Aus dem Abteilung für Medizintechnik

der Medizinischen Fakultät

der Otto-von-Guericke-Universität Magdeburg

Chair of Intelligent Catheter and Image Guided Therapies (INKA)

Thyroid Texture Classification using Machine Learning in conjunction with Autoregressive Modeling and Deep Learning

Dissertation

zur Erlangung des Doktorgrades

Dr. rer. medic.

(doctor rerum medicarum)

an der Medizinischen Fakulät

der Otto-von-Guericke-Universität Magdeburg

Submitted by Prabal Poudel

from Chitwan, Nepal

Magdeburg, 2019

Supervised by:
Prof. Dr. Michael Friebe
Prof. Dr. med. med. Christoph Arens
Dr. Alfredo Illanes

Everybody is a genius. But if you judge a fish by its ability to climb a tree, it will live its whole life believing that it is stupid.

- Albert Einstein

Acknowledgments

First and foremost, I would like to express my sincere gratitude to Prof. Dr. Michael Friebe for providing me an opportunity to work at the Chair of Intelligent Catheter and Image Guided Therapies (INKA), Otto-von-Guericke University and also as a visiting researcher at General Electrics Healthcare, Milwaukee, USA. It is an honour for me to have had these great experiences at Prof. Friebe's chair. I would also like to express my gratitude to Prof. Dr. Christoph Arens for his invaluable supports in the clinical aspects of this thesis.

This thesis would not have been a success if it was not for my supervisor Dr. Alfredo Illanes who was always there to help me with my queries and guide me to the write path no matter what. His constant monitoring and supervision of my progress always guided me to the right path. I would like to extend my gratitude to my colleagues Elmer Ataide, Sathish Balakrishnan, Naghmeh Mahmoodian and Nazilla Esmaeili and my student Onur Gulkokan for their direct involvement with my work in different ways. Similarly, I would like to thank General Electric Healthcare and Mike Washburn and Yelena Tsymbalenko in particular for having me at GE and helping develop my work into a industrial product. Also, I would like to thank all my colleagues at INKA who were my family throughout my stay for their constant support and generating a motivational atmosphere.

Last but not the list, I would like to acknowledge my parents, Eak Narayan and Sabitri, my grand-parents, Prajapati and Ranjati and my sister Prashamsa and all of my family and friends, for encouraging me to work with enthusiasm and dedication in my PhD.

Thyroid Texture Classification using Machine Learning in conjunction with Autoregressive Modelling and Deep Learning

Prabal Poudel

Department of Medical Engineering
Otto-von-Guericke University
Magdeburg, Germany
2019

ABSTRACT

Analysis of medical images play a crucial role in diagnosis and treatment of several diseases in human body. Texture classification is an important tool for segmentation, tissue characterization and organ/boundaries detection in medical images. In this research, we mainly focussed on characterization of tissues in thyroid Ultrasound (US) images so that a thyroid region can be classified from the non-thyroid region. The goal of this thesis was to segment a thyroid region in 2D thyroid US images by characterizing the thyroid and non-thyroid textures using several image and signal based texture classification approaches. The segmented thyroid images could be used for 3D reconstruction and computation of the thyroid volume. The volumetric analysis of thyroid allows for diagnosis of probable thyroid diseases.

The first part of the thesis focuses on using classic image based methods to segment the thyroid. Active Contours Without Edges (ACWE), Graph Cut (GC) and Pixel Based Classifier (PBC) were used for thyroid segmentation in 2D US images. These approaches were compared based on accuracy, computation time, robustness and level of human interactions required.

The second part explains a novel feature extraction technique that parametrically models a signal version of the US image as a data resulting from a dynamical process.

Autoregressive (AR) modelling is used to compute several energy based features which are used to train different machine learning (ML) based classifiers. Similarly, Higher Order Statistical Analysis was also used as another feature extraction technique in a separate study. The extracted features were then used for training several machine learning classifiers. The trained classifiers were later used to classify the thyroid and non-thyroid textures.

