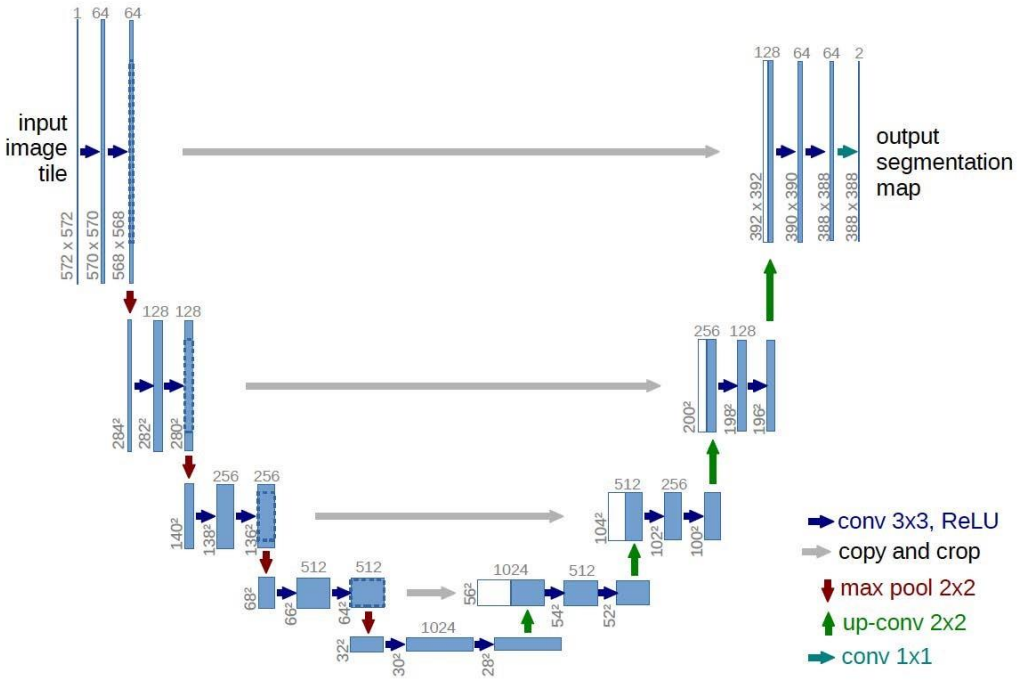

Figure 5.1: A U-NET architecture example, where blue boxes represent multichannel feature maps and white boxes represent copied feature maps. Reproduced from with permission from [205].

Dropout layers

Feedforward neural networks adapt the weights on the incoming connections of non-linear hidden units, between their inputs and outputs. By adapting these weights, the network is able to learn feature detectors that allow it to predict the correct output when a certain input vector is provided [248]. Now, if the relationship between input and output is complex, yet the network has enough hidden units to model this complexity correctly, it is likely that there would exist a di↵erent set of weights for the hidden units that the network can learn, which would produce exactly the same answer. This happens particularly if there is only a limited training datase t[248]. It is also certain

that each of these di↵erent possible weight vectors will produce completely di↵erent predictions on a hold-out test dataset, and will produce a worse performance on the test dataset when compared to the training dataset.

Hinton *et al.* proposed in their work in [248], that in order to reduce such overfitting, a 'dropout' scheme could be introduced in order to prevent complex co-adaptations of the hidden nodes on the training data. This is implemented in a manner whereby, during each training iteration, hidden units are randomly omitted from the network with a probability of 0.5, such that any particular hidden unit is forced not to rely on others being present. This process may be seen as an e cient version of 'averaging' neural network models, since the error on the test set may also be reduced by averaging the predictions produced by a large number of di↵erent networks which have been trained separately. The latter process is of course impractical however, and computationally expensive during both training and testing. Thus, random dropout would make it possible to train a large number of di↵erent networks in a reasonable amount of time. Hinton *et al.* show in their work that dropout achieves improved error rates on a number of benchmark test sets [248].

Application of U-NETs

The U-NET network was constructed using an in-built function available to MATLAB, which is called *createUnet*, allowing the user to define a deep network with U-NET Architecture. The U-NET structure constructed was smaller than that defined by Ronneberger *et al.* in [205]. It was kept largely similar in terms of structure to the network which was finally retained in Chapter 4 – since it was shown here that these structural parameters proved

optimal for the type of segmentation being carried out. Thus, as shown in Figure 5.2 and

5.3, a total of 6 stacks were used, with 3 on the encoder side and 3 on the

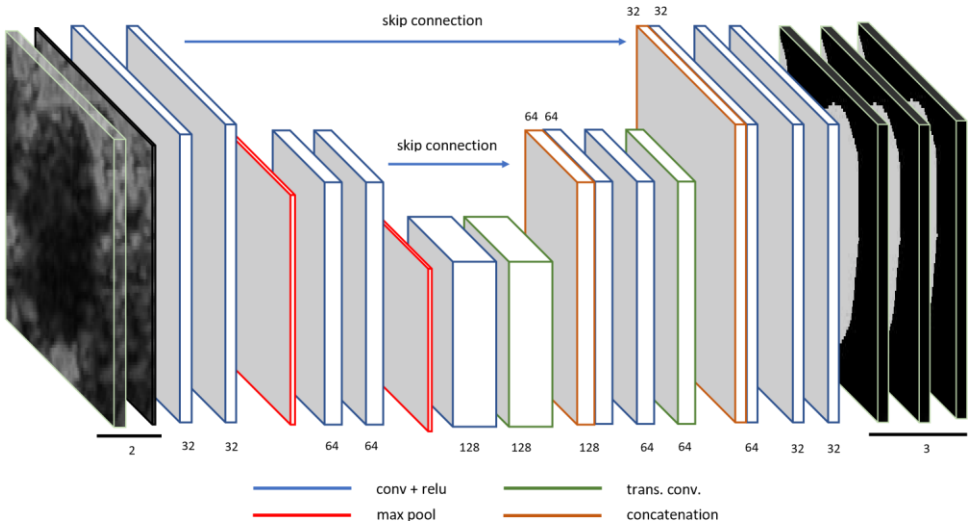

Figure 5.2: A graphical representation of the U-NET architecutre. The colour of the boxes represents the layer type. The number of channels at each layer is given below each box. Reproduced with permission from [92]. Copyright 2020, IEEE.

decoder side. The decoder was built using stacks of convolutional transpose layers with a stride of 2, and convolutional layers for concatenation.

The segmentation task attempted this time was a pixel-by-pixel classification problem to categorise each pixel into one of *three* classes, which are namely, 1) background pixels falling outside the MA contour; 2) pixels falling in between the MA and LI contours, and therefore falling within the vessel walls which may potentially harbour collection of plaque; and 3) pixels falling within the LI contour, and which therefore constitute pixels representing the lumen

of the artery and the blood flowing through. This is in contrast to the segmentation task described in Chapter 4, where the network was designed to segment pixels only in two classes: those outside the MA contour and those within the MA contour. Thus, at the end of this network, the number of output maps were reduced to three and not two, and fed into a softmax classifier, which provided logistic regression for a three-class problem. Additionally, two skip connections were included, joining the final convolutional blocks of the encoder's first and second layer, to concatenation layers at the respective end in the decoder, in order to leverage both high- and low-level features. Drop out layers were included after the second convolutional block of each encoder layer.

5.2.2 Geometrically Constrained Cost Functions

Theory

This section focuses on the objective cost functions being used on the output side of the U-NET, in order to calculate the loss for stochastic gradient descent. All previous experiments carried out in Chapter 4 used the standard, built-in, multi-class logistic regression function to calculate the loss on each iteration. In this section, a novel modification is proposed, which includes a priori information on the carotid anatomy, in the form of geometric constraints, in order to tune the cost function to the type of data that is being segmented.

If we assume that we have a training set of m training examples, defined as $\{(x^{(1)},y^{(1)}),\ldots,(x^{(m)},y^{(m)})\}$, then in a simplistic network carrying out a linear regression, the objective function would be the sum of differences between the

predicted output $h_{\checkmark,b}(x^k)$ and the ground truth labels y^k, over all di↵erent training examples k = 1,2,3...m. This would be expressed as:

$$E_O = \sum_{k=1}^{m} (h_{\theta,b}(x^{(k)}) \quad y^{(k)})^2 \tag{5.1}$$

where the output $h(x)$ is the result of a feedforward operation carried out

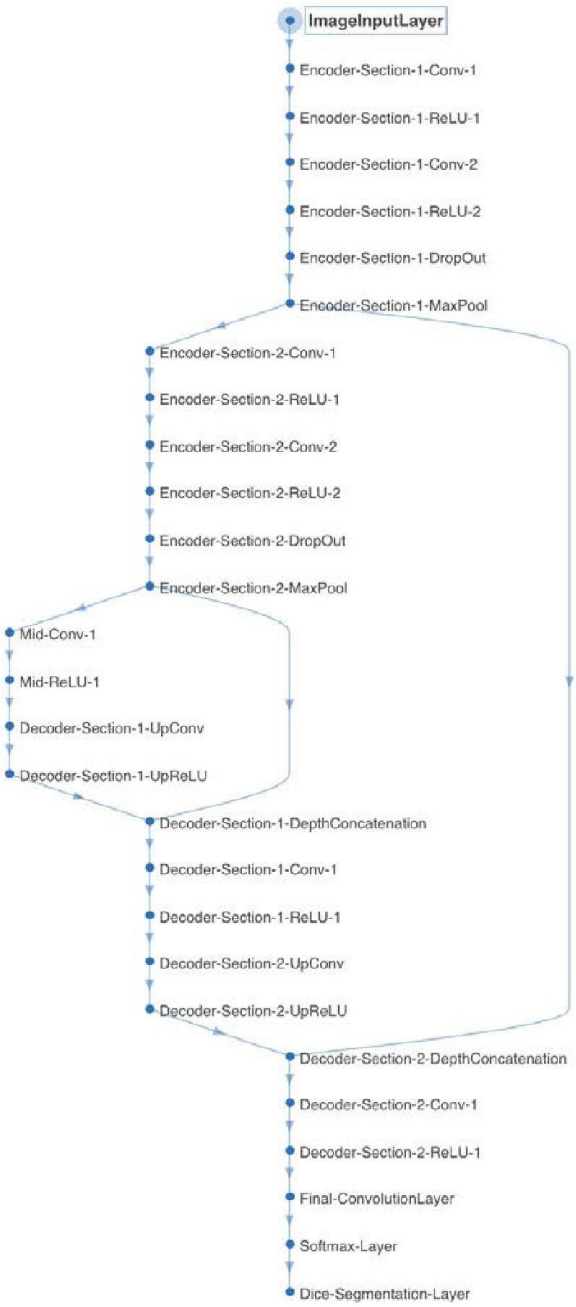

Figure 5.3: A MATLAB reproduction of the U-NET structure using the 'analyzenetwork' function.

through the network with training parameters ✓ and *b*. In a pixel-wise classification problem on the other hand, we perform a multi-class logistic regression to classify each pixel to a category. Thus, at the output of the decoder, the final convolution layer is connected to a softmax layer, which classifies each pixel into one of three categories: the background pixels outside the MAB, pixels between the MAB and LIB and pixels within the LIB. If we assume that $z_{i,j}^c$ is an output neuron's activity computed for channel *c* at position (*i,j*), then the output of the softmax layer is given by:

$$P^c(i,j) = \frac{\exp(z_{i,j}^c)}{\sum_{l=1}^{N} \exp(z_{i,j}^n)} \tag{5.2}$$

where N is the number of classes. To compute the loss, we considered and compared two basic loss functions: the class-weighted cross entropy loss (CE) and the class-weighted generalised DICE loss (GDL). Another recent loss function applied to medical imaging is that of using Boundary loss [249]. However, this is intended for problems with highly unbalanced classes, where the foreground class is of several orders of magnitude smaller than the background classes. In this case, the classes were not unbalanced to this extent, and therefore the author opted to test with the standard loss functions described. The prediction scores *P* and training targets *T* are organised as 3D matrices, with the first 2 dimensions *x* and *y* representing spatial location in the image, and the third dimension *n* representing the class. The weighted cross entropy function for a multi-class problem is then defined as:

$$L_{CE} = \frac{1}{N} \sum_{k=1}^{m} \sum_{n=1}^{N} w_n T_{nk} log(P_{nk}) \tag{5.3}$$

where *m* is the number of training examples, *N* is the number of classes and *w* is a vector of weights for each class. The weighted generalised DICE loss function for a multi-class problem is defined as:

$$L_{DC} = 1 \quad \frac{2\sum_{n=1}^{N} w_n \sum_{f}^{F} P_{nf} T_{nf}}{\sum_{n=1}^{N} w_n \sum_{f}^{F} P_{nf}^2 + T_{nf}^2} \tag{5.4}$$

where F is the number of pixel elements along the first two dimensions (*x* and *y*) in the matrices *P* and *T*, *N* is the number of classes along the third dimension, and w_n is the class specific weighting factor that controls the contribution that each class will make to the loss. Specifically, it is typically the inverse area of the expected region, defined as:

$$w_n = \frac{1}{\left(\sum_{f=1}^{F} T_{nf}\right)^2} \tag{5.5}$$

In order to tune these objective functions and make them more sensitive to the anatomy of the structures that are being segmented, a modification is proposed to the objective functions being tested, by adding three additional cost terms, which are defined as follows:

Curvature.

The radius of curvature at any point on a curve is equal to the radius of the circular arc which best approximates the curve at the said point, as shown in Figure 5.4. It is also defined as the inverse of curvature, and is expressed mathematically as follows:

$$R = \frac{1}{\kappa} \tag{5.6}$$

where is the curvature at a point. If we let the contour of the artery, from which these penalty terms are going to be derived, be expressed in the form of a curve in two dimensions as $u = f(v)$, then the curvature term itself may also be expressed as [250]:

$$\kappa = \frac{\left|\frac{d^2u}{dv^2}\right|}{\left[1 + \left(\frac{du}{dv}\right)^2\right]^{3/2}} \tag{5.7}$$

The penalty term proposed from the above measure of curvature is defined as follows:

$$C(u) = \max\{|\overline{\kappa}|, 0\} \tag{5.8}$$

The absolute value of is used because one may ignore the sign of the value, which is indicative of the direction of curvature, and which is not important for computation of the penalty term. The mean is taken across all the values of worked out on the whole contour. If the contour is largely smooth and without notches, this would generally yield similarly small values of all throughout the curve. Taking the examples shown in Figure 5.4 for instance, in the left panel, is a perfect circle of radius 50 pixels. The osculating circle, which would best approximates the curve at any point in this case, is the same circle of radius 50 pixels. Thus the curvature at all points is 0.02. In the middle panel of Figure 5.4 is a more realistic example of how an artery would be manually delineated by the expert, and where the contour is largely circular and smooth, but has small 'imperfections' resulting from the imperfectly traced contour. The mean value of || over all points in this case is larger than in the first case, but would still be below a certain threshold.

A threshold value is therefore subtracted, because one may ignore curvature values below this value as simple 'noise' created by the imperfect manual delineation process. The right panel of Figure 5.4, shows pronounced notches

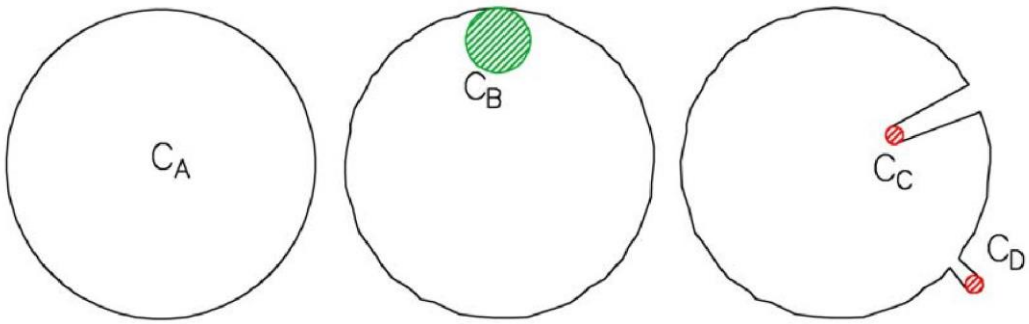

Figure 5.4: [Left] A graphical representation of Curvature C_A in a perfect circle where is equal to the inverse of the radius of circle C_A. [Middle] The penalty at $C_B > C_A$ because the radius of the osculating circle C_B at various points on the imperfect contour is smaller than that of C_A. [Right] Curvature C_C is of opposite sign to that of C_D due to opposite direction of deformation.

marked by C_C and C_D. These would give significantly larger values of of opposing polarity, which would be picked up and penalised by equation 5.8. A zoomed in and more realistic example is shown in Figure 5.5. The left panel shows a section of contour produced by the network, whereby jagged edges may be observed within the blue circle which however follow a regular path. These

produce low values of curvature $|\overline{\ }|$. Following subtraction by , which yields a negative value, the max function of equation 5.8 outputs a 0 value to ensure

no penalty in this case. The right panel in Figure 5.5 shows an irregular notch demarcated by the green circle, which would be anatomically inconsistent with the structure of the Media-Adventitia contour. Such an irregular notch would

elevate the value of |□| which exceed □ and thus count towards the penalty.

Solidity

If we consider the example shown in the left panel of Figure 5.6, one may observe an instance where, although the contour presents an irregularity which

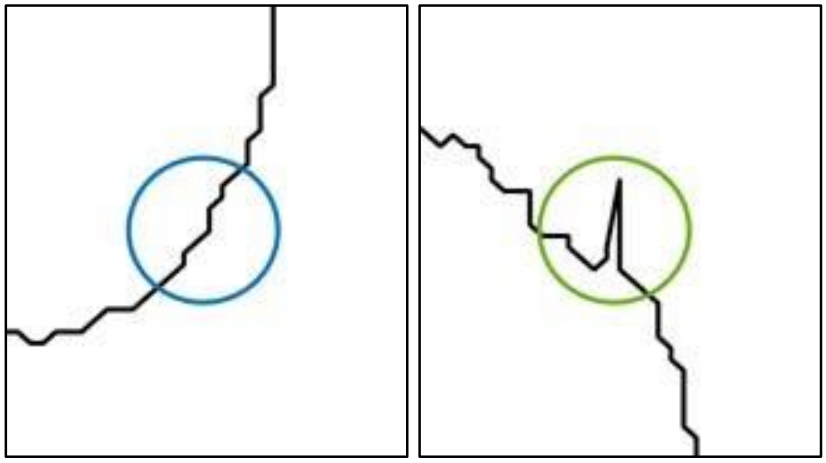

Figure 5.5: [Left] A section of contour produced by the DNN on the MAB outer border, showing jagged edges (blue circle). [Right] A section of contour produced by the DNN with an incorrect, irregular notch (green circle).

would anatomically not be possible, the curvature penalty alone would not impart any significant e↵ect on the overall cost function. This is due to the fact that the irregularity fits osculating circles with a fairly large radius, and which

either fall below the threshold, ✎, or else would still not result in a large penalty. Thus a second penalty is proposed, which measures the Solidity of a contour. Solidity is a measure of morphological roughness and is sensitive to concavities in a shape or structure. It compares the pixel area of the object (C_A in Figure 5.6, right side) to the area of a bounding reference shape (C_B in Figure 5.6, right side), which in this case would be the convex hull. Mathematically therefore, it may be expressed as:

$$SLD = \frac{A(u)}{A_c(u)} \tag{5.9}$$

where $A(u)$ is the area of the object and $A_c(u)$ is the convex area of the shape in question. A solidity of 1 would indicate a perfectly solid shape with an area which is equal to its convex area. An irregular shape with concavities

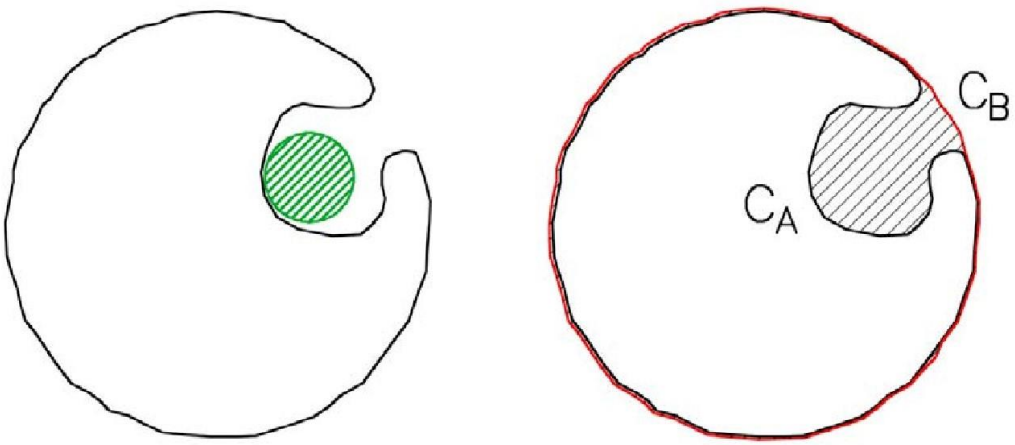

Figure 5.6: [Left] A graphical representation of a contour irregularity with an example osculating circle. [Right] A graphical representation of the Convex hull C_B, shown in red, around the irregular contour C_A. The di↵erence in area

shown by the shaded section, between C_B and C_A is used to calculate the measure of Solidity.

would present with a solidity which is < 1 due to the area being smaller than the convex area. With regards to defining a penalty term, the MediaAdventitia boundary of the carotid is expected to be smoothly circular, without any irregular concavities within the perimeter. Thus one would seek to penalise the objective function in instances where the solidity of the segmented mask would again deviate significantly from 1, as would be the case in Figure 5.6. The cost term proposed is therefore:

$$S(u) = \frac{A_c(u)}{A(u)}\,1 \tag{5.10}$$

where the inverse of SLD is used to have an increasing term in proportion to increased concavities in the shape.