The final part focuses on using current deep-learning (DL) based approaches to segment the thyroid. We trained a U-Net Convolutional Neural Network and a Fully Connected Convolutional Neural Network (FCNN) using several 2D thyroid US images. The trained CNNs were used for segmenting the test thyroid US images. Throughout the study, we saw that all the ML and DL based approaches require large amount of training images to segment the thyroid with significant accuracy. Hence, we explored the possibilities of generating synthetic 2D thyroid US images using Generative Adversarial Networks (GAN).

Resume - Prabal Poudel

Prabal Poudel was born on 24.03.1993 in Chitwan, Nepal. He completed his Bachelors in Science in Electrical Engineering and Computer Science from Jacobs University in 2014. At Jacobs, he had the opportunity to visit Washington State University, Washington, USA as a semester exchange student. During the summer holiday of his second year (June 2013), he got an opportunity to work as a scientific intern at Fraunhofer Mevis, Bremen, Germany. It was during this work at Mevis, he got familiar to medical imaging and ever since, he has been involved in this field. He moved to Bonn to pursue his Masters in Science degree in Computer Science in 2014. At Bonn, he specialized in computer vision, computer graphics and audio processing. During the final year of his masters, he moved to the Chair of Intelligent Catheters and Technology (INKA) at Otto-von-Guericke University Magdeburg under Prof. Michael Friebe for his masters thesis. Immediately, after completion of his masters from Bonn in October 2016, he started his PhD at INKA in January 2017. Towards the end of his second year of PhD, he went to General Electric Healthcare at Milwaukee, USA. At USA, he worked on implementation of thyroid segmentation software (in C++) which is supposed to be incorporated in the later release of GE Logiq E10 Ultrasound System.

Eidesstattliche Erklärung

Ich erkläre, dass ich die der Medizinischen Fakultät der Otto-von-Guericke-Universität zur Promotion eingereichte Dissertation mit dem Titel

"Thyroid Texture Classification using Machine Learning in Conjunction with Autoregressive Modelling and Deep Learning"

im Universitätsklinikum Magdeburg, Otto-von-Guericke Universität Magdeburg

mit Unterstützung durch

Prof. Michael Friebe, Prof. Christoph Arens und Dr. Alfredo Illanes

ohne sonstige Hilfe durchgeführt und bei der Abfassung der Dissertation keine anderen als die dort aufgeführten Hilfsmittel benutzt habe.

Bei der Abfassung der Dissertation sind Rechte Dritter nicht verletzt worden.

Ich habe diese Dissertation bisher an keiner in- oder ausländischen Hochschule zur Promotion eingereicht. Ich übertrage der Medizinischen Fakultät das Recht, weitere Kopien meiner Dissertation herzustellen und zu vertreiben.

Magdeburg, den 08.08.2019

—————————————

Prabal Poudel

Table of Contents

List of Figures

xiv

List of Tables

Outline

A brief description of the main concepts of this thesis are presented in the following section.

Chapter I: Introduction In this chapter, the importance of thyroid in human body as well as US imaging for thyroid disease diagnosis are discussed. Additionally, other imaging modalities are discussed to highlight the safety and cost effectiveness of US compared with these modalities.

Chapter II: Classic Image Based or Data Driven Segmentation Methods for Thyroid Segmentation This chapter presents the existing image based segmentation techniques specially for thyroid segmentation as well as the 3D reconstruction of the segmented thyroid for volumetric analysis. The problems associated with the segmentation approaches are highlighted to find a robust solution.

Chapter III: Thyroid Segmentation with Machine Learning Classifiers Trained using Features from Different Feature Extraction Techniques Based on the problems identified in the previous chapter, several feature extraction techniques are discussed. A novel feature extraction technique by modelling a signal version of an US image is proposed and discussed. Similarly, higher order statistical features as well as features obtained using the apriori information based on the physics of the US imaging process have been discussed. These features are later used to train different machine learning classifiers for thyroid segmentation.

Chapter IV: Deep Learning for Thyroid Segmentation Usually, it is a tough and time consuming task to compute robust hand-crafted features in thyroid US images. This problem can be easily tackled by using some deep learning methods since these methods compute very generalized to high level features. This chapter presents the segmentation of thyroid using two convolutional neural network architectures.

1

Chapter V: Discussion and Conclusion A general explanation of different approaches used for thyroid segmentation along with the results, advantages, drawbacks and possible improvements are discussed in this chapter.

Selected Publications in Chronological Order

[**Poudel *et al.*, 2019**] **P. Poudel**, A. Illanes, E. Ataide, N. Esmaeili, S. Balakrishnan and M. Friebe, "Thyroid Ultrasound Texture Classification Using Autoregressive Features in Conjunction with Machine Learning Approaches.", IEEE Access, June 2019. *Enlisted in the Second Part of the Thesis.* **Impact Factor: 4.098**

[**Illanes *et al.*, 2019**] A. Illanes, N. Esmaeili, **P. Poudel**, S. Balakrishnan and M. Friebe, "Parametrical modelling for texture characterization - A novel approach applied to ultrasound thyroid segmentation," PLOS ONE, 14(1): e0211215, January 2019. *Enlisted in the Second Part of the Thesis.* **Impact Factor: 2.766**

[**Poudel *et al.*, 2018**] **P. Poudel**, A. Illanes, D. Sheet and M. Friebe, "Evaluation of commonly used algorithms for thyroid ultrasound images segmentation and improvement using machine learning approaches," Journal of Healthcare Engineering, Article ID: 8087624, 1-13, September 2018. *Enlisted in the Second Part of the Thesis.* **Impact Factor: 1.295**

[**China *et al.*, 2018**] D. China, A. Illanes, **P. Poudel**, M. Friebe, P. Mitra and D. Sheet, "Anatomical Structure Segmentation in Ultrasound Volumes using Cross Frame Belief Propagating Iterative Random Walks," IEEE Journal of Biomedical and Health Informatics, 1-1. August 2018. *Enlisted in the Second Part of the Thesis.* **Impact Factor: 4.217**

[**Poudel *et al.*, 2019**] **P. Poudel**, A. Illanes and M. Friebe, "Patch Based Texture Classification of Thyroid Ultrasound Images using Convolutional Neural Network", In Proc. IEEE Engineering in Medicine and Biology, Berlin, Germany, July 2019. *Enlisted in the Second Part of the Thesis.*

[**Mahmoodian** *et al.*, **2019**] N. Mahmoodian, **P. Poudel**, A. Illanes and M. Friebe, "Higher Order Statistical Analysis for Thyroid Texture Classification and Segmentation in 2D Ultrasound Images", In Proc. IEEE Engineering in Medicine and Biology, Berlin, Germany, July 2019. *Enlisted in the Second Part of the Thesis.*

[**Poudel** *et al.*, **2018**] **P. Poudel**, A. Illanes, E. Ataide and M. Friebe, "Convolutional Neural Network for Texture Based Thyroid Ultrasound Segmentation and Classification," IEEE International Student Conference, Banglore, India, December 2018.

[**Poudel** *et al.*, **2018**] **P. Poudel**, S. Ghosh, T. Sühn, A. Illanes and M. Friebe, "Classification of thyroid texture in ultrasound images using Bayesian Network and Adaptive Boosting," In Internal Journal of Computer Assisted Radiology and Surgery, Volume 13, Berlin, Germany, June 2018.

[**Poudel** *et al.*, **2018**] **P. Poudel**, A. Illanes and M. Friebe, "Thyroid Texture Classification in Ultrasound Images using Bootstrap Aggregating," IEEE Engineering in Medicine and Biology Conference, Honolulu, USA, June 2018.

[**Poudel** *et al.*, **2018**] **P. Poudel**, E. Ataide, A. Illanes and M. Friebe, "Linear Discriminant Analysis and K-Means Clustering for Classification of Thyroid Texture in Ultrasound Images," IEEE Engineering in Medicine and Biology Conference, Honolulu, USA, June 2018.

[**Poudel** *et al.*, **2017**] **P. Poudel**, A. Illanes, C. Arens, C. Hansen and M. Friebe, "Active contours extension and similarity indicators for improved 3D segmentation of thyroid ultrasound images," In SPIE Medical Imaging, International Society of Optics and Photonics, Orlando, Florida, United States of America, February 2017.