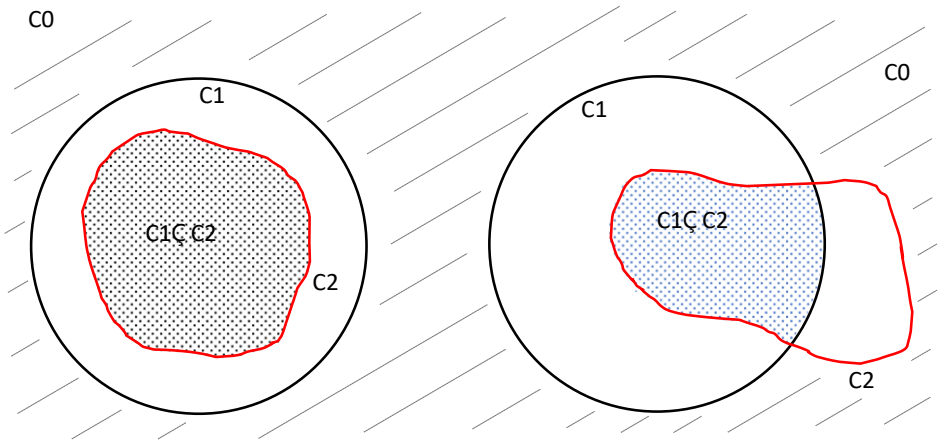

Figure 5.7: A graphical representation of incorrect intersection between the MA outer boundary and LI inner boundary.

Intersection

The third penalty term introduced to the objective cost function is a term which penalises the instances whereby the Lumen-Intima boundary would not be contained within the Media-Adventitia boundary, as is shown in Figure 5.7. This is again a circumstance which may not anatomically occur, and therefore its occurrence is penalised heavily to reduce the possibility of the neural network from converging to such a result. Referring to Figure 5.7, if we consider the MAB to be defined as contour C_1, the LIB to be defined as contour C_2, and their joint intersection as contour $C_1 \setminus C_2$, it easily follows that any occurrence whereby $Area_{C_2} = Area_{C_1 \setminus C_2}$ should be penalised. The proposed penalty term is therefore defined as:

$$I(u) = \frac{Area_{C_2} \quad Area_{C_1 \cap C_2}}{Area_{C_0}} \tag{5.11}$$

where the denominator is used to normalise summation of pixels outside C_1. The overall new modified objective cost function, E_M, may therefore now be represented in the following manner:

$$E_M = L(x^k) + \square C(v) + \qquad S(v) + I(v) \tag{5.12}$$

where the coe cients □, and are scaling coe cients multiplied by the penalty terms $C(v)$, $S(v)$ and $I(v)$ respectively, in order to scale them to a suitable magnitude.

Application of Geometric Penalty Terms to Objective Function

The geometric penalty terms were implemented in MATLAB using functions which were incorporated into the final layer of the U-NET. The block diagram

shown in Figure 5.8 shows the change in training from the traditional networks shown in Chapter 4. The curvature penalty term was computed using the MATLAB function *LineCurvature2D*, written by Kroon *et. al* [238], which returns the curvature value at all points on a contour defined by a 2D vector of points. The value of was set by first computing the average curvature across all points of the manually traced MA contours, and subsequently an average was obtained over all such manually traced contours available in the labelled dataset. This yielded an average = 0.34, and thus was set to 0.5.

For the Solidity penalty term, this was obtained using the built in MATLAB function called *regionprops*, and computed by taking the inverse and subtracting 1. For the Intersection penalty term, the area of the intersection $Area_{C1 \setminus C2}$ was found by first obtaining the contour of said intersection using the built-in MATLAB function *polybool*, and then calculating the area within using the built-in function *polyarea*. The same function was also used to calculate the area of the LI contour. With these areas available, the intersection penalty was calculated as per equation 5.11

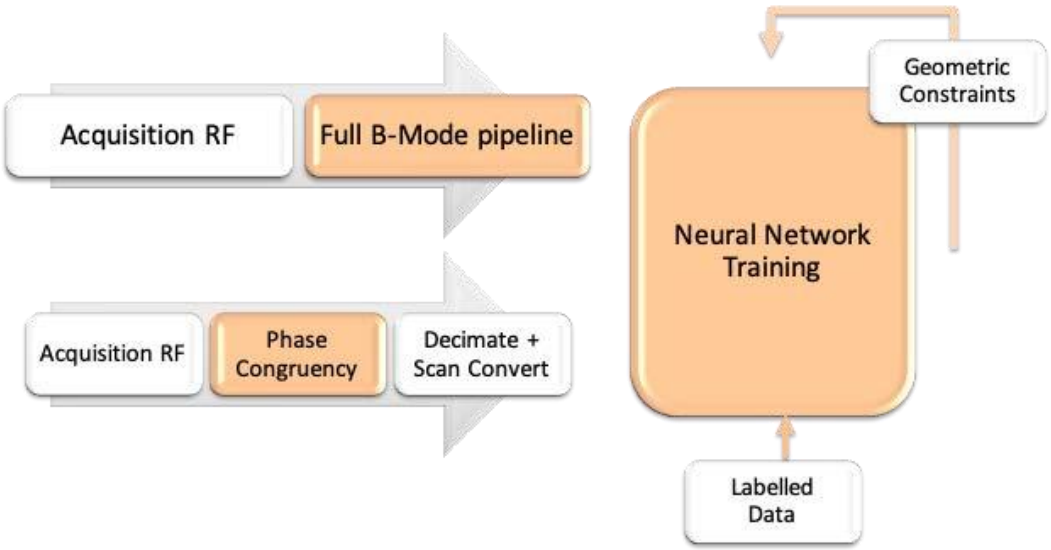

Figure 5.8: A block diagram showing the pipeline of training using Deep Neural Networks and Geometric Constraints.

The final intent of a modified objective cost function as described in equation 5.12, is to firstly require the network to train further if the cost function results in greater, penalised values. Additionally, it must also feed back through the backpropagation algorithm, a set node errors observed at the output, such that the network may direct correction e↵orts towards the right filter map nodes further in. An approach to feeding back the node errors at the output was implemented, whereby the penalty coe cients may be applied locally, and thus spatially, to nodes corresponding to irregular pixel positions.

The U-NET trains on data in sub-batches of frames, given that optimisation is based on batched stochastic gradient descent. The local application of penalty coe cients is however applied separately on a frame by frame basis. For a particular frame therefore, the penalties based on Curvature, Solidity and

Intersection were first computed as described previously. After this, the 2D vector of points representing the convex hull in place around the segmented

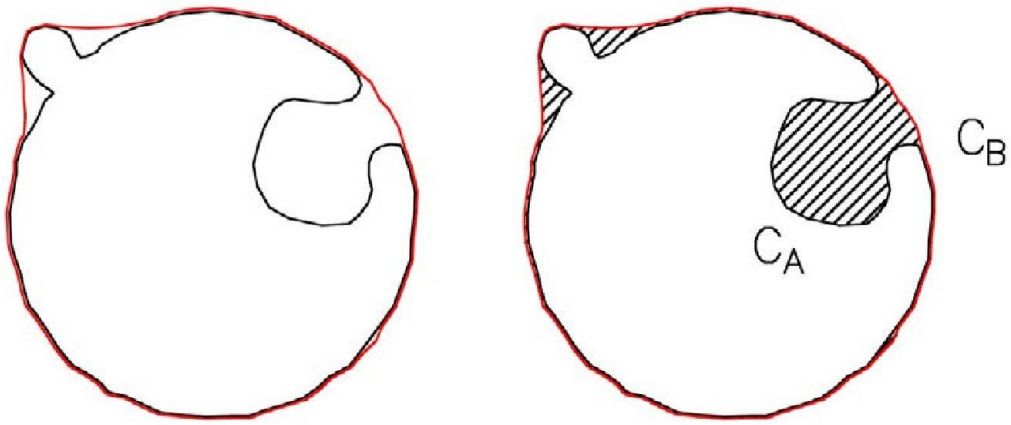

Figure 5.9: [Left] A graphical representation of an irregular contour on the MA boundary, and of the convex hull around this contour shown in red. [RIGHT] The subtraction of the irregular contour from the convex hull leaves the area shaded in black stripes, which is used to locally apply the penalties around these regions.

contour was taken. During training, the network produces at each iteration its best attempt at segmenting the MA and LI contours, and the convex hulls around these contours were obtained using the *regionprops* MATLAB function described earlier. An example with both positive (outward) and negative (inward) contour irregularities is shown in the left panel of Figure 5.9. The convex hull is the bounding contour shown in red, which contains the object contour, inclusive of any irregularities that may be present.

The area within the convex hull was then filled with ’1’s using the MATLAB function *imfill*, and from this the original object contour was subtracted, which was also filled with ’1’s within the contour. What was left was thus a mask of the di↵erence between these two subtractions, as shown in right panel of Figure

5.9 through the shaded section. This process was repeated for both the MA contour and the LI contour, resulting in a 120 x 120 penalty matrix, containing ’1’s at the shaded region and ’0’s elsewhere. Then the three penalties, each multiplied by their own coe cient described in equation 5.12, were summed together, and the resulting value was multiplied against the penalty matrix. In case of the LI contour, the penalty based on solidity was relaxed, since the LI contour should be allowed to have concavities in it, as these may signify the presence of plaque which should not be corrected. The sharp protrusions best picked up by the curvature penalty, or intersection with the MA contour, are however not anatomically possible even in the LI contour, and thus these were included for penalisation at the same coe cient values.

The penalty matrix was then applied to the network as follows. During the feedforward step, the network computes the final Loss value based on the standard objective cost functions used. When testing the standard, unmodified loss functions, both Weighted Cross Entropy Loss and Generalised Dice Loss functions were used and tested as cost functions. Their performance was evaluated in the results and discussion. However for the inclusion of the geometric constraints, all the non-zero pixel values in the penalty matrix were summed, and added to the loss value produced by the standard loss function. Conversely, during the backpropagation step, the penalty matrix was added, on a pixel-by-pixel basis, to the matrix of delta node errors produced by the

network. The matrix of delta node errors, is a 120 x 120 matrix holding an error value, , for each node, which represents how mistaken each node was in producing the correct result. The addition of the penalty matrix thus 'amplified' the error deltas for nodes which were spatially located around or inside the irregularities of the contour. This in turn steepens the gradient at which the coe cients of these nodes are updated into more suitable values.

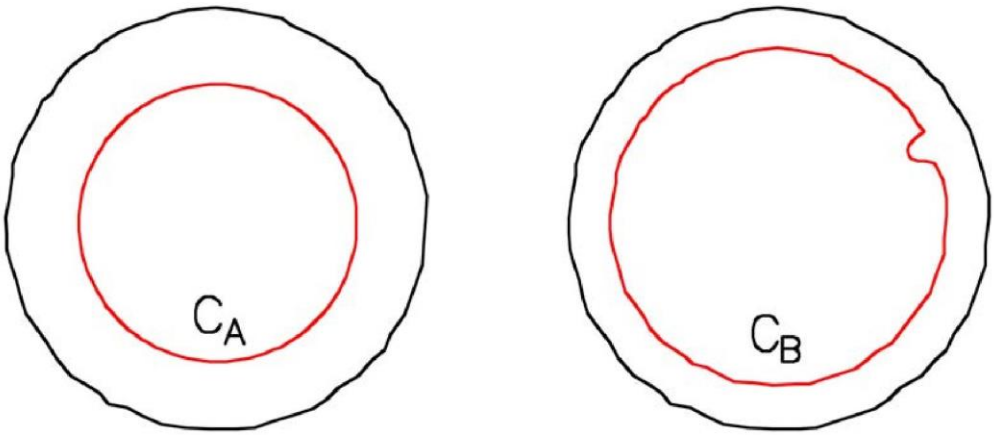

Figure 5.10: [Left] A perfectly circular contour (red), with a smaller degree of overlap with ground truth contour (black). [Right] An contour with an anatomically impossible notch (red) with a higher degree of overlap with ground truth

(black) will ultimately score better using DICE coe cients

The choice of coe cient values, , and multiplied agains the penalty terms in equation 5.12 is discussed in this section rather than through a detailed experimental analysis in the results and discussion. The values for these coefficients were chosen heuristically, since such hyperparameters are very di cult to tune to perfectly optimal values, and also have e↵ects which may be di

cult to qualitatively, or indeed quantitatively, observe. There could be instances where a contour, which visually appears notch-free and reasonably smooth, would have an inferior DICE coe cient to a contour which may be slightly larger, has a notch defect, but ultimately overlaps on more pixels than the previous contour would. An example is shown in Figure 5.10. Such situations stem from the limitations of performance metrics like DICE coe cient, which although give a good indication of performance, are not perfect.

From an experimental point of view, a detailed grid search with the full dataset to determine the optimal scaling coe cients was not considered feasible, since even testing just 4 di↵erent values for each coe cient would necessitate 4 x 4 x 4 x 18 cross fold iterations, resulting in 1,152 training runs. With approximately 97,000 images, and each epoch taking just under 7 hours each, a single training run of around 20 epochs would require an average of 6 days to finally arrive at a result. Thus, a heuristic approach was deemed as a necessary compromise in this case, and the authors acknowledge that the absolutely *optimal* values may yet lie undiscovered. By using MATLAB's debugging function, the author manually observed what typical loss values and corresponding penalties are produced by the network through the initial and later stages of training. The values , and were kept equal, and set such that the amount of penalty added during an iteration would approximate 30% of the full loss value when the network is still considerably o↵ from producing a realistic contour. The compromise is to set the penalties at values not too high that they cause the network not to converge at all, but not too low that they leave no e↵ect on the result. After testing a set of coe cient values across smaller sample datasets, randomly chosen from the available full datasets, we

chose a value of $5e^{7}$ as a coe cient value which produced a large number of positively improved results.

5.2.3 Experiment Construction

Given the size of the dataset available, the data was first augmented using the following strategy. The transverse image sets, as well as the corresponding labelled datasets, were first scaled twice by a factor of 1.2 and 1.5. The scaled images were then cropped back to their original dimension. Each image, having dimensions of 256 x 256 pixels was then patch-wise sampled 9 times along an equally spread 3x3 grid centred about the middle, using a 120 x 120 pixel window with a constant overlap between each region. Finally, the images were rotated through 90, 180 and 270 degrees. All the additionally generated image sets and corresponding labels were concatenated into separate augmented dataset-pairs. From a dataset of 18 di↵erent patients, originally having a total of 50 images each, the dataset was augmented to a total of 97,200 images

A number of schemes were then employed in order to prevent over-fitting and in order to ensure validity of the results. Firstly, during the training of the neural network, a basic early-stopping technique described by Prechelt in [251] was implemented in order to account for over-fitting. In accordance with the technique described by Prechelt, the dataset was split into two sub-datasets: a training dataset and a validation dataset. This split was implemented in a ratio of 66% training data and 33% validation data. Training took place only on the training set, and the error was evaluated on the validation set after each epoch. Once the training is stopped, the weights that the network had in the previous training run are used. In these experiments,

training was carried out using stochastic gradient descent, with batches of 8 ultrasound images, a learning rate of 1→10 2 which decreased by a factor of 10 every 5 epochs, a momentum of 0.90 and a weight decay of 1→10 4. The weights and bias terms were randomly initialised. The early stopping was applied after 20 epochs.

An 18-fold leave-one-out cross validation scheme was also utilised to have some assurances on the validity of the results. The data available from the acquisitions was split at patient-scan level during each iteration into two categories: training + validation datasets (described earlier for the early stopping technique) and testing datasets. The training + validation datasets were created by concatenating the various ultrasound images obtained from di↵erent patient scans, and then randomising their sequence. The testing dataset in each iteration was then always made up of a hold-out set of ultrasound images from a particular patient scan, which were not present in the training set.

Two experiments were carried out to assess the performance of the network. Firstly, an ablation study was run across the di↵erent sources of input available, namely: amplitude data, phase congruency maps, and a fusion at input-level of both. This serves to evaluate the di↵erence in performance when using the fused amplitude and phase data while the network is segmenting both the MAB and LIB contours. The previous work in Chapter 4 carried out a similar ablation study, but on MAB contours only, with a much smaller dataset, and with a regular end-to-end CNN network. A second ablation study was carried out to test the performance with and without the additional geometric constraints applied to both the basic Cross Entropy and Generalised Dice Loss

objective functions. 18-Cross fold validation exercises on a hold out test set were run in both experiments. Furthermore, a third experiment was carried out to compare performance across the techniques developed in Chapters 3, 4 and this chapter. Since the PCM based segmentation with active contours developed in Chapter 3 was developed to segment only the MAB, the comparison was limited to testing this contour. Thus, the latter technique was compared against the traditional CNN (without geometric constraints) of Chapter 4, and with the U-NET with constraints proposed in this chapter.

The U-NET was built using the MATLAB Deep Learning toolbox and trained on an Intel Core i7 with a Radeon Pro 560 GPU. At the time of writing of this dissertation, a previously available computer, with an NVIDIA CUDA enabled Geforce GPU was no longer available. Thus, training was carried out using multi-core CPU parallel processing, but without harnessing the facilities of the GPU. The segmentation results obtained from the U-NET were compared against a manually labelled ground truth dataset, which was manually and independently traced under the supervision of 2 radiographers, and used as labelled training data. Each radiographer manually traced samples from the dataset and supervised the manual delineation of the rest. The process was repeated twice, with a period of 2 weeks in between sessions and an average across both labelling sessions was retained as the final ground truth.

5.2.4 Performance Metrics

A number of evaluation metrics were used in order to quantify the performance of the U-NET, and in order to allow comparison against other methods in the literature. The performance metrics used are the same as the

ones used in Chapters 3 and 4, but for convenience's sake only, these are listed down again here with a simple definition of each.

DICE Similarity. The similarity between the segmented result and the ground truth is computed using the Dice Coe cient of Similarity. This gauges the degree of overlap between two boundaries.

Modified Hausdor Distance. The Hausdor Distance is a measure of distance between two point sets. It provides the largest mismatched points between two boundaries. The Modified Hausdor distance (MHD) [246] on the other hand finds the mean distance between two boundaries.

Sensitivity. The Sensitivity term is defined as the ability of a method or test to correctly classify a sample as being diseased [244], or in other words, the probability of a test being positive when the disease is indeed present.

Specificity. The Specificity term, is defined as the ability of a method or test to correctly classify a sample as being free from disease [244], or in other words, the probability of a test being negative when the disease is indeed absent.

5.3 Results and Discussion

As a novel application over the previous work in Chapter 4, the network was trained to identify the contours of both the Media-Adventitia boundary as well as the Lumen-Intima boundary contemporarily. Thus, the segmentation performance was quantified individually for these two boundaries and reported in the results. The rationale behind this is because the problem of segmenting these two boundaries poses di erent levels of challenge for the network. The Media-Adventitia boundary normally presents itself as an

approximately circular structure with well-defined contours as evidenced by two examples shown in Figure 5.7 [LEFT]. The Lumen-Intima boundary on the other hand is prone to having an irregular shape, partly because of the plaque which may be sited between the walls of the Intima, and partly because the walls of the Intima might not have clearly defined contours and poor contrast.

Table 5.1: Averaged results with standard deviation for testing DICE coe cient, MHD, Sensitivity (Sens) and Specificity (Spec) on di↵erent inputs, using weighted cross entropy loss, on [18]-fold Leave-one-out cross validation on the hold-out test datasets

Metric	**MAB DICE**	**MAB MHD** in mm	**LIB DICE**	**LIB MHD** in mm	**MAB Sens**	**MAB Spec**	**LIB Sens**	**LIB Spec**
Method								
UNET+A+CE	.880 ± .031	5.020 ± .263	.869 ± .085	5.778 ± .245	.863 ± .051	.894 ± .016	.889 ± .079	.907 ± .044
UNET+P+CE	.832 ± .069	6.038 ± .821	.794 ± .083	7.120 ± .150	.855 ± .066	.859 ± .073	.866 ± .035	.876 ± .084
UNET+A/P+CE	.936 .033 ±	.215 .090 ±	.901 .045 ±	.349 .157 ±	.939 .037 ±	.953 .036 ±	.936 .046 ±	.960 .047 ±

U-NET is used with the following: (A): Amplitude; (P): Phase Congruency; (A/P): Amplitude and Phase Congruency; (CE): Weighted cross entropy loss.

Table 5.1 presents the averaged performance metrics for the unmodified (no geometric constraints) network when trained with amplitude data alone, phase congruency maps alone, and lastly, with the fusion of amplitude and phase congruency maps supplied as a 2-channel input. The basis loss function used

Table 5.2: Averaged results with standard deviation for testing DICE coe cient, MHD, Sensitivity (Sens) and Specificity (Spec) with di↵erent loss functions, across [18]-fold Leaveone-out cross validation on the hold-out test datasets

Metric	MAB DICE	MAB MHD in mm	LIB DICE	LIB MHD in mm	MAB Sens	MAB Spec	LIB Sens	LIB Spec
Method								
UNET+CE	.936 ± .033	.215 ± .090	.901 ± .045	.349 ± .157	.939 ± .037	.953 ± .036	.936 ± .046	.960 ± .047
UNET+GDL	.941 ± .037	.210 ± .071	.902 ± .060	.341 ± .152	.933 ± .041	.952 ± .039	.932 ± .049	.958 ± .040
UNET+CE+GC	.955 ± .040	.208 ± .072	.917 ± .051	.284 ± .182	**.950 ± .050**	.968 ± .014	.934 ± .043	**.973 ± .052**
UNET+GDL+GC	**.959 ± .029**	**.200 ± .071**	**.920 ± .044**	**.270 ± .189**	.949 ± .036	**.973 ± .036**	**.938 ± .039**	.965 ± .040

U-NET is used with the following: (CE): Weighted Cross Entropy Loss; (GDL): Weighted Generalised Dice Loss; (GC): Geometric Constraints

in this experiment is the weighted cross-entropy loss function. The results obtained show that the inclusion of both amplitude and phase information improve the DICE performance by approximately 4% and 10% over amplitude alone and phase alone respectively for the MAB, and 3% and 10% respectively for the LIB. A Holm-Bonferroni corrected paired sample T-test showed that for both MAB and LIB DICE, the di↵erences between averages were statistically significant for all three methods ($p < 0.05$). Improvements are also noted in the MHD, sensitivity and specificity, when the fused data is used. The performance di↵erence of the MHD is however noted to be larger than one order of magnitude when comparing the fused input to the individual inputs. The reason for this is that at lower values of DICE coe cients, the segmented output of the network is often consistently irregular in shape. This results in markedly elevated values of MHD, despite it being designed to mitigate such e↵ects over the regular Hausdor↵ distance.

Table 5.2 and Figures 5.11 and 5.12, present the averaged performance metrics quantified for training the U-NET with di↵erent loss functions and with or without the geometric constraints. We observe initially that in agreement with the noted di culty of the segmentation task, the performance

of the network in delineating the MAB is consistently higher than the performance for delineating the LIB across all methods. The average DICE coe cient for the MAB is between 2.7% to 4.0% higher for the various training methods. Correspondingly, the average MHD is noted to be 0.07mm to 0.13mm lower in the MAB than the LIB, for the various methods.

Furthermore, when observing the performance of the network when using different loss functions, one notes that the weighted GDL slightly outperforms the weighted CE function across DICE and MHD metrics. The GDL however underperforms slightly on specificity of the MAB, and both sensitivity and specificity of the LIB. With the addition of geometric constraints to both loss functions, one notes that the average DICE coe cients for MAB and LIB respectively improve by approximately 2% for the MAB and the LIB. This performance increase of 2 %, despite being moderate, is of comparable magnitude to other studies such as Zhou *et al.* in 2019 [88] who experiment with modified CNNs and U-Nets and register performance increases of 0.5% to 3.3%, and Oktay *et al.* in 2017 [252] who experiment with anatomically constrained networks and register performance increases of 1.2 to 3.1%. Paired sample T-tests carried out for both MAB and LIB DICE, showed that the addition of geometric constraints respectively to CE and GDL functions yielded statistically significant di↵erences ($p < 0.05$). Conversely, paired sample T-tests carried out *between* CE and GDL functions, both with and without geometric constraints, did not yield statistically significant di↵erences ($p > 0.05$). The addition of geometric constraints also provides improvements to the average MHD coe cients for both MAB and LIB by approximately 0.010 to

0.070 mm respectively. The overall best performing combination is that of training with

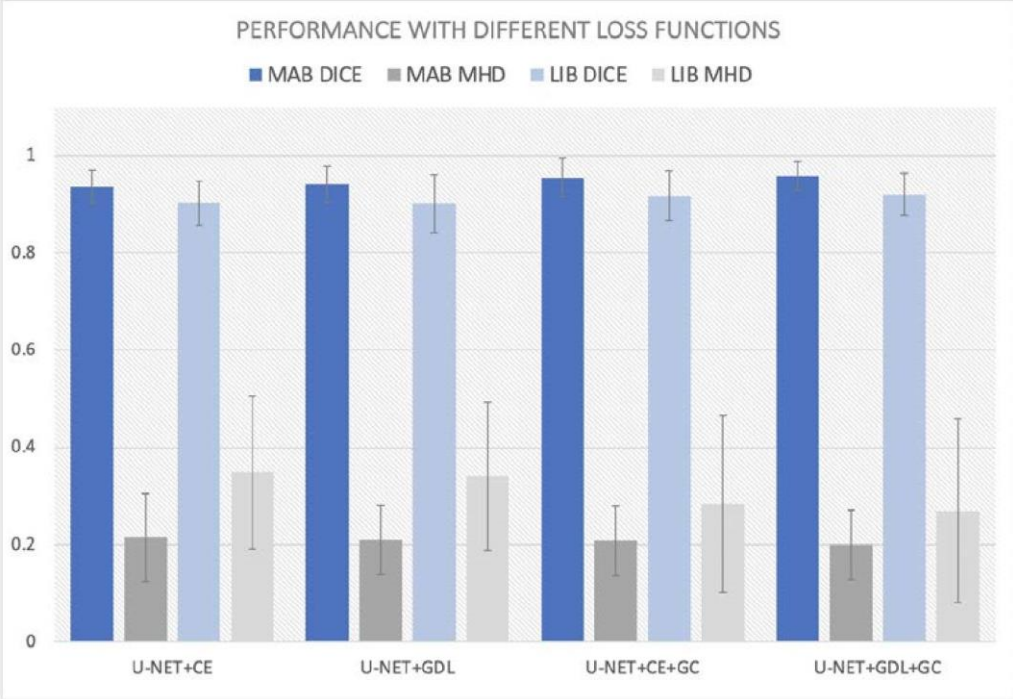

Figure 5.11: Perfomance metrics, DICE and MHD, from Table 5.2

Figure 5.12: Perfomance metrics, Sensitivity and Specificity, from Table 5.2 GDL and geometric constraints as opposed to training with CE and geometric constraints. The slight improvement is potentially due to the performance metric and loss function being both based on the same DICE coe cient, thus inherently biasing the result in this direction.

The results noted in Table 5.2 provide further comparative analyses with additional performance metrics of Sensitivity and Specificity. Here a general improvement is noted in average performance across sensitivity and specificity when adding the geometric constrains to both loss functions. One notes however that performances between CE and GDL with the additions of the constraints perform very similarly, with both techniques outperforming the other in di↵erent instances.

The images in Figure 5.13 show a qualitative result on some examples, comparing the outputs of the ground truth with that of the U-NET with GDL loss function and U-Net with GDL and geometric constraints. The results obtained exhibit a DICE similarity of between 93 - 95 % in case of the MAB and 88 to 91% in case of the LIB, in relation to the manual labelling. Although these are high values, they are prone to show imperfections in the segmentation process, particularly around the border of the MAB. In the first and second column, we may observe two instances where the U-NET with geometric constrains achieves an improved, 'fuller' result and without 'notches' in the LIB areas, in similar fashion to the ground truth. This is in contrast to the method without constraints, where we may observe notches or concavities along the contours. In the first, second and third column, we may observe three additional clear examples of improvements on the MAB contour. In all instances the technique without constraints produces a contour with notch defects, which are filled in or smoothened by the constrained network. These qualitative results

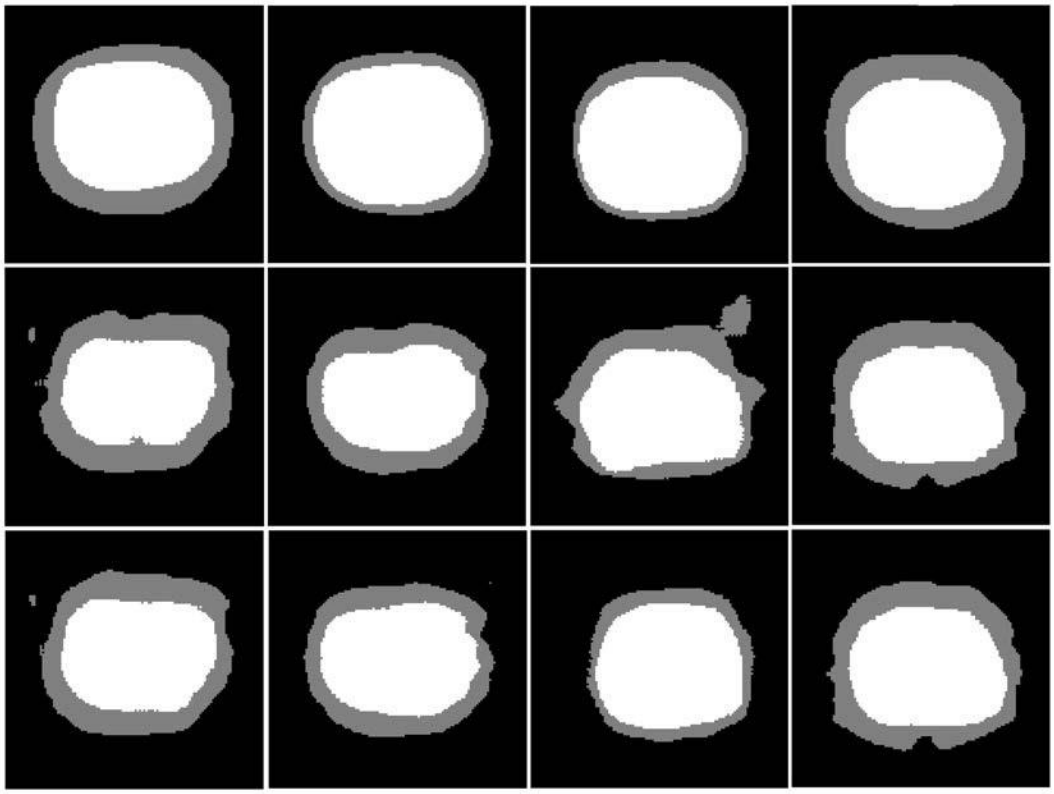

Figure 5.13: A sample of segmentation results of outer MAB and inner LIB borders. [Row 1] - Averaged Ground truth; [Row 2] - Segmentation with U-Net and GDL only; [Row 3] - Segmentation with U-Net + GDL + GC

are important, because although the 'quantitative' DICE performance of both methods is still fairly high and with largely similar coe cients, the geometric constraints imposed have implications on the shape produced, and result in smoother contours at the output. Of course, the results produced are not consistently perfect, and there are instances where both networks fail, such as that shown in the fourth column. Here, a slightly fuller LIB is produced, but the contour along the MAB in both techniques exhibits notch defects. Another set of examples is shown in Figure 5.14, which shows both the automatically obtained MAB and LIB contours (in green) and the manually obtained con-

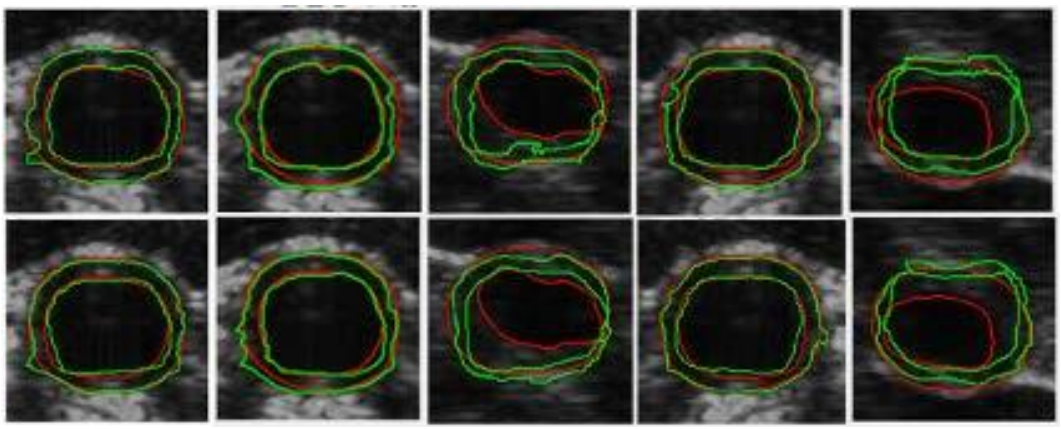

Figure 5.14: A sample of segmentation results of outer MAB and inner LIB borders. [Top row] - Examples without geometric constraints; [Row 2] - Same examples but with geometric constraints; Network obtained contours are shown in green while manually labelled contours are shown in red.

tours (in red) superimposed on the original B-mode image. The top row shows examples obtained without geometric constrains and the bottom row shows examples with the geometric constraints. As may be noticed, the application of geometric constraints reduces the notches and defects.

An interesting study is that of the filters which are produced during the training process of both U-NETs without constraints and with constraints. Unfortunately, with the networks having 28 layer-stacks, and several layers more for the banks of filters used within the stacks themselves (the filter bank in the middle convolution stack has 128 layers!) - it becomes a di cult task to reproduce and compare all these against each other. Instead therefore, a bank of 10 x 10 filters present on the final convolution layer stack is reproduced. This stack has three banks of 32 layers of 10 x 10 pixels each, one for each of

the output segmentation masks produced by the U-NET. Figure 5.15 and Figure 5.16 show two corresponding banks of 32 filters, for the U-NET without

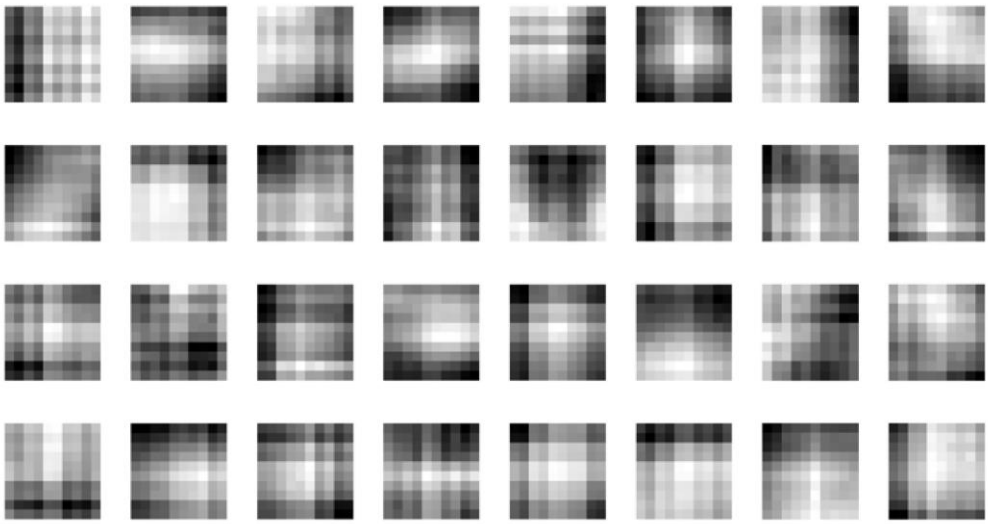

Figure 5.15: A bank of 32 output filters for U-NET without constraints.

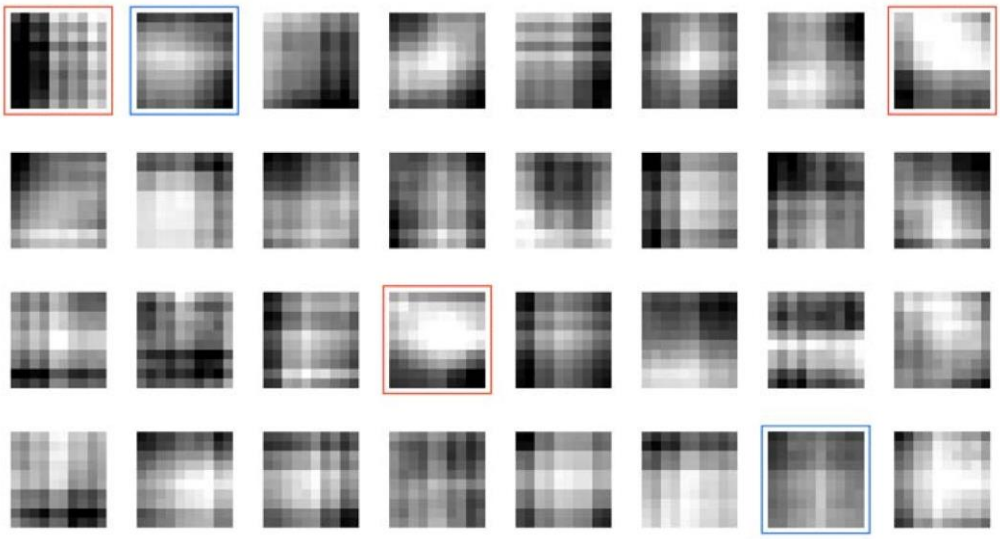

Figure 5.16: A bank of 32 output filters for U-NET with constraints.

constraints and the U-NET with constraints, respectively. The filter images are all normalised to the maximum and minimum values of the first filter bank;

that of the U-NET without constraints. This is done so that the di↵erences between them are more visible.

A comparison of the filters across both figures yields large similarities across the board, making it somewhat di cult to visualise any di↵erences at all. However, closer inspection of particular filters in Figure 5.16, specifically those marked with rectangular boxes, shows subtle di↵erences. The filters marked in red boxes show increased intensities over larger areas. The filters marked in blue boxes on the other hand, show a more blurred, softened version of their counterparts in the non-constrained filters. These types of di↵erences, and many others throughout the network which are not visualised here, could indicate the mechanisms in which the constrained network attempts to smoothen the contour edges or fill in gaps - akin to a low pass filtering e↵ect.

Table 5.3: Averaged results with standard deviation across author's own di↵erent techniques which attempt to segment MAB.

Method	**PCM-AC**	**CNN-CE**	**U-NET-CE-GC**
	.875 ± .067	$.955 \pm .040$	
±	.815 .150	$.208 \pm .072$	±
		0.411 .089	

PCM-AC: Phase Congruency Maps with Active Contours; CNN-CE: Traditional CNN network with Cross Entropy; U-NET-CE-GC: U-NET network with Cross Entropy and Geometric Constraints.

The results noted in Table 5.3 show the results of a performance comparison carried out between the author's proposed techniques in Chapter 3, 4, and this chapter. The Phase congruency maps with active contours yielded a DICE coe cient of 87.5%, which is worse than the performance registered in chapter 3.

This is because the technique was now applied on a much larger dataset of 18 patients as opposed to the small dataset used in Chapter 3. The traditional CNN network achieved a relatively good performance which is similar to that of the U-NET without geometric constraints reported in Table 5.1. The UNET with geometric constraints gave the best performance. Holm-Bonferroni corrected paired T-tests yielded statistically significant di↵erences across all three results ($p < 0.05$). The performance registered with MHD shows similar patterns to that for DICE.

Table 5.4: Averaged results with standard deviation across di↵erent studies which attempt to segment MAB and LIB.

Method	LSM [84]	SFL [73]	ASM [72]	DSL [63]	UNet [87; 88]	DCNN/MUNet [88]	T-CNN [88]	This work
Metric								
DSC MAB	.954 ± .016	.950 ± .017	.918 ± .035	.915 ± .035	.907 ± .062	.965 ± .022	.960 ± .026	.959 ± .029
DSC LIB	.931 .031 ±	.920 .042 ±	.936 .026 ±	.735 .169 ±	.895 .049 ±	.928 .045 ±	.863 .067 ±	.920 .044 ±

LSM: Level set method; SFL: Sparse field level sets method; ASM: Active shape models method; DSL: Distance regularised level set method; UNet: Traditional U-Net method; DCNN/MUNet: Dynamic convolutional neural network with modified U-Net; T-CNN: Traditional convolutional neural network.

The results noted in Table 5.4 and in Figure 5.17 (which shows only the best performing techniques in comparison to ours) show a comparative assessment of performance between the technique in this work and the techniques of other studies whose work similarly addressed the problem of MAB and LIB segmentation. Table 5.3 shows that this work's average results yield similar performance to that of Zhou *et al.* in their study in [88] using deep learning methods. Additionally, one notes that when comparing the performance of U-NETS in comparison to traditional CNNs, a general

improvement in performance is registered. However one must consider that this comes at increased computational cost. Figure 5.18 shows the increase in computational time incurred when using the proposed U-NET, in comparison to the CNN proposed in Chapter 4, which has the same amount of 'layer stacks', but only one set of convolutions per stack and no skip layers. The computational time of the U-NET is approximately double that of the CNN. The addition of the geometric constraints is tested in both CNN and U-NETs, and is also shown to marginally increase computational cost.