[**Wunderling** *et al.*, **2017**] T. Wunderling, B. Golla, **P. Poudel**, M. Friebe and C. Hansen, "Comparison of thyroid segmentation techniques for 3D ultrasound," In SPIE Medical Imaging, International Society of Optics and Photonics, Orlando, Florida,

4

United States of America, February 2017.

[**Poudel *et al.,* 2016**]   **P. Poudel**, C. Hansen, J. Sprung and M. Friebe, "3D segmentation of Thyroid Ultrasound Images using Active Contours," Current Directions in Biomedical Engineering, 2(1): 467-470, September 2016.

List of Abbreviations

MRI   Magnetic Resonance Imaging

US   Ultrasound

CT   Computed Tomography

PET   Positron Emission Tomography

SPECT   Single Photon Emission Computed Tomography

ML   Machine Learning

DL   Deep Learning

2D   2-Dimensional

CNN   Convolutional Neural Networks

RF   Radio Frequency

GE   General Electric

ACWE   Active Contours Without Edges

GC   Graph Cut

GMM   Gaussian Mixture Model

DC   Dice Coefficient

PBC   Pixel Based Classifier

SNR   Signal to Noise Ratio

SVM	Support Vector Machine
AR	Autoregressive
HOSA	Higher Order Statistical Analysis
IRW	Iterative Random Walks
GVF	Gradient Vector Flow
CWT	Continuous Wavelet Transformation
D1	Dataset 1
D2	Dataset 2
RBF	Radial Basis Function
CNN	Connected Convolutional Neural Network
FCNN	Fully Connected Convolutional Neural Network
ReLU	Rectifier Linear Unit
HD	Hausdorff Distance
GAN	Generative Adversarial Networks

CHAPTER 1

Introduction

History of Medical Imaging: Medical Imaging is the process of visualizfing the interior body parts for non-invasive medical intervention and clinical analysis. The history of medical imaging dates back to 1895 when Wilhelm Conrad Roentgen discovered X-ray for the first time [13]. Roentgen made it possible to visualize the internal body structures without any human intervention. This way of non-invasive medical intervention revolutionized the way of diagnosing and treating injuries and diseases in human body. Since then, many new medical imaging technologies like Magnetic Resonance Imaging (MRI), Ultrasound (US), Computed Tomography (CT), Positron Emission Tomography (PET), Single Photon Emission Computed Tomography (SPECT), Elastography, Echocardiography, Spectroscopy, etc. have been discovered.

However, the advent of computers in the world of medical imaging was seen in early 1970s when CT and MRI were introduced. Since then, numerous approaches have been developed to make medical imaging an easily accessible and easy to use technology. With the introduction of computers and new medical imaging technologies, understanding of human physiology and anatomy has increased significantly. The term 'Machine Learning' was coined in 1959 by Arthur Samuel and its first use was seen in early 1970s when Edward Shortliffe developed an early backward chaining expert system called MYCIN to identify bacteria causing severe infections such as bacteremia and meningitis.

Over the span of last fifty years, medical and technological advancements have enabled the growth of healthcare-related applications. The priorities have been on correct diag-

nosis of diseases, assistance and validation of the treatment and monitoring the possible recurrence of the diseases.