When comparing the methods used in this work to non-deep learning based methods, the technique outperforms that of Ukwatta *et al.* [84], Yang *et al.* [72] and that of Hossain *et al.*[63] when segmenting the MAB. But both this work and that of Zhou *et al.* underperforms against [84] and [72] in segmenting the LIB. In their work, Ukwatta *et al.* propose a semi-automated technique based on a level-set method, whereby the operator was asked to provide anchor points as high-level domain knowledge. In [72], the authors use active shape models to segment the MAB and LIB interfaces, but however require user intervention to provide anchor points during training. Zhou *et al.* also re-

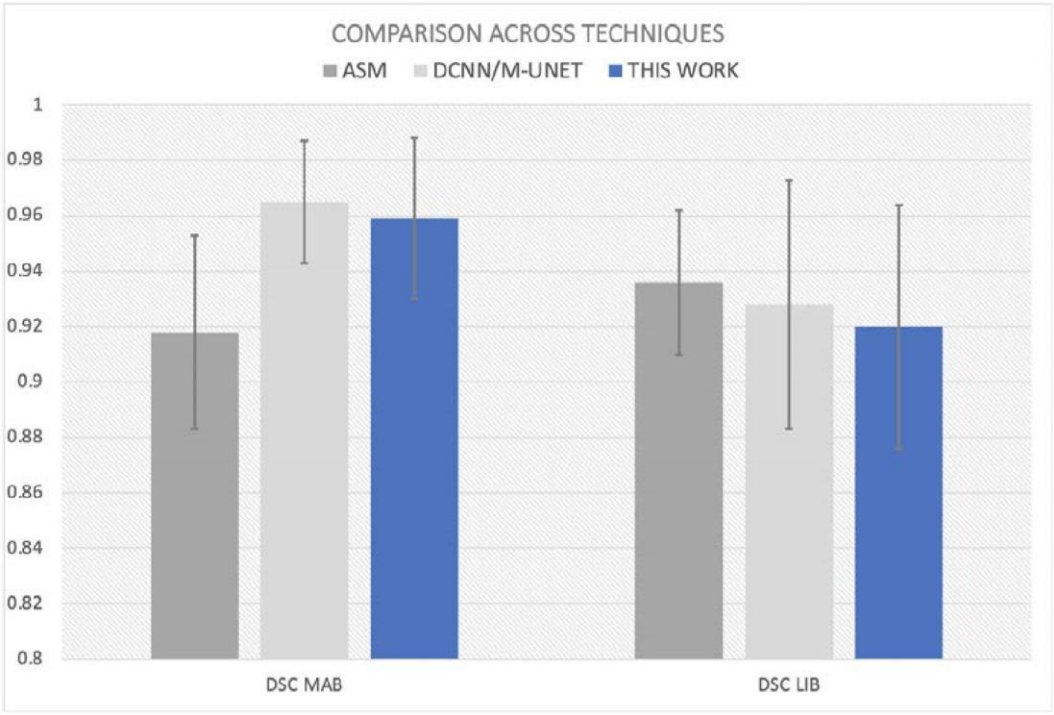

Figure 5.17: A graph comparing the top performing techniques for: LIB; based on ASM [72], and MAB; based on DCNN/MNET [88], to this work.

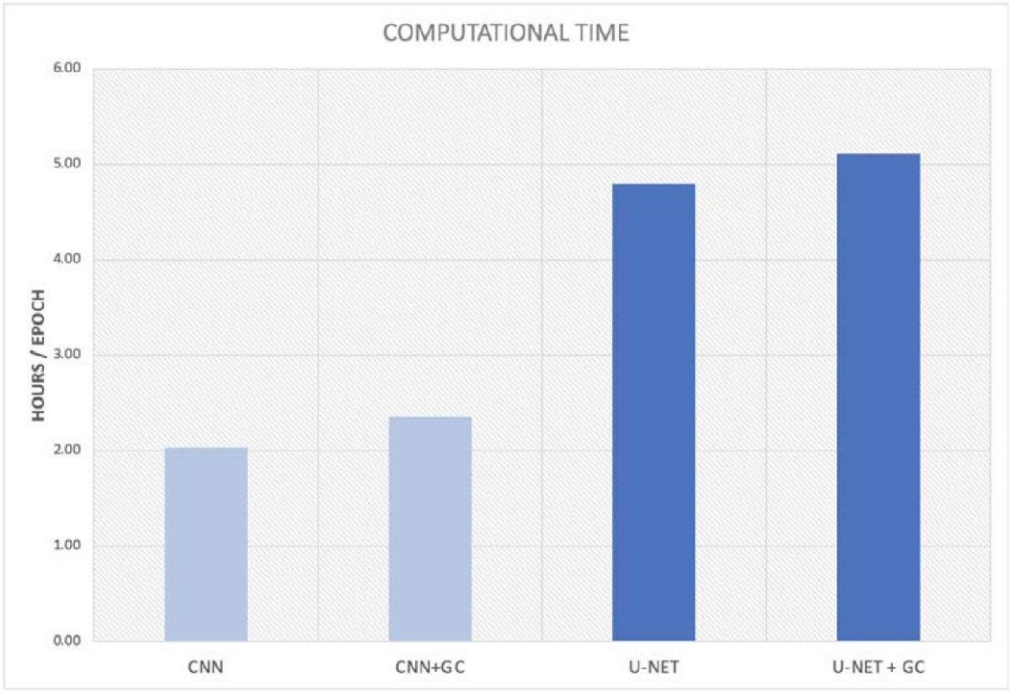

Figure 5.18: Cost of computational time: CNN vs U-NET

quire some degree of user interaction to provide anchor points to the dynamic CNN. In comparison, the technique proposed in this work is a fully automated technique, which requires no user intervention, barring the initial training on a manually segmented dataset that would have presumably been carried out once in the beginning with a large enough and representative dataset.

Some limitations are of course also present in this study. The size of the dataset was limited to only 18 patients, and thus there was not a significant number of patients within each stenotic category. A larger data set of highly stenotic and irregular images might likely result in a decrease in performance. Additionally, despite use of phase information to decrease dependence on intensity gain settings, the source of data is still largely dependent on the imaging protocol. Di↵erent imaging platforms employing di↵erent and proprietary image pre-

processing steps invariably result in di↵erent quality of images, and potentially, di↵erent performance outcomes for the algorithm. In an ideal scenario, the network would be trained on datasets obtained from di↵erent platforms and di↵erent users to increase generalisability. Imaging protocols are themselves also subject to several parameters, variables and user preferences, aside from simply brightness, which also a↵ect the quality of the data. These may also be considered as hyper-parameters of the overall system, aside from the hyper-parameters of the deep network itself which may be adjusted or tuned. Creating a system which is invariant to changes in all such hyper-parameters is a non-trivial task, and an exhaustive testing sequence which determines the e↵ect of each parameter, is impractical. An additional limitation is the dependence on expert labelling. The segmentation performance of the system is only as good as the manual labelling that is used for training. And inter-expert variability across studies naturally plays a role in the comparative assessment of performances across studies.

5.4 Conclusion

In this Chapter, a novel system has been developed based on a deep convolutional U-NET, for segmenting both the Media-Adventitia Boundary and the Lumen-Intima Boundary in transverse carotid Ultrasound images, using a fully automated approach. The previously proposed novel fusion of amplitude and phase data was used as a bimodal source of data at the input level, and a novel, geometrically constrained objective function was developed for the training of the U-NET. The author has shown that the addition of the geometric constraints results in an improvement of approximately 3% in terms of DICE coe cient of performance, in comparison to a U-NET trained with a standard

generalised dice loss function. Furthermore, the author has demonstrated that although the technique is fully automated, it retains a good performance of 95.9% and 92.0% for MAB and LIB borders respectively, and that this retains comparable performance to techniques from other studies.

6

Clinical Application and Proof of Concept

6.1 Introduction

In Chapter 5, a novel algorithm was proposed for the accurate measurement of vessel geometry to take place, allowing one to obtain contours representing the Media-Adventitia interface and the Lumen-Intima interface. This chapter shows how the proposed methodologies achieve superior performance in comparison to similar neural network-based methodologies, when it comes to computing the Vessel Wall Volume. Initially the present state-of-the-art biomarkers for evaluating clinical risks and outcomes associated with atherosclerosis are discussed. Subsequently the on-going paradigm shift from classic stenosis measurement techniques to plaque-burden measurement techniques involving metrics like Vessel Wall Volume are discussed. Finally, it is then shown how the outputs obtained from the proposed algorithm in Chapter 5 may be used to compute such metrics, and their reliability is evaluated and discussed.

6.2 Measures of Cardiovascular Risk

The measure of Carotid Intima Media Thickness (IMT) as a predictor of cardiovascular risk was proposed in the 1980s, and it has since become the *defacto* imaging marker to be used to gauge cardiovascular risk associated with atherosclerosis [74]. Many reports have relied on this marker for studies associated with risk factors, genetics, and evaluation of lipid-lowering medication. Important endarterectomy trials (ECST, NASCET, ACAS) carried out in the 1980s to mid-1990s established the basis for considering the degree of stenosis as the gold standard metric of stratifying subsequent stroke risk [64]. Even until recently, studies still show that carotid IMT together with the presence of carotid plaque are accurate predictors of cardiovascular events in patients with critical limb ischaemia [74].

Concurrently however, the concept of measuring the degree of luminal stenosis as a marker of atherosclerotic disease severity has been criticised by experts, due to the fact that some plaques producing only mild-to-moderate stenosis may still lead to cerebral infarctions [64]. Plaque is known to progress at a much faster rate along the vessel wall (in the direction of blood flow), in comparison to the rate at which it thickens the vessel wall [63]. Carotid stenosis alone is therefore a weak indicator of the volume and extension of carotid plaque. There are additional plaque features that are closely tied to ischaemic events, and these features significantly increase the risk of stroke, regardless of the degree of stenosis [64].

According to Kosmas *et al.* [74], it is a great mystery as to why investigators so commonly keep using carotid IMT for risk prediction, when it is no longer recommended by the American Journal of Cardiology and American Heart

Association - being deemed inferior to other metrics, such as coronary artery calcium score or carotid plaque detection. There is need for a paradigm shift to more advanced 3D technologies, which may be used to assess plaque burden in terms of plaque volume, area, or vessel volume. Indeed, numerous recent studies are investigating the use of volumetric methods for vulnerable plaque imaging [64].

Two such volumetric methods are those of computing the Total Plaque Volume (TPV) and Vessel Wall Volume (VWV). The total plaque volume had already been proposed as a metric to monitor plaque progression in 2004, by Ainsworth *et al.* [48]. In their work, plaque area was measured by manual planimetry using an in-house software across several slices of a 3D image. The areas were summed and multiplied by the interslice distance. The authors concluded that TPV was more sensitive to change in plaque over time than methods limited to measurement of change of thickness. In 2007, the group of Fenster *et al.* [68] however, argued that while TPV provides valuable quantitative information, obtaining its measure was di cult and required highly trained experts. It also did not provide information regarding the locations where the vessel wall is changing. On this basis, they proposed the Vessel Wall Volume, which is a measure of the vessel wall thickness and plaque together, between the LumenIntima and Media-Adventitia. These two interfaces are more straightforward to interpret and more reproducible. The authors showed that VWV had lower intra-observer coe cients of variation in comparison to TPV [68].

In line with recent e↵orts to shift to volume based assessment strategies, recent technical literature, which also studied carotid artery segmentation,

proposed to segment the MAB and LIB interfaces for the purpose of the computation of Vessel Wall Volume [63], [88]. As a proof of concept, the author thus applies the outputs of the algorithm in Chapter 5 towards the computation of VWV.

6.3 Computation of Vessel Wall Volume

Vessel wall volume is a 3D measurement of vessel wall thickness, inclusive of plaque where present. It is measured by first segmenting the MAB and the LIB, and determining the area of the region in between these two contours. Previous methods would calculate the area using the trapezoidal rule [253], however with MATLAB, we simply sum over the number of pixels within the region. In order to do this, we must determine the lateral or axial square dimension of each pixel. For acquisition of the images, the probe used was a model L14-5/38 linear probe, which is known to have 128 piezoelectric crystal elements, each with a lateral pitch of 300 μm. The overall scanning width of the probe therefore, which also corresponds to the width of the B-mode image width.

Figure 6.1: Probe width corresponds to 429 pixels

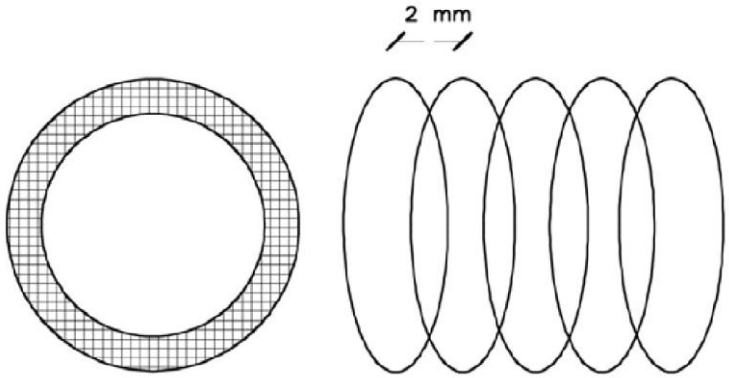

Figure 6.2: The vessel wall volume is computed by summing the product of individual frame area (hatched area) by the inter-slice distance.

the image, is 128 x 300 μm, which is 38.4mm. This is shown in Figure 6.1. We additionally know that the B-mode images are scan-converted to a final image dimension of 429 (lateral) x 342 (axial) pixels. Thus, 429 pixels must equal 38.4mm, leading to a single pixel having a square dimensions of 0.0895 mm. We further confirm this by computing the axial dimension of 342 pixels x 0.0895 mm/pix, which is equal to 30mm. This is in agreement with the depth setting of 3.0 cm set during the acquisition.

Proof of Concept. In order for Vessel Wall Volume to be computed, it is necessary for a number of image slices to be acquired in transverse view over some length *x*. The acquisition over this distance is normally acquired either using a regular freehand 2D probes, or else using 4D probes which sweep a short distance mechanically within the probe while acquiring scans. In case of freehand 2D probes, the position is tracked using electromagnetic, optical, or

accelerometer-based position trackers, or else by fixing the probe to a physical linear gauge placed over the patient, and carefully moving it along a linear

Figure 6.3: The electromagnetic transmitter is positioned at the head of the couch, closest to the imaging area as possible.

distance. Either method allows the user to acquire transverse slices along the length of the artery, and thus knowing the distance between each slice as shown in Figure 6.2. This distance is referred to as the inter-slice distance.

The work of this project focused on the image segmentation aspect of the problem, and therefore developing a rigorous position tracking scheme was beyond the scope of the work. This would have entailed devising calibration methods which apply geometric transforms from the position sensor's coordinate system to the 3D coordinate system within the imaging space.

Instead the author opted, merely as a basic proof of concept, to utilise an electromagnetic position tracking system developed by NDI, trakSTAR, to sense the position of the probe in the direction along the length of the carotid artery. Yaw, pitch

Figure 6.4: [Left] The trakstar base amplifier unit. [Right] The position sensor a xed to the ultrasound linear probe with double-sided tape.

and roll movements were manually kept fixed as much as possible, as were movements in the *y* and *z* plane. The EM transmitter was placed at the head of the couch, close to the head of the patient, since this was the site closest to the examination area. This is shown in Figure 6.3. Any equipment unrelated to the scanning was kept far away to reduce potential interference. The small EM receiver, which is used to track the position of the actual object, was a xed to the linear probe using double sided tape as shown in Figure 6.4. Both transmitters and receivers were connected to a base unit which feeds the data into a computer. The probe was gently moved along the direction of the carotid artery in 2mm steps, using the EM position tracker as guidance, and with the help of an assistant, images were saved at each interval. The probe's relative

position and yaw, pitch and roll were read o↵ in real time from the tracker software, shown in Figure 6.5. The probe was moved through a distance of 40 mm, thus resulting in 20 images. The process was repeated for a sample of 5 healthy participants from the laboratory, which formed part of the cohort of 18 participants used for the whole project.

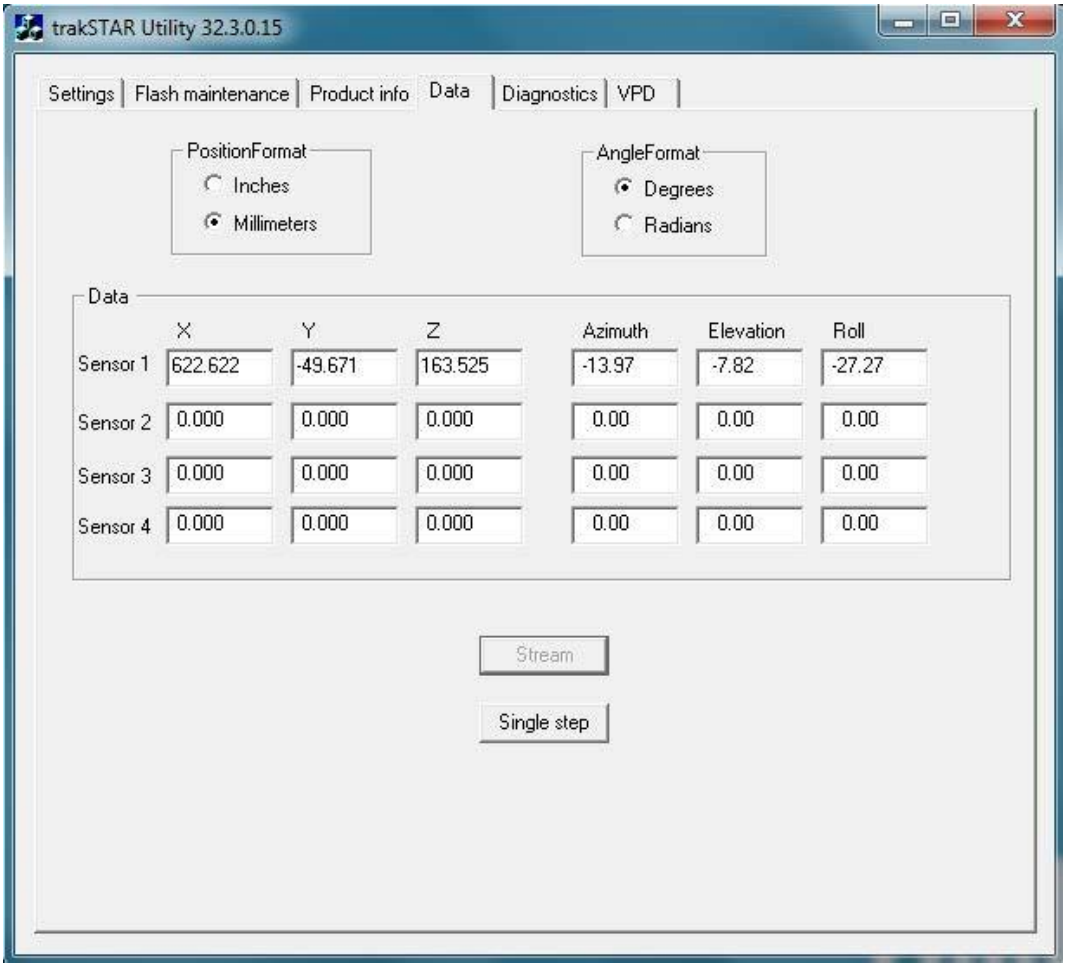

Figure 6.5: The trakstar software provides the relative position in 3D space.

6.4 Discussion

Since Vessel Wall Volume is reported to be sensitive to temporal changes in plaque found in the carotid, we report the performance of the algorithm in computing this metric from the automatically delineated contours, and compare this against the same metric obtained from the manually delineated contours.

Table 6.1 shows the performance of the algorithm by evaluating $\overline{VWV}$ and comparing against results from three other recent studies by Ukwatta *et al.* [73], Hossain *et al.* [63] and Zhou *et al.* [88]. The latter metric is computed by first taking the di↵erence between the automatically generated volume VWV_a^i and manually obtained volume VWV_m^i, divided by the manually obtained volume (ground truth), over a number of measurements for the same patient *i* = 1

[84]:

$$\Delta VWV^i = \frac{\overline{VWV_a^i} \quad \overline{VWV_m^i}}{\overline{VWV_m^i}} \tag{6.1}$$

where $\overline{VWV_a^i}$ and $\overline{VWV_m^i}$ are the mean VWV measurements computed from three repeated measurements of VWVs for the algorithm and manual method respectively, for one particular patient. The overall mean VWV di↵erence for

all patients, $\overline{VWV}$, is then computed as [84]:

$$\overline{\Delta VWV} = \frac{\sum_{i=1}^{N} \Delta VWV^i}{N} \tag{6.2}$$

where N is now the total number of patients. Additionally, we also compute the absolute VWV di↵erence, $|VWV^i|$ for patient i, which is simply the absolute term of VWV^i. We then report the geometric mean over all patients, $|\overline{VWV}|$. Furthermore, in Figure 6.6 we plot a correlation plot of the algorithm and manually generated VWV for the 5 data sets on which this is tested. In Figure 6.7, we show a Bland-Altman plot of the di↵erence between automatically and manually generated volumes, whereby the bias line is plotted along with the upper and lower limits of agreement as 1.96 x standard deviation.

Table 6.1: Averaged results with standard deviation of percentage $\overline{VWV}$ and $|\overline{VWV}|$ across di↵erent studies, including this work.

Method	**3D-SFLS** [73]	**DRLS [63]**	**DCNN [88]**	**This work**
Metric				
$\overline{VWV}$ (%)	0.56 ± .12.42	0.645 ± 4.93	1.62 ± 8.70	4.91 ± 2.17
$\lvert\overline{VWV}\rvert$ (%)	5.64 8.1 ±	3.43 3.42 ±	6.48 6.14 ±	4.45 1.68 ±

3D-SFLS: 3D Sparse Field Level set method; DRLS: Distance regularised Level set method; DCNN: Dynamic CNN and U-NET

From the table of results, one notices that the mean di↵erence in VWV,

$\overline{VWV}$, is higher in the proposed algorithm in comparison to that of other studies. The reason for this is that in other studies, the VWV computed by the algorithm is sometimes larger and sometimes smaller than that computed manually. This leads to positive and negative values, which when averaged lead to a small mean value. In this work's proposed algorithm a bias is present, whereby the automated algorithm's VWVs are always slightly larger than that computed manually. This is also evident from the Bland-Altman plot in Figure 6.7, which shows a mean bias of $29mm^3$. However, when one compares the mean absolute difference $|\overline{VWV}|$, we notice that here the proposed algorithm surpasses the performance of the techniques by Zhou *et al.* and Ukwatta *et al.*. The comparison against the technique of Zhou *et al.* is particularly relevant since the latter also use deep networks in their methodology of segmentation and VWV estimation. In Figure 6.6, we see that the values for VWVs obtained from the algorithm and manual methods are highly correlated.

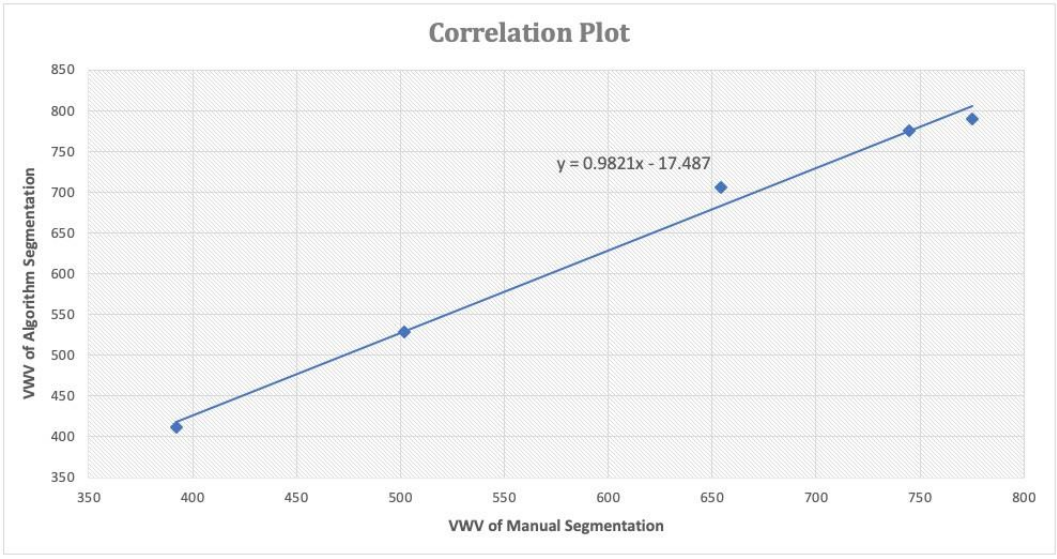

Figure 6.6: Correlation plot for automated and manually generated 3D vessel wall volumes.

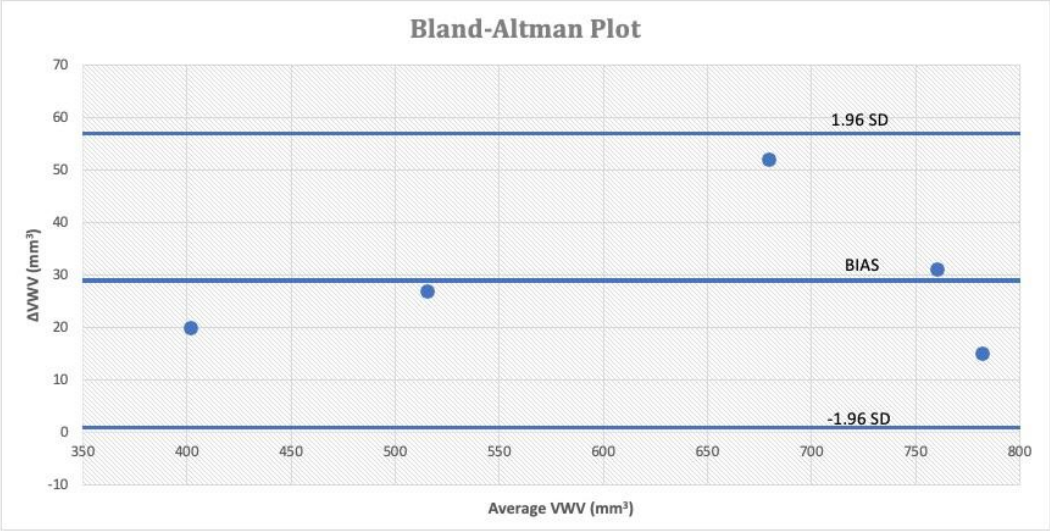

Figure 6.7: Bland-Altman plot of the di↵erence between the VWVs generated by the automated algorithm and the manual method.

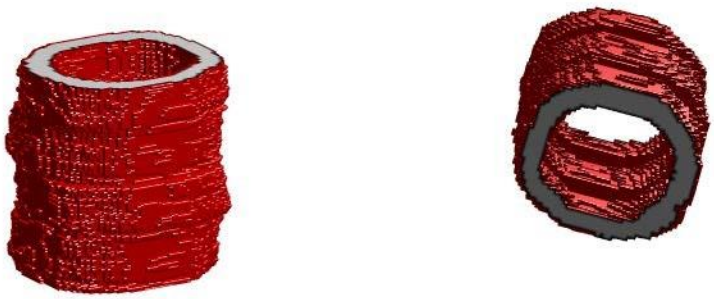

Figure 6.8: A volume render of the two segmented contours being shown from di↵erent orientations

Lastly, Figure 6.8 shows a volume render from two perspectives of the vessel wall volume taken from a particular dataset. This volume is generated in MATLAB, and shows how a qualitative reproduction of the vessel wall volume would be like for visual analysis by clinicians. A visual representation such as that in Figure 6.8 would be useful for the clinicians to visualise the location of any protruding plaque areas, and to therefore plan subsequent treatment accordingly. Since the acquisition of images was not gated to the subjects' ECG signal (and neither was that of comparable studies which the author looked at the time of preparing this experiment [84; 72; 85]), the vessel wall in 3D is seen to present with slight undulations. This is an artefact that arises from the expanding and contracting nature of the vessel wall in tandem with the cardiac cycle, and it may have a bearing on the final volume measured by the algorithm.

Cheng *et al.* [85] argue that gating is not necessary, since the arteries of subjects su↵ering from atherosclerosis are not compliant and exhibit very little radial changes. However, while this may be true, if the metric would be used as a risk prediction tool it would be envisaged that it would be used also on subjects with complaint arteries, and that the undulations would then be present and generate artifactual readings. The ideal solution to avoid such an artefact impacting the volume measurement would be to gate the image acquisition with the same point in the cardiac cycle, to ensure that the volume is constructed accurately. Alternatively, when acquiring natively in 3D, if a volumetric cine loop is acquired, this would allow post processing algorithms to select images at their peak of extension or contraction to have them selected at the same point in time.

6.5 Conclusion

This chapter has shown that the proposed algorithm yields superior results when used to produce measures of Vessel Wall Volume, in comparison to other neural network based techniques. It has described how the vessel wall volume, a novel metric proposed in literature to monitor plaque progression, is constructed from the outputs of the algorithm. Its use has been highlighted, particularly justifying its need in comparison to other standard methods of plaque assessment. Furthermore, it has been shown that the proposed automated UNET-based algorithm is capable of producing delineations, which result in a VWV that is in agreement with that produced manually by the expert. Finally, it has been shown that the performance obtained by the automated algorithm in producing VWVs, is comparable to that of other techniques in the literature but superior to those also based on neural networks. The next chapter will provide concluding thoughts on the whole work, summarising contributions, and discussing the limitations of this project and future work.

7

Conclusion

7.1 Introduction

This final chapter shall summarise the contributions produced in this project. It shall discuss the limitations encountered during the various stages of the project, and propose some future work which could be looked into in order to bring the algorithm closer to clinical implementation.

7.2 Summary of Contributions

The overarching target of this project was to quantify atherosclerosis in vasculature using ultrasound imaging. In order to achieve this goal, the author set out to develop an algorithm, which is capable of generating a 3D volume of a vascular structure. From this, a volume-base metric is produced, which is associated with presence and progression of plaque. The said metric is the vessel wall volume, which has been shown in literature to be a valid tool to monitor plaque progression.

In order to generate the 3D volume of the artery, a technique, which is both robust to inter-user variations in settings and which is clinically simple enough to implement, is required. For this reason, the research objectives included: a) the requirement of developing a technique which is invariant to amplitude variations that may come about from di↵erent user-gain settings, and b) the requirement that it should also be as far as possible, automated, such that it requires minimal user intervention. These objectives have been met through the following contributions.

1. **Developing RF derived Phase Congruency Maps of Carotid Ultrasound images**

   The author has initially established whether phase information present in ultrasound RF data is useful for representing anatomical structures within the data. This was achieved by asserting that phase di↵erence information taken from RF data produces images, which qualitatively show tissue structure. An algorithm has then been proposed, which produces Phase Congruency Maps from the ultrasound data. Phase Congruency Maps are arrays of phase information, which are invariant to amplitude / gain settings, and also provide structural information about the underlying anatomy. It has also been shown that the latter are superior to Phase Di↵erence Matrices for the purpose of representing structural information.

   A part of the research in this section has been published in: C. Azzopardi, K. P. Camilleri and Y. A. Hicks, ”Carotid Ultrasound Segmentation using Radio-Frequency derived Phase information and Gabor filters,” 2015

37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Milan, 2015, pp. 6338-6341.

2. **Carotid Artery Segmentation using Deep Convolutional Networks and Phase Congruency Maps**

   In order to produce the vessel wall volumes, a simple to use segmentation technique was required to delineate the vessel contours. The author has thus proposed the use of deep neural networks, the present state-of-theart in segmentation strategies, to delineate the vessel in an automated manner, which requires only the selection of a region of interest. Selecting ROIs is already a commonly available tool in commercial ultrasound machines. Apart from being the first to propose the application of DCNs to transverse ultrasound carotid image segmentation, the author proposed their application on the fusion of amplitude and phase information. It has been demonstrated that this yields improved results over using either alone.

   A part of the research in this section has been published in: C. Azzopardi, Y. A. Hicks and K. P. Camilleri, "Automatic Carotid Ultrasound Segmentation using Deep Convolutional Neural Networks and Phase Congruency Maps," 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017), Melbourne, VIC, 2017, pp. 624-628.

3. **Bimodal Automated Carotid Ultrasound Segmentation using Geometrically Constrained Convolutional Neural Networks**

Finally, the author has proposed a novel modification to the optimisation function of the neural network, by applying a priori knowledge concerning

the shape of vessels, and by also adapting the neural network to simultaneously segment both the Media-Adventitia interface and the Lumen-Intima interface, thus having both contours necessary for obtaining the vessel wall volume. These contours, obtained along the length of the carotid artery, are used to generate the vessel wall volume. The author further shows that the vessel wall volume generated shows good agreement with that generated manually by experts.

A part of the research in this section has been published in: C. Azzopardi, Y. A. Hicks and K. P. Camilleri, "A Bimodal Automated Carotid Ultrasound Segmentation using Geometrically Constrained Convolutional Neural Networks," 2019, IEEE Journal of Biomedical and Health Informatics.

7.3 Limitations

The principal limitation in this work was the lack of availability of subject volunteers with registered symptomatic presentation of plaque in carotid arteries. Although the help of a local vascular surgeon was enlisted for this purpose, the author was advised that patients who, upon examination at the local government hospital, were found to have significant presentation of atherosclerotic plaque in their carotid arteries, were immediately booked for carotid endarterectomies. Due to data protection issues, the author was also not allowed to contact patients who had known presence of plaque but who were as yet ineligible for surgical intervention. Additionally, since the use of the Ultrasonix ultrasound device for extraction of RF data was required, it proved inconvenient for the vascular surgeon to either use the latter machine (which was inferior to his clinical one), or to have the author accompany him

in his clinics to repeat the scans of fresh patients coming in for examination and who were found on the spot to have significant presentation of plaque.

Thus, the author resorted to recruiting at random, both young and elderly patients, without knowing whether these would have presence of plaque or otherwise. From the pool of volunteers recruited, it was indeed established that elderly patients had an increased vessel wall volume in comparison to younger patients. This is consistent with the presence of accumulated plaque build-up. However, only one patient was ultimately found to have a significant occlusion. The limitation that the above poses was that the author was thus unable to fully test the algorithm on known symptomatic patients with stenosis > 60%, and to therefore establish the performance of the technique in automatically flagging patients which might need treatment. Furthermore, the author was unable to include larger quantities of ultrasound data in the network's training set with significant stenosis. These typically pose a more di cult segmentation problem due to shadowing caused by the plaque on vessel wall distal to the probe.

An additional limitation in the project was the hardware used for training and testing the deep neural networks. Despite the machines used having a CUDA enabled graphical processing unit, or parallel processing capabilities, the performance of the machines were such that training and testing of the deep networks was a lengthy and laborious process, making it highly impractical to fully test the available hyper-parameter space of the network. In future work, this may be resolved through the use of cloud computing subscriptions.

Finally, from the perspective of ultrasound hardware, it is to be noted that modern ultrasound platforms are highly superior to the research ultrasound

platform used, which was already partially dated at the time of commencement of research. Modern platforms provide improved resolution, and volumetric matrix probes, which would be able to acquire volumes natively, without the need for position tracking hardware. This would naturally provide a more accurate reconstruction of segmented volume, which in turn yields more accurate vessel wall volumes. Additionally, with the availability of raw data already in volumetric format, one may do away with segmenting individual slices and consider the native segmentation of the volumetric space with the deep neural networks.

7.4 Future Work

The author believes that there are three important areas which merit immediate further attention in order to bring automated ultrasound plaque visualisation and evaluation techniques to market:

Firstly, from a technical perspective, with the resurgence of artificial intelligence and its excellent performance in most fields of computer science; with the increasingly readily available parallel processing architectures made available through powerful graphical processing units; and with volumetric matrix probes with single crystal technology achieving enhanced resolution and signalto-noise ratio, future work should focus on using voxel-based convolutional neural networks to segment the carotid structures natively in the volumetric space. Indeed, research published as recently as September 2020 by Zhou *et al.* [89], is already exploring the use of voxel-based fully convolutional networks for this application. The geometric constraints proposed in this work could potentially be adapted to reflect new constraints

and a priori knowledge present in a volumetric problem. Additionally, the fusion strategy adopted in this work limited itself to using input level fusion. The performance of layer-level or decision-level fusion of amplitude and phase information is intriguing and should also be explored to evaluate if improved performances are registered.

Secondly, also from a technical perspective, in order for such algorithms to be truly considered e↵ective, there is the need for the establishment of a large and common dataset, to be made available to the research community at large which is exploring the same problem. A large and continuously updated dataset containing manually labelled ultrasound images / volumes would provide a common reference for research groups to be able to test their algorithms against it. Without this, research groups presently continue to develop and test algorithms against di↵erent datasets, which undoubtedly vary in quality and acquisition protocol, and this may have an e↵ect on the final performance reported.

Finally, from a clinical perspective, and as mentioned by Kosmas *et al.* in [74], the clinical research community needs to actively transition away from the classic stenosis risk prediction strategies based on carotid IMT, given that this is no longer recommended by the American Journal of Cardiology or American Heart Association. Future work in the clinical community should therefore focus more on carrying out clinical studies on the use of 3D volumetric parameters, as it is in this manner that su cient momentum and confidence in such metrics may be achieved for them to become widely implemented in commercial platforms.

Bibliography

[1] W. S. O. World Health Organisation, World Heart Federation, “Global atlas on cardiovascular disease prevention and control,” World Health Organization, Tech. Rep., 2011.

[2] R. L. S. M. B. W. H. S. M. J. P. V. F. M Patrice Lindsay, Bo Norrving, “Global stroke fact sheet 2019,” World Stroke Organization, Tech. Rep., 2019.

[3] W. H. Organisation, “Cardiovascular diseases,” March 2020. [Online]. Available: https://www.who.int/health-topics/cardiovasculardiseasestab=tab1

[4] H. C. Stary, a. B. Chandler, S. Glagov, J. R. Guyton, W. Insull, M. E. Rosenfeld, S. a. Scha↵er, C. J. Schwartz, W. D. Wagner, and R. W. Wissler, “A definition of initial, fatty streak, and intermediate lesions of atherosclerosis. A report from the Committee on Vascular Lesions of the Council on Arteriosclerosis, American Heart Association,” *Arterioscler.*

Thromb. Vasc. Biol., vol. 14, no. 5, pp. 840–856, may 1994. [Online]. Available: http://atvb.ahajournals.org/cgi/doi/10.1161/01.ATV.14.5.840

[5] H. C. Stary, A. B. Chandler, R. E. Dinsmore, V. Fuster, S. Glagov, W. Insull, M. E. Rosenfeld, C. J. Schwartz, W. D. Wagner, and R. W. Wissler, “A Definition of Advanced Types of Atherosclerotic Lesions and a Histological Classification of

Atherosclerosis : A Report From the Committee on Vascular Lesions of the Council on Arteriosclerosis, American Heart Association," *Circulation*, vol. 92, no. 5, pp. 1355–1374, sep 1995. [Online]. Available: http://circ.ahajournals.org/cgi/doi/10.1161/01.CIR.92.5.1355

[6] K. Skagen, M. Skjelland, M. Zamani, and D. Russell, "Unstable carotid artery plaque: new insights and controversies in diagnostics and treatment," *Croatian medical journal*, vol. 57, no. 4, pp. 311–320, 08 2016. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/27586546

[7] V. S. S., A. Alvaro, A. H. J., B. E. J., B. M. S., C. C. W., C. A. P., C. A. M., C. Susan, D. F. N., E. M. S. V., E. K. R., F. J. F., G. D. K., K. S. S., K. B. M., K. K. L., L. C. D., L. Ten´eT., L. Junxiu, L. M. Shane, L. P. L., M. Jun, M. Jason, M. S. S., M. D. B., M. M. E., N. S. D., P. A. Marma, R. G. A., S. Zainab, S. G. M., S. E. B., S. S. H., S. C. M., S. Andrew, V. L. B., W. Nae-Yuh, T. C. W., and null null, "Heart disease and stroke statistics—2021 update," *Circulation*, vol. 143, no. 8, pp. e254–e743, 2021/03/14 2021. [Online]. Available: https://doi.org/10.1161/CIR.0000000000000950

[8] L. Waite and J. Fine, *Applied biofluid mechanics*. New York: Mc Graw Hill, 2007.

[9] T. Truelsen and S. a. Begg, "The global burden of cerebrovascular disease," *World Health Organization*, 01 2006.

[10] A. Naser, K. E. A., M. R. D., C. Dong, T. J. I., G. D. E., H. R. M., M. T. G., F. P. A., and W. W. E., "Left atrial blood stasis and von willebrand factor– adamts13 homeostasis in atrial fibrillation," *Arteriosclerosis, Thrombosis, and Vascular*

Biology, vol. 31, no. 11, pp. 2760–2766, 2021/03/14 2011. [Online]. Available: https://doi.org/10.1161/ATVBAHA.111.232991

[11] E. J. Donald, S. J. L., A. G. W., A. M. J., C. Seemant, F. Edward, H. T. S., H. R. T., J. S. Claiborne, K. C. S., L. H. L., M. Elaine, and S. R. L., "Definition and evaluation of transient ischemic attack," *Stroke*, vol. 40, no. 6, pp. 2276–2293, 2021/03/14 2009. [Online]. Available: https://doi.org/10.1161/STROKEAHA.108.192218

[12] V. Scanlon and T. Sanders, *Understanding Human Structure and Function*. F.A. Davis, 1997. [Online]. Available: https://books.google.com.mt/books?id=7EyRHAAACAAJ

[13] G. J. Tortora, B. Derrickson, and J. W. . Sons, *Principles of anatomy physiology*, 2017.

[14] G. D. Hammer, *Pathophysiology of disease : an introduction to clinical medicine*, seventh edition ed., S. J. McPhee, Ed. New York: McGraw-Hill Education Medical, 2014. [Online]. Available: https://search.library.wisc.edu/catalog/9910207956102121

[15] A. Jaminon, K. Reesink, A. Kroon, and L. Schurgers, "The role of vascular smooth muscle cells in arterial remodeling: Focus on calcification-related processes," *International Journal of Molecular Sciences*, vol. 20, no. 22, 2019. [Online]. Available: https://www.mdpi.com/1422-0067/20/22/5694

[16] F. C. G. van Bussel, B. C. T. van Bussel, A. P. G. Hoeks, J. Op 't Roodt, R. M. A. Henry, I. Ferreira, F. H. M. Vanmolkot, C. G. Schalkwijk, C. D. A. Stehouwer, and K. D. Reesink, "A control systems approach to quantify wall shear stress normalization by flow-mediated dilation in the brachial artery," *PLOS ONE*, vol.

10, no. 2, pp. e0115977–, 02 2015. [Online]. Available: https://doi.org/10.1371/journal.pone.0115977

[17] P. Libby, “Inflammation and Atherosclerosis,” *Circulation*, vol. 105, no. 9, pp. 1135–1143, mar 2002.

[18] R. J. Henning, M. Bourgeois, and R. D. Harbison, “Poly(adp-ribose) polymerase (parp) and parp inhibitors: Mechanisms of action and role in cardiovascular disorders,” *Cardiovascular Toxicology*, vol. 18, no. 6, pp. 493– 506, 2018. [Online]. Available: https://doi.org/10.1007/s12012-018-9462-2

[19] M. F. O’Rourke and J. Hashimoto, “Mechanical factors in arterial aging: a clinical perspective.” *J Am Coll Cardiol*, vol. 50, no. 1, pp. 1–13, Jul 2007.