1.1 Background and Motivation

Context: Thyroid is one of the largest endocrine glands in human body. It is a butterfly shaped organ located in the neck anterior to the trachea and below the Adam's apple. It is a two lobed gland with each lobe connected by a bridge (isthmus) in the middle as shown in Fig. 1.1. It weighs about 15 to 20 gram in adults. Thyroid is involved in a lot of body mechanisms like regulating the metabolic functions in the body including heart rate, cardiac output, lipid metabolism, heat regulation and skeletal growth, controlling energy sources usage, synthesis of proteins and controlling body's sensitivity to other hormones [5].
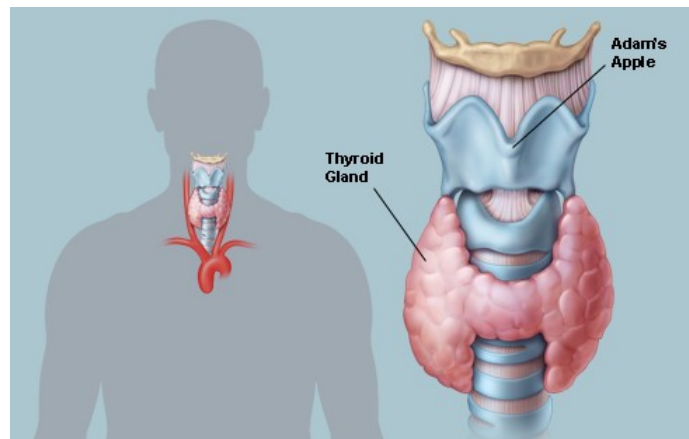
Fig. 1.1: Thyroid Gland in Human Body [4]

Due to these important functionalities in the human body, thyroid is one of the important organs. However, thyroid is susceptible to many thyroid diseases like the Graves' (excessive production of thyroid hormones), subacute thyroiditis (inflammation of thyroid), thyroid cancer, goiter (thyroid swelling), thyroid nodules (small abnormal lump

grows in the thyroid), etc [5, 53].

Problem Statement: Most of the thyroid diseases often involve the change in the shape and size of thyroid over time. Hence it is essential to monitor and track these changes. Thyroid segmentation and volume computation of thyroid are the two important imaging tools that can be used to monitor the different stages of thyroid. Improved diagnosis, treatment, follow-up and monitoring of thyroid diseases like thyroid nodule, Goiter, Graves', thyroiditis, thyroid cancer, etc. have been made possible by the recent advancements in the medical imaging sector.

US imaging is used widely for imaging the thyroid in human body. However, the quality of images in US is not optimal for analysing different anatomical structures due to the presence of speckle noise and low contrast. This makes it a very challenging task to process US images compared to MRI or CT. Similarly, the manual segmentation/annotation of thyroid region is highly subjective as it can differ from person to person. Hence, it is essential that an automated approach has to be used to identify the thyroid region in an US image.

Proposal: US, MRI, CT and scintigraphy have been used for imaging of the thyroid gland. A detailed analysis of these imaging modalities are presented in the next section. US is the preferred imaging modality since it is much cheaper, safer and painless for patients compared to other modalities. Hence, we propose a 2D freehand thyroid data acquisition approach compared to the acquisition of the 3D volume directly. This is because, segmenting the individual 2D thyroid US images and reconstructing them to a volume possess greater advantages than segmenting the volume directly as segmenting the 3D image is very complex and requires larger computation power. Additionally, 2D segmentation of thyroid allows a detailed analysis of the shape of the thyroid. For the segmentation of the thyroid, we propose two novel feature extraction methods combined with different machine learning approaches to classify the textures inside and outside

10

the thyroid and compare these approaches with several classical image based and deep learning based approaches. Similarly, two different tools are proposed to 3D reconstruct the segmented thyroid for volumetric analysis.

Hypothesis: The presence of speckle noise, low contrast and signal to noise ratio (SNR) makes it difficult to segment the thyroid in US images [25]. In order to tackle these issues, a novel feature extraction technique has been proposed. This approach analyzes the textures in the thyroid US images as signals resulting from a dynamical process. The main hypothesis of this approach is that, the textures from thyroid and non-thyroid regions have completely different dynamical patterns and analysing these dynamical patterns using a parametrical modelling (i.e. autoregressive modelling) allows classification of different textures in thyroid US images. This approach works even with less amount of data and is not affected by the presence of speckle, low contrast and SNR as seen in US images.

As mentioned earlier, US images contains speckle noise which is known to have a non-Gaussian distribution and the formation of these noise result from a non-linear process. Hence, speckle noise is dependent on the behaviour of the sound propagation in different body structures depending on the texture [26, 50]. Thus, the speckle noise formation from the thyroid and non-thyroid regions should also be different in characteristics. A method (i.e. higher order statistical analysis) that could process these non-linear data (i.e. speckle noise) and model the differences in the characteristics of speckle noise along with different textures in a thyroid US image would allow us to classify different regions with better accuracy.