[20] N. A. Herity, M. R. Ward, S. Lo, and A. C. Yeung, “Review: Clinical aspects of vascular remodeling.” *J Cardiovasc Electrophysiol*, vol. 10, no. 7, pp. 1016–1024, Jul 1999.

[21] H. C. Stary, D. H. Blankenhorn, A. B. Chandler, S. Glagov, W. Insull, M. Richardson, M. E. Rosenfeld, S. A. Scha↵er, C. J. Schwartz, and W. D. Wagner, “A definition of the intima of human arteries and of its atherosclerosis-prone regions. A report from the Committee on Vascular Lesions of the Council on Arteriosclerosis, American Heart Association,” *Circulation*, vol. 85, no. 1, pp. 391–405, jan 1992. [Online]. Available: http://circ.ahajournals.org/cgi/doi/10.1161/01.CIR.85.1.391

[22] J. R. Guyton, “Phospholipid Hydrolytic Enzymes in a ’Cesspool’ of Arterial Intimal Lipoproteins : A Mechanism for Atherogenic Lipid Accumulation,” *Arterioscler. Thromb. Vasc. Biol.*, vol. 21, no. 6, pp. 884–886, jun 2001. [Online]. Available: http://atvb.ahajournals.org/cgi/doi/10.1161/01.ATV.21.6.884

[23] P. T. Kovanen, “Atheroma formation: defective control in the intimal round-trip of cholesterol.” *Eur. Heart J.*, vol. 11 Suppl E, pp. 238–46, aug 1990. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/2226531

[24] L. National Heart and B. Institute. (2018) Atherosclerosis. [Online]. Available: https://www.nhlbi.nih.gov/health-topics/atherosclerosis

[25] N. I. for Health and C. Excellence, “Nice pathways - stroke overview,” March 2021. [Online]. Available: https://pathways.nice.org.uk/pathways/stroke

[26] I.-. C. Group, “The benefits and harms of intravenous thrombolysis with recombinant tissue plasminogen activator within 6 h of acute ischaemic stroke (the third international stroke trial [ist-3]): a randomised controlled trial,” *The Lancet*, vol. 379, no. 9834, pp. 2352 – 2363, 2012. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0140673612607685

[27] A. R. Naylor, J. B. Ricco, G. J. de Borst, S. Debus, J. de Haro, A. Halliday, G. Hamilton, J. Kakisis, S. Kakkos, S. Lepidi, H. S. Markus, D. J. McCabe, J. Roy, H. Sillesen, J. C. van den Berg, F. Vermassen, E. G. Committee, P. Kolh, N. Chakfe, R. J. Hinchli↵e, I. Koncar, J. S. Lindholt, M. Vega de Ceniga, F. Verzini, E. G. Reviewers, J. Archie, S. Bellmunt, A. Chaudhuri, M. Koelemay, A. K. Lindahl, F. Padberg, and M. Venermo, “Management of atherosclerotic carotid and vertebral artery disease: 2017 clinical practice guidelines of the european society for vascular surgery (esvs),” *European Journal of Vascular and Endovascular Surgery*, vol. 55, no. 1, pp. 3–81, 2021/03/28 2018.

[28] C. Oates, A. Naylor, T. Hartshorne, S. Charles, T. Fail, K. Humphries, M. Aslam, and P. Khodabakhsh, “Joint recommendations for reporting carotid ultrasound investigations in the united kingdom,” *European Journal of Vascular and*

Endovascular Surgery, vol. 37, no. 3, pp. 251–261, 2009. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1078588408005522

[29] J. Walker and A. R. Naylor, "Ultrasound based measurement of ‘carotid stenosis >70%’: An audit of uk practice," *European Journal of Vascular and Endovascular Surgery*, vol. 31, no. 5, pp. 487–490, 2021/03/28 2006. [Online]. Available: https://doi.org/10.1016/j.ejvs.2005.11.029

[30] S. S. Y. Ho, "Current status of carotid ultrasound in atherosclerosis," *Quantitative imaging in medicine and surgery*, vol. 6, no. 3, pp. 285–296, 06 2016. [Online]. Available: https://www.ncbi.nlm.nih.gov/pubmed/27429912

[31] T. M. C., G. Amanda, K. C. E., G. A. D., K. J. D., L. K. J., A. B. C., S. Lianne, K. R. A., and S. J. H., "Predictors of carotid thickness and plaque progression during a decade," *Stroke*, vol. 45, no. 11, pp. 3257–3262, 2014.

[32] L. E. Chambless, G. Heiss, A. R. Folsom, W. Rosamond, M. Szklo, A. R. Sharrett, and L. X. Clegg, "Association of coronary heart disease incidence with carotid arterial wall thickness and major risk factors: the atherosclerosis risk in communities (aric) study, 1987-1993." *Am J Epidemiol*, vol. 146, no. 6, pp. 483–494, Sep 1997.

[33] D. H. O'Leary, J. F. Polak, R. A. Kronmal, T. A. Manolio, G. L. Burke, and S. K. Wolfson, "Carotid-artery intima and media thickness as a risk factor for myocardial infarction and stroke in older adults," *New England Journal of Medicine*, vol. 340, no. 1, pp. 14–22, 1999, pMID: 9878640. [Online]. Available: https://doi.org/10.1056/NEJM199901073400103

[34] J. R. r. Crouse, J. S. Raichlen, W. A. Riley, G. W. Evans, M. K. Palmer, D. H. O'Leary, D. E. Grobbee, and M. L. Bots, "E↵ect of rosuvastatin on progression of carotid intima-media thickness in low-risk individuals with subclinical atherosclerosis: the meteor trial." *JAMA*, vol. 297, no. 12, pp. 1344–1353, Mar 2007.

[35] J. F. Polak, M. J. Pencina, K. M. Pencina, C. J. O'Donnell, P. A. Wolf, and R. B. D'Agostino, "Carotid-wall intima–media thickness and cardiovascular events," *New England Journal of Medicine*, vol. 365, no. 3, pp. 213–221, 2011, pMID: 21774709.

[36] C. J. J., A. A. M., M. T. A., P. J. F., P. B. M., H. C. H., K. L. H., and C. Mary, "Association of carotid artery intima-media thickness, plaques, and c-reactive protein with future cardiovascular disease and all-cause mortality," *Circulation*, vol. 116, no. 1, pp. 32–38, 2021/04/02 2007.

[37] M. Plichart, D. S. Celermajer, M. Zureik, C. Helmer, X. Jouven, K. Ritchie, C. Tzourio, P. Ducimeti`ere, and J.-P. Empana, "Carotid intima-media thickness in plaque-free site, carotid plaques and coronary heart disease risk prediction in older adults. the three-city study," *Atherosclerosis*, vol. 219, no. 2, pp. 917–924, 2021/04/02 2011. [Online]. Available: https://doi.org/10.1016/j.atherosclerosis.2011.09.024

[38] J. Yeboah, R. McClelland, T. Polonsky, G. Burke, C. Sibley, D. O'Leary, J. Carr, D. Go↵, P. Greenland, and D. Herrington, "Comparison of novel risk markers for improvement in cardiovascular risk assessment in intermediate-risk individuals," *JAMA : the journal of the American Medical Association*, vol. 308, pp. 788–95, 08 2012.

[39] A. F. Members:, J. Perk, G. De Backer, H. Gohlke, I. Graham, Z. Reiner,ˇ M. Verschuren, C. Albus, P. Benlian, G. Boysen, R. Cifkova, C. Deaton,S. Ebrahim, M. Fisher, G. Germano, R. Hobbs, A. Hoes, S. Karadeniz, A. Mezzani, E. Prescott, L. Ryden, M. Scherer, M. Syva¨nne, W. J. Scholte Op Reimer, C. Vrints, D. Wood, J. L. Zamorano, F. Zannad, O. experts who contributed to parts of the guidelines:, M. T. Cooney, E. C. for Practice Guidelines (CPG):, J. Bax, H. Baumgartner, C. Ceconi, V. Dean, C. Deaton, R. Fagard, C. Funck-Brentano, D. Hasdai, A. Hoes, P. Kirchhof, J. Knuuti, P. Kolh, T. McDonagh, C. Moulin, B. A. Popescu, Z. Reiner, U. Sechtem, P. A. Sirnes, M. Tendera, A. Torbicki, A. Vahanian,ˇ S. Windecker, D. Reviewers:, C. Funck-Brentano, P. A. Sirnes, V. Aboyans, E. A. Ezquerra, C. Baigent, C. Brotons, G. Burell, A. Ceriello, J. De Sutter, J. Deckers, S. Del Prato, H.-C. Diener, D. Fitzsimons, Z. Fras, R. Hambrecht, P. Jankowski, U. Keil, M. Kirby, M. L. Larsen, G. Mancia, A. J. Manolis, J. McMurray, A. Pajak, A. Parkhomenko, L. Rallidis, F. Rigo, E. Rocha, L. M. Ruilope, E. van der Velde, D. Vanuzzo, M. Viigimaa, M. Volpe, O. Wiklund, and C. Wolpert, "European Guidelines on cardiovascular disease prevention in clinical practice (version 2012): The Fifth Joint Task Force of the European Society of Cardiology and Other Societies on Cardiovascular Disease Prevention in Clinical Practice (constituted by representatives of nine societies and by invited experts)Developed with the special contribution of the European Association for Cardiovascular Prevention amp; Rehabilitation (EACPR)," *European Heart Journal*, vol. 33, no. 13, pp. 1635–1701, 05 2012. [Online]. Available: https://doi.org/10.1093/eurheartj/ehs092

[40] M. F. Piepoli, A. W. Hoes, S. Agewall, C. Albus, C. Brotons, A. L. Catapano, M.-T. Cooney, U. Corra`, B. Cosyns, C. Deaton, I. Graham,
M. S. Hall, F. D. R. Hobbs, M.-L. Løchen, H. L¨ollgen, P. Marques-Vidal,
J. Perk, E. Prescott, J. Redon, D. J. Richter, N. Sattar, Y. Smulders, M. Tiberi, H. B. van der Worp, I. van Dis, W. M. M. Verschuren, S. Binno, and E. S. D. Group, "2016 european guidelines on cardiovascular disease prevention in clinical practice: The sixth joint task force of the european society of cardiology and other societies on cardiovascular disease prevention in clinical practice (constituted by representatives of 10 societies and by invited experts)developed with the special contribution of the european association for cardiovascular prevention amp; rehabilitation eacpr," *European Heart Journal*, vol. 37, no. 29, pp. 2315–2381, 05 2016. [Online]. Available: https://doi.org/10.1093/eurheartj/ehw106

[41] G. D. C., L.-J. D. M., B. Glen, C. Sean, D. R. B., G. Raymond,
G. Philip, L. D. T., L. Daniel, O. C. J., R. J. G., S. J. Sanford, S. S. T., S. S. C., S. Paul, S. N. J., and W. P. W. F., "2013 acc/aha guideline on the assessment of cardiovascular risk," *Circulation*, vol. 129, no. 25 suppl 2, pp. S49–S73, 2021/04/02 2014. [Online]. Available: https://doi.org/10.1161/01.cir.0000437741.48606.98

[42] J. F. Polak, "Sancta simplicitas!" *Journal of the American Society of Echocardiography*, vol. 25, no. 10, pp. 1137–1139, 2021/04/02 2012. [Online]. Available: https://doi.org/10.1016/j.echo.2012.08.004

[43] V. Nambi, L. Chambless, A. R. Folsom, M. He, Y. Hu, T. Mosley, K. Volcik, E. Boerwinkle, and C. M. Ballantyne, "Carotid intimamedia thickness and

presence or absence of plaque improves prediction of coronary heart disease risk: The aric (atherosclerosis risk

in communities) study," *Journal of the American College of Cardiology*, vol. 55, no. 15, pp. 1600–1607, 2010. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0735109710005103

[44] Y. Inaba, J. A. Chen, and S. R. Bergmann, "Carotid plaque, compared with carotid intima-media thickness, more accurately predicts coronary artery disease events: A meta-analysis," *Atherosclerosis*, vol. 220, no. 1, pp. 128–133, 2021/04/02 2012. [Online]. Available: https://doi.org/10.1016/j.atherosclerosis.2011.06.044

[45] T. Wannarong, G. Parraga, and D. Buchanan, "Progression of carotid plaque volume predicts cardiovascular events," *Journal of Vascular Surgery*, vol. 58, no. 4, p. 1142, 2019/06/25 2013. [Online]. Available: https://doi.org/10.1016/j.jvs.2013.08.015

[46] J. D. Spence, M. Eliasziw, M. DiCicco, D. G. Hackam, R. Galil, and T. Lohmann, "Carotid plaque area: a tool for targeting and evaluating vascular preventive therapy." *Stroke*, vol. 33, no. 12, pp. 2916–2922, Dec 2002.

[47] J. Spence, "Carotid ultrasound phenotypes are biologically distinct." *Arteriosclerosis, thrombosis, and vascular biology*, vol. 35 9, pp. 1910–3, 2015.

[48] C. D. Ainsworth, C. C. Blake, A. Tamayo, V. Beletsky, A. Fenster, and J. D. Spence, "3d ultrasound measurement of change in carotid plaque volume: a tool for rapid evaluation of new therapies." *Stroke*, vol. 36, no. 9, pp. 1904– 1909, Sep 2005.

[49] A. Huibers, G. J. de Borst, R. Bulbulia, H. Pan, and A. Halliday, "Plaque echolucency and the risk of ischaemic stroke in patients with asymptomatic carotid stenosis within the first asymptomatic carotid surgery trial (acst-1)," *European Journal of Vascular and Endovascular Surgery*,

vol. 51, no. 5, pp. 616–621, 2021/04/02 2016. [Online]. Available: https://doi.org/10.1016/j.ejvs.2015.11.013

[50] S. Kozue, N. Kazuyuki, I.-U. Hatsue, W. Akihiro, K. Hideaki, and I. Koji, "Contrast-enhanced ultrasound for the evaluation of neovascularization in atherosclerotic carotid artery plaques," *Stroke*, vol. 45, no. 10, pp. 3073–3075, 2021/04/02 2014. [Online]. Available: https://doi.org/10.1161/STROKEAHA.114.006483

[51] M. Muraki, T. Mikami, T. Yoshimoto, S. Fujimoto, M. Kitaguchi, S. Kaga, T. Sugawara, K. Tokuda, S. Kaneko, and T. Kashiwaba, "Sonographic detection of abnormal plaque motion of the carotid artery: Its usefulness in diagnosing high-risk lesions ranging from plaque rupture to ulcer formation," *Ultrasound in Medicine and Biology*, vol. 42, no. 2, pp. 358–364, 2021/04/02

2016. [Online]. Available: https://doi.org/10.1016/j.ultrasmedbio.2015.09.007

[52] W. McDicken and T. Anderson, "Chapter 1 - basic physics of medical ultrasound," in *Clinical Ultrasound (Third Edition)*, third edition ed.,

P. L. Allan, G. M. Baxter, and M. J. Weston, Eds. Edinburgh: Churchill Livingstone, 2011, pp. 3 – 15. [Online]. Available: http://www.sciencedirect.com/science/article/pii/B9780702031311000018

[53] D. Hykes, W. Hedrick, and D. Starchman, *Ultrasound Physics and Instrumentation*. Mosby-Year Book, 1992. [Online]. Available: https://books.google.com.mt/books?id=K2VrAAAAMAAJ

[54] K. Martin, "Chapter 2 - basic equipment, components and image production," in *Clinical Ultrasound (Third Edition)*, third edition ed., P. L. Allan, G. M. Baxter, and M. J. Weston, Eds. Edinburgh: Churchill Livingstone, 2011, pp. 16 – 30. [Online]. Available: http://www.sciencedirect.com/science/article/pii/B978070203131100002X

[55] J. D. Kirsch, M. Mathur, M. H. Johnson, G. Gowthaman, and L. M. Scoutt, "Advances in transcranial doppler us: Imaging ahead," *RadioGraphics*, vol. 33, no. 1, pp. E1–E14, 2021/04/10 2013. [Online]. Available: https://doi.org/10.1148/rg.331125071

[56] K. N. Thomas, N. C. S. Lewis, B. G. Hill, and P. N. Ainslie, "Technical recommendations for the use of carotid duplex ultrasound for the assessment of extracranial blood flow," *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*, vol. 309, no. 7, pp. R707–R720, 2015, pMID: 26157060. [Online]. Available: https://doi.org/10.1152/ajpregu.00211.2015

[57] B. Sigel, "A Brief History of Doppler Ultrasound in the Diagnosis of Peripheral Vascular Disease," *Ultrasound Med. Biol.*, vol. 24, no. 2, pp. 169–176, feb 1998. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S0301562997002640

[58] Z. Guo and A. Fenster, “Three-dimensional power Doppler imaging: a phantom study to quantify vessel stenosis,” *Ultrasound Med. Biol.*, vol. 21, no. X, 1996. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0301562996001251

[59] L. Christiaens, D. Coisne, J. Allal, P. Blouin, F. Dubreuil, E. Donal, and R. Barraine, “Three-dimensional power Doppler imaging: volume reconstruction of pulmonary artery flow with an in vitro pulsatile flow system,” *Eur. J. Ultrasound*, vol. 6, no. 2, pp. 135–139, oct 1997. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S0929826697100027

[60] S. Peng, J. Lin, W. Lee, and P. Tsao, “3-D power Doppler cerebral angiography in neonates and young infants: comparison with 2-D power Doppler angiography,” *Ultrasound Med. ...*, vol. 25, no. 6, pp. 947–951, 1999. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0301562999000472

[61] D. C. Barratt, B. B. Ari↵, K. N. Humphries, S. a. M. Thom, and A. D. Hughes, “Reconstruction and quantification of the carotid artery bifurcation from 3-D ultrasound images.” *IEEE Trans. Med. Imaging*, vol. 23, no. 5, pp. 567–83, may 2004. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/15147010

[62] G. Cloutier, Z. Qin, D. Garcia, and G. Soulez, “Assessment of arterial stenosis in a flow model with power Doppler angiography: accuracy and observations on blood echogenicity,” *Ultrasound Med. ...*, vol. 26, no. 9, pp. 1489–1501, 2000. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0301562900003008

[63] M. M. Hossain, K. AlMuhanna, L. Zhao, B. K. Lal, and S. Sikdar, “Semiautomatic segmentation of atherosclerotic carotid artery wall volume using 3D ultrasound imaging.” *Med. Phys.*, vol. 42, no. 4, p. 2029, 2015.

[64] L. Saba, C. Yuan, T. Hatsukami, N. Balu, Y. Qiao, J. DeMarco, T. Saam, A. Moody, D. Li, C. Matouk, M. Johnson, H. J¨ager, M. Mossa-Basha, M. Kooi, Z. Fan, D. Saloner, M. Wintermark, D. Mikulis, and B. Wasserman, “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,” *American Journal of Neuroradiology*, vol. 39, no. 2, pp. E9–E31, 2018. [Online]. Available: http://www.ajnr.org/content/39/2/E9

[65] a. Fenster, D. B. Downey, and H. N. Cardinal, “Three-dimensional ultrasound imaging.” *Phys. Med. Biol.*, vol. 46, no. 5, 2001.

[66] T. R. Nelson and D. H. Pretorius, “Three-dimensional ultrasound imaging.” *Ultrasound Med. Biol.*, vol. 24, no. 9, pp. 1243–70, nov 1998. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/10385948

[67] M. I. Daoud, A. Alshalalfah, F. Awwad, and M. Al-Najar, “Freehand 3d ultrasound imaging system using electromagnetic tracking,” in *2015 International Conference on Open Source Software Computing (OSSCOM)*, Sep. 2015, pp.
1–5.

[68] M. Egger, J. D. Spence, A. Fenster, and G. Parraga, “Validation of 3d ultrasound vessel wall volume: an imaging phenotype of carotid atherosclerosis.” *Ultrasound Med Biol*, vol. 33, no. 6, pp. 905–914, Jun 2007.

[69] B. Chiu, E. Ukwatta, S. Shavakh, and A. Fenster, "Quantification and visualization of carotid segmentation accuracy and precision using a 2d standardized carotid map," *Physics in medicine and biology*, vol. 58, pp. 3671–3703, 05 2013.

[70] D. Buchanan, T. Lindenmaier, S. McKay, Y. Bureau, D. Hackam, A. Fenster, and G. Parraga, "The relationship of carotid three-dimensional ultrasound vessel wall volume with age and sex: Comparison to carotid intima-media thickness," *Ultrasound in medicine biology*, vol. 38, pp. 1145–53, 05 2012.

[71] T. Lindenmaier, D. Buchanan, D. Pike, T. Hartley, R. Reid, J. D. Spence, R. Chan, M. Sharma, P. Prior, N. Suskin, and G. Parraga, "One, two and three-dimensional ultrasound measurements of carotid atherosclerosis before and after cardiac rehabilitation: Preliminary results of a randomized controlled trial," *Cardiovascular ultrasound*, vol. 11, p. 39, 11 2013.

[72] X. Yang, J. Jin, W. He, M. Yuchi, and M. Ding, "Segmentation of the common carotid artery with active shape models from 3D ultrasound images," B. van Ginneken and C. L. Novak, Eds., feb 2012, p. 83152H.

[73] E. Ukwatta, J. Yuan, D. Buchanan, B. Chiu, J. Awad, W. Qiu, G. Parraga, and a. Fenster, "Three-dimensional segmentation of three-dimensional ultrasound carotid atherosclerosis using sparse field level sets." *Med. Phys.*, vol. 40, no. May, p. 052903, 2013.

[74] K. I. Paraskevas, H. H. Sillesen, T. Rundek, E. B. Mathiesen, and J. D. Spence, "Carotid intima–media thickness versus carotid plaque burden for predicting cardiovascular risk," *Angiology*, vol. 71, no. 2, pp. 108–111, 2020, pMID: 31569951. [Online]. Available: https://doi.org/10.1177/0003319719878582

[75] H. Sillesen, S. Sartori, B. Sandholt, U. Baber, R. Mehran, and V. Fuster, "Carotid plaque thickness and carotid plaque burden predict future cardiovascular events in asymptomatic adult Americans," *European Heart Journal - Cardiovascular Imaging*, vol. 19, no. 9, pp. 1042–1050, 10 2017. [Online]. Available: https://doi.org/10.1093/ehjci/jex239

[76] M. E. B., J. S. Harald, W. Tom, B. K. H., L. Maja-Lisa, and N. Inger, "Carotid plaque area and intima-media thickness in prediction of first-ever ischemic stroke," *Stroke*, vol. 42, no. 4, pp. 972–978, 2021/04/17 2011. [Online]. Available: https://doi.org/10.1161/STROKEAHA.110.589754

[77] J. D. Spence, "Uses of ultrasound in stroke prevention," *Cardiovascular diagnosis and therapy*, vol. 10, no. 4, pp. 955–964, 08 2020. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/32968653

[78] NHS. Atherosclerosis. [Online]. Available: https://www.nhs.uk/conditions/atherosclerosis/

[79] NICE. Cardiovascular disease: risk assessment and reduction, including lipid modification. [Online]. Available: https://www.nice.org.uk/guidance/cg181/chapter/1-recommendations

[80] A. M. Johri, V. Nambi, T. Z. Naqvi, S. B. Feinstein, E. S. H. Kim, M. M. Park, H. Becher, and H. Sillesen, "Recommendations for the assessment of carotid arterial plaque by ultrasound for the characterization of atherosclerosis and evaluation of cardiovascular risk: From the american society of echocardiography," *Journal of the American Society of Echocardiography*, vol. 33, no. 8, pp. 917–933, 2021/04/18 2020. [Online]. Available: https://doi.org/10.1016/j.echo.2020.04.021

[81] A. D. K., B. R. S., A. M. A., B. A. B., G. Z. D., H. E. J., H. C. Dennison, K. Amit, L.-J. Donald, M. J. William, M. E. D., M. M. D., M. Daniel, S. S. C., V. S. S., W. K. A., Y. Joseph, and Z. Boback, "2019 acc/aha guideline on the primary prevention of cardiovascular disease: A report of the american college of cardiology/american heart association task force on clinical practice guidelines," *Circulation*, vol. 140, no. 11, pp. e596–e646, 2021/04/18 2019. [Online]. Available: https://doi.org/10.1161/CIR.0000000000000678

[82] J. D. Spence, "Coronary calcium is not all we need: Carotid plaque burden measured by ultrasound is better," *Atherosclerosis*, vol. 287, pp. 179–180, 2021/04/18 2019. [Online]. Available: https://doi.org/10.1016/j.atherosclerosis.2019.04.214

[83] P. K. Shah, "Screening asymptomatic subjects for subclinical atherosclerosis: Can we, does it matter, and should we?" *Journal of the American College of Cardiology*, vol. 56, no. 2, pp. 98–105, 2010. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0735109710016931

[84] E. Ukwatta, J. Awad, a. D. Ward, D. Buchanan, J. Samarabandu, G. Parraga, and a. Fenster, "Three-dimensional ultrasound of carotid atherosclerosis: semiautomated segmentation using a level set-based method." *Med. Phys.*, vol. 38, no. 5, pp. 2479–2493, 2011.

[85] J. Cheng, H. Li, F. Xiao, A. Fenster, X. Zhang, X. He, L. Li, and M. Ding, "Fully Automatic Plaque Segmentation in 3-D Carotid Ultrasound Images," *Ultrasound Med. Biol.*, vol. 39, no. 12, pp. 2431–2446, 2013.

[86] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. van der Laak, B. van Ginneken, and C. I. S´anchez, "A survey on deep learning in medical image analysis," *Medical Image Analysis*, vol. 42, pp. 60 – 88, 2017.

[87] A. F. M. D. Ran Zhou, Wei Ma, "U-net based automatic carotid plaque segmentation from 3d ultrasound images," 2019. [Online]. Available: https://doi.org/10.1117/12.2511932

[88] R. Zhou, A. Fenster, Y. Xia, J. D. Spence, and M. Ding,
"Deep learning-based carotid media-adventitia and lumen-intima boundary segmentation from three-dimensional ultrasound images," *Medical Physics*, vol. 46, no. 7, pp. 3180–3193, 2019. [Online]. Available: https://aapm.onlinelibrary.wiley.com/doi/abs/10.1002/mp.13581

[89] R. Zhou, F. Guo, M. R. Azarpazhooh, J. D. Spence, E. Ukwatta, M. Ding, and A. Fenster, "A voxel-based fully convolution network and continuous max-flow for carotid vessel-wall-volume segmentation from 3d ultrasound images," *IEEE Transactions on Medical Imaging*, vol. 39, no. 9, pp. 2844–2855, Sep. 2020.

[90] K. P. C. C. Azzopardi and Y. A. Hicks, "Carotid ultrasound segmentation using radio-frequency derived phase information and gabor filters," in *37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. EMBC, 2015, pp. 6338–6341.

[91] C. Azzopardi, Y. A. Hicks, and K. P. Camilleri, "Automatic carotid ultrasound segmentation using deep convolutional neural networks and phase congruency maps," in *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*, April 2017, pp. 624–628.

[92] C. Azzopardi, K. P. Camilleri, and Y. A. Hicks, “Bimodal automated carotid ultrasound segmentation using geometrically constrained deep neural networks,” *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 4, pp. 1004–1015, April 2020.

[93] A. S. Dar and D. Padha, “Medical image segmentation a review of recent techniques, advancements and a comprehensive comparison,” *International Journal of Computer Sciences and Engineering*, vol. 7, pp. 114–124, 07 2019.

[94] L. K. Lee, S. C. Liew, and W. J. Thong, “A review of image segmentation methodologies in medical image,” in *Advanced Computer and Communication Engineering Technology*, H. A. Sulaiman, M. A. Othman, M. F. I. Othman, Y. A. Rahim, and N. C. Pee, Eds. Cham: Springer International Publishing, 2015, pp. 1069–1080.

[95] J. Beveridge, J. Gri th, R. Kohler, A. Hanson, and E. Riseman, “Segmenting images using localized histograms and region merging,” *International Journal of Computer Vision*, vol. 2, pp. 311–347, 01 1989.

[96] Y. Feng, H. Zhao, X. Li, X. Zhang, and H. Li, “A multi-scale 3d otsu thresholding algorithm for medical image segmentation,” *Digital Signal Processing*, vol. 60, pp. 186 – 199, 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1051200416301191

[97] C. Zhou, L. Tian, H. Zhao, and K. Zhao, “A method of two-dimensional otsu image threshold segmentation based on improved firefly algorithm,” in *2015 IEEE International Conference on Cyber Technology in Automation, Control, and Intelligent Systems (CYBER)*, 2015, pp. 1420–1424.

[98] N. Greggio, A. Bernardino, C. Laschi, P. Dario, and J. Santos-Victor, "Fast estimation of gaussian mixture models for image segmentation," *Machine Vision and Applications*, vol. 23, no. 4, pp. 773–789, 2012. [Online]. Available: https://doi.org/10.1007/s00138-011-0320-5

[99] A. M. Taori, A. K. Chaudhari, S. S. Patankar, and J. V. Kulkami, "Segmentation of macula in retinal images using automated seeding region growing technique," in *2016 International Conference on Inventive Computation Technologies (ICICT)*, vol. 2, 2016, pp. 1–5.

[100] A. Javadpour and A. Mohammadi, "Improving brain magnetic resonance image (mri) segmentation via a novel algorithm based on genetic and regional growth," *Journal of Biomedical Physics and Engineering*, vol. 6, no. 2, pp. –, 2016. [Online]. Available: https://jbpe.sums.ac.ir/article$_4$3221.*html*

[101] E. Hancer and D. Karaboga, "A comprehensive survey of traditional, merge-split and evolutionary approaches proposed for determination of cluster number," *Swarm and Evolutionary Computation*, vol. 32, pp. 49 – 67, 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S2210650216300475

[102] G. N. Chaple, R. D. Daruwala, and M. S. Gofane, "Comparisions of robert, prewitt, sobel operator based edge detection methods for real time uses on fpga," in *2015 International Conference on Technologies for Sustainable Development (ICTSD)*, 2015, pp. 1–4.

[103] A. Kalra and R. L. Chhokar, "A hybrid approach using sobel and canny operator for digital image edge detection," in *2016 International Conference on Micro-Electronics and Telecommunication Engineering (ICMETE)*, 2016, pp.

305–310.

[104] M. Nikolic, E. Tuba, and M. Tuba, “Edge detection in medical ultrasound images using adjusted canny edge detection algorithm,” 11 2016, pp. 1–4.

[105] D. Marr and E. Hildreth, “Theory of edge detection,” *Proceedings of the Royal Society of London. Series B, Containing papers of a Biological character. Royal Society (Great Britain)*, vol. 207, pp. 187–217, 02 1980.

[106] S. Beucher and F. Meyer, *The Morphological Approach to Segmentation: The Watershed Transformation*, 01 1993, vol. Vol. 34, pp. 433–481.

[107] N. Dhanachandra, K. Manglem, and Y. J. Chanu, “Image segmentation using k - means clustering algorithm and subtractive clustering algorithm,” *Procedia Computer Science*, vol. 54, pp. 764 – 771, 2015, eleventh International Conference on Communication Networks, ICCN 2015, August 21-23, 2015, Bangalore, India Eleventh International Conference on Data Mining and Warehousing, ICDMW 2015, August 21-23, 2015, Bangalore, India Eleventh International Conference on Image and Signal Processing, ICISP 2015, August 21-23, 2015, Bangalore, India. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1877050915014143

[108] M. Sharma, G. N. Purohit, and S. Mukherjee, “Information retrieves from brain mri images for tumor detection using hybrid technique k-means and artificial neural network (kmann),” in *Networking Communication and Data Knowledge Engineering*, G. M. Perez, K. K. Mishra, S. Tiwari, and M. C. Trivedi, Eds. Singapore: Springer Singapore, 2018, pp. 145–157.

[109] A. Funmilola, “Fuzzy k-c-means clustering algorithm for medical image segmentation,” *Journal of Information Engineering and Applications*, vol. Vol 2, no. No .6, pp. 21–32, 2012.

[110] R. Meena Prakash and R. Shantha Selva Kumari, “Spatial fuzzy c means and expectation maximization algorithms with bias correction for segmentation of mr brain images,” *Journal of Medical Systems*, vol. 41, no. 1, p. 15, 2016. [Online]. Available: https://doi.org/10.1007/s10916-016-0662-7

[111] S. Osher and J. A. Sethian, “Fronts propagating with curvature-dependent speed: Algorithms based on hamilton-jacobi formulations,” *Journal of Computational Physics*, vol. 79, no. 1, pp. 12 – 49, 1988. [Online]. Available: http://www.sciencedirect.com/science/article/pii/0021999188900022

[112] R. Malladi, J. A. Sethian, and B. C. Vemuri, “Shape modeling with front propagation: a level set approach,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 17, no. 2, pp. 158–175, 1995.

[113] X. Wang, L. He, and W. Wee, “Deformable contour method: A constrained optimization approach,” *International Journal of Computer Vision*, vol. 59, no. 1, pp. 87–108, 2004. [Online]. Available: https://doi.org/10.1023/B:VISI.0000020672.14006.ad

[114] Z. Kato and J. Zerubia, *Markov Random Fields in Image Segmentation*. Hanover, MA, USA: Now Publishers Inc., 2012.

[115] W. Wang, J. Dong, and T. Tan, “Exploring dct coe cient quantization e↵ects for local tampering detection,” *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 10, pp. 1653–1666, 2014.

[116] Y. Boykov and G. Funka-Lea, “Graph cuts and e cient n-d image segmentation,” *International Journal of Computer Vision*, vol. 70, no. 2, pp. 109–131, 2006. [Online]. Available: https://doi.org/10.1007/s11263-006-7934-5

[117] E. Zanaty, “Medical image segmentation techniques: An overview,” *International Journal of informatics and medical data processing (JIMDP)*, vol. 1, p. 1, 01 2016.

[118] J. M. Lo¨tjo¨nen, R. Wolz, J. R. Koikkalainen, L. Thurfjell, G. Waldemar, H. Soininen, and D. Rueckert, “Fast and robust multi-atlas segmentation of brain magnetic resonance images,” *NeuroImage*, vol. 49, no. 3, pp. 2352–2365, 2010. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1053811909010970

[119] J. A. Noble and D. Boukerroui, “Ultrasound image segmentation: a survey.” *IEEE Trans. Med. Imaging*, vol. 25, no. 8, pp. 987–1010, aug 2006. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/16894993

[120] V. Shrimali, R. S. Anand, and V. Kumar, “Current trends in segmentation of medical ultrasound b-mode images: A review,” *IETE Technical Review*, vol. 26, no. 1, pp. 8–17, 2009. [Online]. Available: https://www.tandfonline.com/doi/abs/10.4103/0256-4602.48464

[121] F. Molinari, G. Zeng, and J. S. Suri, “A state of the art review on intimamedia thickness (IMT) measurement and wall segmentation techniques for carotid ultrasound.” *Comput. Methods Programs Biomed.*, vol. 100, no. 3, pp. 201–21, dec 2010.

[122] P. Pignoli and T. Longo, "Evaluation of atherosclerosis with B-mode ultrasound imaging," *J. Nucl. Med. Allied Sci. 32*, vol. 3, no. 3, pp. 166–173, 1988.

[123] P. J. Touboul, P. Prati, P. Y. Scarabin, V. Adrai, E. Thibout, and P. Ducimeti`ere, "Use of monitoring software to improve the measurement of carotid wall thickness by B-mode imaging." *J. Hypertens. Suppl.*, vol. 10, pp. S37–S41, 1992.

[124] C. Liguori, A. Paolillo, and A. Pietrosanto, "An automatic measurement system for the evaluation of carotid intima-media thickness," *IEEE Trans. Instrum. Meas.*, vol. 50, no. 6, pp. 1684–1691, 2001.

[125] J. H. Stein, C. E. Korcarz, M. E. Mays, P. S. Douglas, M. Palta, H. Zhang, T. LeCaire, D. Paine, D. Gustafson, and L. Fan, "A semiautomated ultrasound border detection program that facilitates clinical measurement of ultrasound carotid intima-media thickness," *J. Am. Soc. Echocardiogr.*, vol. 18, no. 3, pp. 244–251, 2005.

[126] F. Faita, V. Gemignani, E. Bianchini, C. Giannarelli, L. Ghiadoni, and M. Demi, "Real-time measurement system for evaluation of the carotid intimamedia thickness with a robust edge operator." *J. Ultrasound Med.*, vol. 27, no. 9, pp. 1353–1361, 2008.

[127] C. P. Loizou, C. S. Pattichis, M. Pantziaris, T. Tyllis, and a. Nicolaides, "Snakes based segmentation of the common carotid artery intima media." *Med. Biol. Eng. Comput.*, vol. 45, no. 1, pp. 35–49, 2007.

[128] F. Molinari, G. Zeng, and J. S. Suri, "An integrated approach to computerbased automated tracing and its validation for 200 common carotid arterial wall

ultrasound images: a new technique." *J. Ultrasound Med.*, vol. 29, no. 3, pp. 399–418, mar 2010.

[129] M. Gutierrez, P. Pilon, S. Lage, L. Kopel, R. Carvalho, and S. Furuie, "Automatic measurement of carotid diameter and wall thickness in ultrasound images," in *Comput. Cardiol.* IEEE, 2002, pp. 359–362.

[130] D. chuan Cheng, A. Schmidt-Trucksa¨ss, K. sheng Cheng, and H. Burkhardt, "Using snakes to detect the intimal and adventitial layers of the com-

mon carotid artery wall in sonographic images," *Comput. Methods Programs Biomed.*, vol. 67, no. 1, pp. 27–37, 2002.

[131] S. Delsanto, F. Molinari, P. Giustetto, W. Liboni, S. Badalamenti, and J. S. Suri, "Characterization of a completely user-independent algorithm for carotid artery segmentation in 2-D ultrasound images," *IEEE Trans. Instrum. Meas.*, vol. 56, no. 4, pp. 1265–1274, 2007.

[132] I. Wendelhag, Q. Liang, T. Gustavsson, and J. Wikstrand, "A new automated computerized analyzing system simplifies readings and reduces the variability in ultrasound measurement of intima-media thickness." *Stroke.*, vol. 28, no. 11, pp. 2195–200, nov 1997.

[133] Q. Liang, I. Wendelhag, J. Wikstrand, and T. Gustavsson, "A multiscale dynamic programming procedure for boundary detection in ultrasonic artery images." *IEEE Trans. Med. Imaging*, vol. 19, no. 2, pp. 127–142, 2000.

[134] F. Destrempes, J. Meunier, M. F. Giroux, G. Soulez, and G. Cloutier, "Segmentation in ultrasonic B-mode images of healthy carotid arteries using

mixtures of Nakagami distributions and stochastic optimization," *IEEE Trans. Med. Imaging*, vol. 28, no. 2, pp. 215–229, 2009.

[135] S. Golemati, J. Stoitsis, E. G. Sifakis, T. Balkizas, and K. S. Nikita, "Using the Hough Transform to Segment Ultrasound Images of Longitudinal and Transverse Sections of the Carotid Artery," *Ultrasound Med. Biol.*, vol. 33, no. 12, pp. 1918–1932, 2007.

[136] F. Destrempes, J. Meunier, M.-f. Giroux, G. Soulez, and G. Cloutier, "Segmentation of Plaques in Sequences of Ultrasonic B-Mode Images of Carotid Arteries Based on Motion Estimation and a Bayesian Model," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 8, pp. 2202–2211, 2011.

[137] J. C. R. Seabra, L. M. Pedro, J. F. E Fernandes, and J. M. Sanches, "A 3-D ultrasound-based framework to characterize the echo morphology of carotid plaques," *IEEE Trans. Biomed. Eng.*, vol. 56, no. 5, pp. 1442–1453, 2009.

[138] J. Guerrero, S. E. Salcudean, J. A. McEwen, B. A. Masri, and S. Nicolaou, "Real-time vessel segmentation and tracking for ultrasound imaging applications," *IEEE Trans Med Imaging*, vol. 26, no. 8, pp. 1079–1090, 2007.

[139] F. Mao, J. Gill, D. Downey, and A. Fenster, "Segmentation of carotid artery in ultrasound images: Method development and evaluation technique," *Med. Phys.*, vol. 27, no. 8, p. 1961, 2000.

[140] A. Zahalka and A. Fenster, "An automated segmentation method for threedimensional carotid ultrasound images," *Phys. Med. Biol.*, vol. 46, no. 4, pp. 1321–1342, 2001.

[141] P. Abolmaesumi, M. Sirouspour, and S. Salcudean, "Real-time extraction of carotid artery contours from ultrasound\nimages," *Proc. 13th IEEE Symp. Comput. Med. Syst. CBMS 2000*, pp. 2–7, 2000.

[142] J. D. Gill, H. M. Ladak, D. a. Steinman, and a. Fenster, "Accuracy and variability assessment of a semiautomatic technique for segmentation of the carotid arteries from three-dimensional ultrasound images." *Med. Phys.*, vol. 27, no. 6, pp. 1333–42, 2000.

[143] I. Solovey, "Segmentation of 3D Carotid Ultrasound Images Using Weak Geometric Priors," Master of Applied Science, University of Waterloo, 2010.

[144] M. M. Hossain, K. AlMuhanna, L. Zhao, B. K. Lal, and S. Sikdar, "Semiautomatic segmentation of atherosclerotic carotid artery lumen using 3d ultrasound imaging," *SPIE*, vol. 8869, p. 86694A, 2013.

[145] M. Egger, J. D. Spence, A. Fenster, and G. Parraga, "Validation of 3D Ultrasound Vessel Wall Volume: An Imaging Phenotype of Carotid Atherosclerosis," *Ultrasound Med. Biol.*, vol. 33, no. 6, pp. 905–914, 2007.

[146] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Semantic Image Segmentation with Deep Convolutional Nets and Fully Connected CRFs," *Iclr*, pp. 1–14, 2014.

[147] R. M. Menchon-Lara and J. L. Sancho-Gomez, "Fully automatic segmentation of ultrasound common carotid artery images based on machine learning," *Neurocomputing*, vol. 151, pp. 161–167, 2015.

[148] N. Tajbakhsh, J. Y. Shin, R. T. Hurst, C. B. Kendall, and J. Liang, "Chapter 5 - automatic interpretation of carotid intima–media thickness videos using

convolutional neural networks," in *Deep Learning for Medical Image Analysis*, S. K. Zhou, H. Greenspan, and D. Shen, Eds. Academic Press, 2017, pp. 105 – 131. [Online]. Available: http://www.sciencedirect.com/science/article/pii/B9780128104088000079

[149] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, "Backpropagation Applied to Handwritten Zip Code Recognition," pp. 541–551, 1989.

[150] Y. LeCun, L. Bottou, Y. Bengio, and P. Ha↵ner, "Gradient-Based Learning Applied to Document Recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278– 2324, 1998.

[151] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015. [Online]. Available: https://doi.org/10.1038/nature14539

[152] J. Gu, Z. Wang, J. Kuen, L. Ma, A. Shahroudy, B. Shuai, T. Liu, X. Wang, L. Wang, G. Wang, J. Cai, and T. Chen, "Recent advances in convolutional neural networks," arXiv, 1512.07108 2017.

[153] G. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," *Neural computation*, vol. 18, pp. 1527–54, 08 2006.

[154] G. Hinton and R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science (New York, N.Y.)*, vol. 313, pp. 504–7, 08 2006.

[155] Y. Bengio, P. Lamblin, D. Popovici, and H. Larochelle, "Greedy layer-wise training of deep networks," *Adv. Neural Inf. Process. Syst.*, vol. 19, pp. 153– 160, 01 2007.

[156] Y. Bengio, L. Yao, G. Alain, and P. Vincent, "Generalized denoising autoencoders as generative models," *Advances in Neural Information Processing Systems*, 05 2013.

[157] S. Rifai, P. Vincent, X. Muller, X. Glorot, and Y. Bengio, "Contractive autoencoders: Explicit invariance during feature extraction," 01 2011.

[158] G. Hinton, A. Krizhevsky, and S. Wang, "Transforming auto-encoders," vol. 6791, 06 2011, pp. 44–51.

[159] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, and P.-A. Manzagol, "Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion," *Journal of Machine Learning Research*, vol. 11, pp. 3371–3408, 12 2010.

[160] G. Hinton, "A practical guide to training restricted boltzmann machines[j]," *Momentum*, vol. 9, pp. 926–947, 01 2010.

[161] H. Larochelle, M. Mandel, R. Pascanu, and Y. Bengio, "Learning algorithms for the classification restricted boltzmann machine," *The Journal of Machine Learning Research*, vol. 13, pp. 643–669, 03 2012.

[162] A. Ng, "Convolutional neural networks (lenet)."

[163] K. Smagulova and A. James, *Overview of Long Short-Term Memory Neural Networks*, 01 2020, pp. 139–153.

[164] (2020, November) Image classification on imagenet. [Online]. Available: https://paperswithcode.com/sota/image-classification-on-imagenet

[165] A. Krizhevsky, I. Sutskever, and G. Hinton, "Imagenet classification with deep convolutional neural networks," *Neural Information Processing Systems*, vol. 25, 01 2012.

[166] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *Journal of Machine Learning Research*, vol. 15, pp. 1929–1958, 2014. [Online]. Available: http://jmlr.org/papers/v15/srivastava14a.html

[167] Y. Bengio, "Practical recommendations for gradient-based training of deep architectures," *Arxiv*, 06 2012.

[168] K. Simonyan and A. Zisserman, "Very deep convolutional networks for largescale image recognition," *arXiv 1409.1556*, 09 2014.

[169] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," *ArXiv e-prints*, 09 2014.

[170] M. Lin, Q. Chen, and S. Yan, "Network in network," *ArXiv e-prints*, vol. 132.4400, 2014.

[171] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," 06 2016, pp. 770–778.

[172] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An image is worth 16x16 words: Transformers for image recognition at scale," 10 2020.

[173] T. Zhou, S. Ruan, and S. Canu, "A review: Deep learning for medical image segmentation using multi-modality fusion," *Array*, vol. 3-4, p. 100004, 2019. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S2590005619300049

[174] S. Pereira, A. Pinto, V. Alves, and C. Silva, "Brain tumor segmentation using convolutional neural networks in mri images," *IEEE Transactions on Medical Imaging*, vol. 35, pp. 1–1, 03 2016.

[175] F. Isensee, P. Kickingereder, W. Wick, M. Bendszus, and K. Maier-Hein, "Brain tumor segmentation and radiomics survival prediction: Contribution to the brats 2017 challenge," 02 2018.

[176] ——, *No New-Net: 4th International Workshop, BrainLes 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16, 2018, Revised Selected Papers, Part II*. MICCAI, 01 2019, pp. 234–244.

[177] S. Cui, L. Mao, J. Jiang, C. Liu, and S. Xiong, "Automatic semantic segmentation of brain gliomas from mri images using a deep cascaded neural network," *Journal of Healthcare Engineering*, vol. 2018, pp. 1–14, 03 2018.

[178] G. Wang, W. Li, S. Ourselin, and T. Vercauteren, *Automatic Brain Tumor Segmentation Using Cascaded Anisotropic Convolutional Neural Networks*, 09 2018, pp. 178–190.

[179] K. Kamnitsas, C. Ledig, V. Newcombe, J. Simpson, A. Kane, D. Menon, D. Rueckert, and B. Glocker, "E cient multi-scale 3d cnn with fully connected crf for accurate brain lesion segmentation," 01 2017.

[180] Z. Xiaomei, Y. Wu, G. Song, l. Zhenye, Y. Zhang, and Y. Fan, "A deep learning model integrating fcnns and crfs for brain tumor segmentation," *Medical Image Analysis*, vol. 43, 02 2017.

[181] A. Myronenko, "3d mri brain tumor segmentation using autoencoder regularization," 10 2018.

[182] A. Cl`erigues, S. Valverde, J. Bernal, J. Freixenet, A. Oliver, and X. Llado, "Sunet: a deep learning architecture for acute stroke lesion segmentation and outcome prediction in multimodal mri," 10 2018.

[183] J. Dolz, K. Gopinath, J. Yuan, H. Lombaert, C. Desrosiers, and I. Ben Ayed, "Hyperdense-net: A hyper-densely connected cnn for multi-modal image segmentation," *IEEE Transactions on Medical Imaging*, vol. PP, 04 2018.

[184] J. Dolz, C. Desrosiers, and I. Ben Ayed, *IVD-Net: Intervertebral Disc Localization and Segmentation in MRI with a Multi-modal UNet*, 01 2019, pp. 130–143.

[185] L. Chen, Y. Wu, A. Dsouza, A. Abidin, A. Wismuller, and C. Xu, "Mri tumor segmentation with densely connected 3d cnn," 01 2018.

[186] L. Rokach, "Ensemble-based classifiers," *Artif. Intell. Rev.*, vol. 33, pp. 1–39, 02 2010.

[187] D. Nie, L. Wang, Y. Gao, and D. Sken, "Fully convolutional networks for multimodality isointense infant brain image segmentation," vol. 2016, 04 2016, pp.
1342–1345.

[188] P. Mlynarski, H. Delingette, A. Criminisi, and N. Ayache, "3d convolutional neural networks for tumor segmentation using long-range 2d context," *Computerized Medical Imaging and Graphics*, vol. 73, 02 2019.

[189] M. Aygu¨n, Y. S¸ahin, and G. Unal, "Multi modal convolutional neural networks¨ forbrain tumor segmentation," 09 2018.

[190] Y. Guo, Y. Liu, T. Georgiou, and M. Lew, "A review of semantic segmentation using deep neural networks," *International Journal of Multimedia Information Retrieval*, vol. 7, 06 2018.

[191] S. Geman and D. Geman, "Geman, d.: Stochastic relaxation, gibbs distribution, and the bayesian restoration of images. ieee trans. pattern anal. mach. intell. pami-6(6), 721-741," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 6, pp. 721–741, 11 1984.

[192] P. Arbelaez, M. Maire, C. Fowlkes, and J. Malik, "Contour detection and hierarchical image segmentation," *IEEE transactions on pattern analysis and machine intelligence*, vol. 33, pp. 898–916, 05 2011.

[193] P. Yu, K. Qin, and D. Clausi, "Unsupervised polarimetric sar image segmentation and classification using region growing with edge penalty," *IEEE Transactions on Geoscience and Remote Sensing - IEEE TRANS GEOSCI REMOT SEN*, vol. 50, pp. 1302–1317, 04 2012.

[194] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 11 2013.

[195] J. Carreira, R. Caseiro, J. Batista, and C. Sminchisescu, “Semantic segmentation with second-order pooling,” in *Computer Vision – ECCV 2012*. Springer Berlin Heidelberg, 10 2012, pp. 430–443.

[196] B. Hariharan, P. Arbela´ez, R. Girshick, and J. Malik, “Simultaneous detection and segmentation,” in *Computer Vision – ECCV 2014*, D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, Eds. Cham: Springer International Publishing, 2014, pp. 297–312.

[197] B. Hariharan, P. Arbelaez, R. Girshick, and J. Malik, “Hypercolumns for object segmentation and fine-grained localization,” arXiv, 11 2014.

[198] S. Liu, X. Qi, J. Shi, H. Zhang, and J. Jia, “Multi-scale patch aggregation (mpa) for simultaneous detection and segmentation,” 06 2016, pp. 3141–3149.

[199] S. Ren, K. He, R. Girshick, and J. Sun, “Faster r-cnn: Towards real-time object detection with region proposal networks,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137–1149, 2017.

[200] J. Long, E. Shelhamer, and T. Darrell, “Fully Convolutional Networks for Semantic Segmentation,” *2015 IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 3431–3440, 2014. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=7298965\nhttp://arxiv.org/

[201] D. Eigen and R. Fergus, “Predicting depth, surface normals and semantic labels with a common multi-scale convolutional architecture,” in *2015 IEEE International Conference on Computer Vision (ICCV)*, 2015, pp. 2650–2658.

[202] Y. Liu, Y. Guo, and M. Lew, “On the exploration of convolutional fusion networks for visual recognition,” 01 2017, pp. 277–289.

[203] V. Dumoulin and F. Visin, “A guide to convolution arithmetic for deep learning,” 2016, arXiv 1603.07285.

[204] C. Wang, X. Yan, M. Smith, K. Kochhar, M. Rubin, S. M. Warren, J. Wrobel, and H. Lee, “A unified framework for automatic wound segmentation and analysis with deep convolutional neural networks.” *Conf. Proc. ... Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. IEEE Eng. Med. Biol. Soc. Annu. Conf.*, vol. 2015, pp. 2415–2418, 2015.

[205] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” *CoRR*, vol. abs/1505.04597, 2015. [Online]. Available: http://arxiv.org/abs/1505.04597

[206] O. C¸i¸cek, A. Abdulkadir, S. Lienkamp, T. Brox, and O. Ronneberger, “3d¨ u-net: Learning dense volumetric segmentation from sparse annotation,” 06 2016.

[207] F. Milletari, N. Navab, and S.-A. Ahmadi, “V-net: Fully convolutional neural networks for volumetric medical image segmentation,” 10 2016, pp. 565–571.

[208] M. Drozdzal, E. Vorontsov, G. Chartrand, S. Kadoury, and C. Pal, “The importance of skip connections in biomedical image segmentation,” 08 2016.

[209] J. Dai, K. He, and J. Sun, “Boxsup: Exploiting bounding boxes to supervise convolutional networks for semantic segmentation,” in *2015 IEEE International Conference on Computer Vision (ICCV)*, 2015, pp. 1635–1643.

[210] G. Papandreou, L. Chen, K. P. Murphy, and A. L. Yuille, “Weakly-and semi-supervised learning of a deep convolutional network for semantic image segmentation,” in *2015 IEEE International Conference on Computer Vision (ICCV)*, 2015, pp. 1742–1750.

[211] A. Khoreva, R. Benenson, J. Hosang, M. Hein, and B. Schiele, "Simple does it: Weakly supervised instance and semantic segmentation," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 1665–1674.

[212] P. Pinheiro and R. Collobert, "From image-level to pixel-level labeling with convolutional networks," in *CVPR*, 06 2015, pp. 1713–1721.

[213] D. Pathak, E. Shelhamer, J. Long, and T. Darrell, "Fully convolutional multiclass multiple instance learning," 12 2014.

[214] D. Pathak, P. Kra¨henbu¨hl, and T. Darrell, "Constrained convolutional neural networks for weakly supervised segmentation," in *2015 IEEE International Conference on Computer Vision (ICCV)*, 2015, pp. 1796–1804.

[215] A. Bearman, O. Russakovsky, V. Ferrari, and L. Fei-Fei, "What's the point: Semantic segmentation with point supervision," in *Computer Vision – ECCV 2016*, B. Leibe, J. Matas, N. Sebe, and M. Welling, Eds. Cham: Springer International Publishing, 2016, pp. 549–565.

[216] Y. Wei, X. Liang, Y. Chen, Z. Jie, Y. Xiao, Y. Zhao, and S. Yan, "Learning to segment with image-level annotations," *Pattern Recognition*, vol. 59, pp. 234 – 244, 2016, compositional Models and Structured Learning for Visual Recognition. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0031320316000364

[217] F. Saleh, M. S. Aliakbarian, M. Salzmann, L. Petersson, S. Gould, and J. M. Alvarez, "Built-in foreground/background prior for weakly-supervised semantic

segmentation," in *Computer Vision – ECCV 2016*, B. Leibe, J. Matas, N. Sebe, and M. Welling, Eds. Cham: Springer International Publishing, 2016, pp. 413–432.

[218] W. Shimoda and K. Yanai, "Distinct class-specific saliency maps for weakly supervised semantic segmentation," in *Computer Vision – ECCV 2016*, B. Leibe, J. Matas, N. Sebe, and M. Welling, Eds. Cham: Springer International Publishing, 2016, pp. 218–234.

[219] B. Jin, M. V. Ortiz Segovia, and S. Su¨sstrunk, "Webly supervised semantic segmentation," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 1705–1714.

[220] K. M. Meiburger, U. R. Acharya, and F. Molinari, "Automated localization and segmentation techniques for b-mode ultrasound images: A review," *Computers in Biology and Medicine*, vol. 92, pp. 210 – 235, 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0010482517303888

[221] D. Boukerroui, O. Basset, a. Baskurt, and G. Gimenez, "A multiparametric and multiresolution segmentation algorithm of 3-D ultrasonic data." *IEEE Trans. Ultrason. Ferroelectr. Freq. Control*, vol. 48, no. 1, pp. 64–77, 2001.

[222] O. Bernard, J. D'hooge, and D. Friboulet, "Segmentation of echocardiographic images based on statistical modelling of the radio-frequency signal," *Eur. Signal Process. Conf.*, no. Eusipco, 2006.

[223] I. Dydenko, D. Friboulet, J. M. Gorce, J. D'hooge, B. Bijnens, and I. E. Magnin, "Towards ultrasound cardiac image segmentation based on the radiofrequency signal," *Med. Image Anal.*, vol. 7, no. 3, pp. 353–367, 2003.

[224] M. M. Nillesen, R. G. P. Lopata, H. J. Huisman, J. M. Thijssen, L. Kapusta, and C. L. De Korte, "Correlation Based 3-D Segmentation of the Left Ventricle in Pediatric Echocardiographic Images Using Radio-Frequency Data," *Ultrasound Med. Biol.*, vol. 37, no. 9, pp. 1409–1420, 2011.

[225] M. Mulet-Parada and J. a. Noble, "2D+T acoustic boundary detection in echocardiography." *Med. Image Anal.*, vol. 4, no. 1, pp. 21–30, 2000.

[226] A. Belaid, D. Boukerroui, Y. Maingourd, and J. F. Lerallut, "Implicit active contours for ultrasound images segmentation driven by phase information and local maximum likelihood," *Proc. - Int. Symp. Biomed. Imaging*, pp. 630–635, 2011.

[227] I. Despotovic, "Using phase information in ultrasound rf-signals for tissue characterization," *Annu. Work.*, pp. 314–317, 2008.

[228] P. Kovesi, "Image Features from Phase Congruency," *Videre*, vol. 1, no. 3, 1999.

[229] (2020, December). [Online]. Available: https://commons.wikimedia.org/wiki/File:Carotid*ultrasound.jpg*

[230] D. Sheet, S. Pal, and A. Chakraborty, "Image quality assessment for performance evaluation of despeckle filters in Optical Coherence Tomography of human skin," *Iecbes, 2010 Ieee*, no. December, pp. 499–504, 2010.

[231] N. Feng, J. Zhang, and W. Wang, "A quadrature demodulation method based on tracking the ultrasound echo frequency," *Ultrasonics*, vol. 44, pp. e47 – e50, 2006, proceedings of Ultrasonics International (UI'05) and World Congress on Ultrasonics (WCU). [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0041624X06002988

[232] J. Malik and P. Perona, “Preattentive texture discrimination with early vision mechanisms,” *Journal of the Optical Society of America A: Optics and Image Science, and Vision*, vol. 7, no. 5, pp. 923–932, 1990.

[233] A. K. Jain and F. Farrokhnia, “Unsupervised texture segmentation using gabor filters,” in *IEEE International Conference on Systems, Man, and Cybernetics Conference Proceedings*, Nov 1990, pp. 14–19.

[234] (2006, November). [Online]. Available: http://note.sonots.com/Matlab/cvprtoolbox.html

[235] Jainguo Zhang, Tieniu Tan, and Li Ma, “Invariant texture segmentation via circular gabor filters,” in *Object recognition supported by user interaction for service robots*, vol. 2, 2002, pp. 901–904 vol.2.

[236] D. J. Williams and M. Shah, “A fast algorithm for active contours and curvature estimation,” *CVGIP: Image Understanding*, vol. 55, no. 1, pp. 14 – 26, 1992. [Online]. Available: http://www.sciencedirect.com/science/article/pii/104996609290003L

[237] J. Stoitsis, S. Golemati, S. Kendros, and K. S. Nikita, “Automated detection of the carotid artery wall in B-mode ultrasound images using active contours initialized by the Hough Transform.” *Conf. Proc. IEEE Eng. Med. Biol. Soc.*, vol. 2008, pp. 3146–3149, 2008.

[238] D.-J. Kroon. 2d line curvature and normals. MATLAB Central File Exchange. [Online]. Available: (https://www.mathworks.com/matlabcentral/fileexchange/32696-2dline-curvature-and-normals)

[239] M. Morrone and R. Owens, “Feature detection from local energy,” *Pattern Recognit. Lett.*, vol. 6, no. 5, pp. 303–313, dec 1987.

[240] E. W. Weisstein. (2020, June) ”fourier series–square wave.” from mathworld-a wolfram web resource. [Online]. Available: https://mathworld.wolfram.com/FourierSeriesSquareWave.html

[241] P. Kovesi, “Phase congruency detects corners and edges,” in *Digital Image Computing: Techniques and Applications: Proceedings of the VIIth Biennial Australian Pattern Recognition Society Conference, DICTA 2003*, C. Sun, H. Talbot, S. Ourselin, and T. Adriaansen, Eds., vol. Vol 1., 2003.

[242] ——. Matlab and octave functions for computer vision and image processing. [Online]. Available: https://www.peterkovesi.com/matlabfns/

[243] H. Boche and V. Pohl, “Peak value blowup of approximations of the hilbert transform of signals with finite energy,” in *2017 American Control Conference (ACC)*, 2017, pp. 5738–5743.

[244] R. Parikh, A. Mathai, S. Parikh, G. Sekhar, and R. Thomas, “Understanding and using sensitivity, specificity and predictive values,” *Indian journal of ophthalmology*, vol. 56, pp. 45–50, 03 2008.

[245] A. Vedaldi and K. Lenc, “Matconvnet – convolutional neural networks for matlab,” in *Proceeding of the ACM Int. Conf. on Multimedia*, 2015.

[246] M. Dubuisson and A. K. Jain, “A modified hausdor↵ distance for object matching,” in *Proceedings of 12th International Conference on Pattern Recognition*, vol. 1, Oct 1994, pp. 566–568.

[247] R. K. Srivastava, K. Gre↵, and J. Schmidhuber, "Training Very Deep Networks," in *Adv. Neural Inf. Process. Syst.*, jul 2015, pp. 2377–2385. [Online]. Available: http://arxiv.org/abs/1507.06228

[248] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov, "Improving neural networks by preventing co-adaptation of feature detectors," arXiv, 2012.

[249] H. Kervadec, J. Bouchtiba, C. Desrosiers, E. Granger, J. Dolz, and I. Ben´ Ayed, "Boundary loss for highly unbalanced segmentation," *arXiv e-prints*, p. arXiv:1812.07032, Dec 2018.

[250] E. Abbena, S. Salamon, and A. Gray, *Modern Di□erential Geometry of Curves and Surfaces with Mathematica, Third Edition*, ser. Textbooks in Mathematics. Taylor & Francis, 2006.

[251] L. Prechelt, *Early Stopping — But When?* Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 53–67.

[252] O. Oktay, E. Ferrante, K. Kamnitsas, M. P. Heinrich, W. Bai, J. Caballero, R. Guerrero, S. A. Cook, A. de Marvao, T. Dawes, D. P. O'Regan, B. Kainz, B. Glocker, and D. Rueckert, "Anatomically constrained neural networks (ACNN): application to cardiac image enhancement and segmentation," *CoRR*, vol. abs/1705.08302, 2017. [Online]. Available: http://arxiv.org/abs/1705.08302

[253] J. D. S. Anthony Landry and A. Fenster, "Research article pdf/epub tools share jump to abstract materials and methods results discussion conclusions footnotes references measurement of carotid plaque volume by 3-dimensional ultrasound," *Stroke*, vol. 35, no. 4, pp. 864–869, 2004.