These two feature extraction techniques combined with machine learning should perform similarly if not outperform the classic segmentation approaches as well as the deep learning based approaches.

Objectives: The main objectives of this thesis will be to propose two novel feature extraction techniques combined with machine learning to classify the textures in thyroid US images. These approaches in conjunction with machine learning will be compared with the pre-existing classic image segmentation techniques as well as with the recent deep learning approaches to prove the robustness of the feature extraction process. Similarly, two tools for 3D reconstruction of segmented thyroid US images will be introduced for volumetric analysis.

Contributions: The main contributions of this thesis are (i) segmentation of the thyroid using classic image based approaches and identifying their drawbacks (Chapter 2), (ii) developing novel and robust feature extraction techniques for thyroid segmentation (Chapter 3) and (iii) applying deep learning methods for thyroid segmentation and generation of synthetic thyroid images (Chapter 4):

(i) The segmentation of thyroid US images using three widely used image based (data-driven) approaches are presented in Chapter 2. These approaches directly use the pixel information from the images to segment the thyroid. Similarly, these approaches are used for segmentation of a two-dimensional (2D) US image at a time. Our contribution to this work is that we extend these approaches to segment not only one image at a time, but a series of freehand thyroid US images. This allows the medical experts and the physicians to segment a large number of thyroid US images at once instead of segmenting them one by one. We have also identified the limitations of these approaches for thyroid segmentation, proposed two machine learning based approaches to address the limitations of the image based approaches and compared their results with the approaches from the state of art.

(ii) A novel and robust feature extraction method using the parametrized signal based

version of the thyroid US images has been presented in Chapter 3. Along with it, another approach of higher order statistical features extraction has been presented. These robust features are used for training several ML based classifiers. A comparison with different methods in the literature proves that the extracted features are very robust and allow the ML classifiers to segment the thyroid US images with high accuracy.

(iii) Despite of the high accuracy obtained using our features in Chapter 2, it is a cumbersome task to compute the hand-crafted features. Hence, we have used Convolutional Neural Networks (CNN) to segment the thyroid US images in Chapter 3. These approaches however require a large amount of US images for training. Hence, computation of synthetic thyroid US images using Generative Adversarial Network (GAN) could be a possible solution to solve the problem of data availability. This work on GAN has been presented in the discussion and conclusion section in the last chapter of the thesis.

1.2 Thyroid Imaging

There have been four major modalities being used for thyroid imaging: (1) Scintigraphy, (2) Computed Tomography (3) Magnetic Resonance Imaging and (4) Ultrasound [18, 51, 58]. All of the four techniques provide structural as well as the location and size information of the thyroid glands. The first technique provides the spatial distribution of the thyroid functional attributes (i.e. how the tissues interact with the external elements for example a radioactive isotope)whereas the last three techniques provide the spatial distribution of the structural attributes in the thyroid such as the varying degree of echogenicity of the tissues under examination

Scintigraphy: It is a diagnostic test where specialized scanners called Gamma Camera

capture the emitted radiation by the radioisotopes either by themselves or tagged to a protein or other molecules travel to some specific organs or tissues to form a 2D image. Several iodine-based radionuclide (Iodine-131, Iodine-123) or technetium-99m pertechnetate are widely used isotopes for thyroid scintigraphy. Two new scintigraphy techniques, PET and SPECT have been used recently for the imaging of the thyroid. Both of these approaches produce 3D images using a gamma camera to detect gamma ray photons emitted from the radioisotopes inside the body.

Scintigraphy is specially used for imaging and evaluation of the thyroid nodule as hot, warm or cold depending on the amount of radioactive isotopes taken by the nodules. This allows the determination of the gland size, localizing of the thyroid tissues, evaluation of the nodules, differentiation of various forms of goitre and identifying non functioning cancers [51]. Due to these factors scintigraphy has been widely used for imaging the pathological thyroids.

Computed Tomography (CT): It is an imaging technique that combines the X-ray images taken from different angles and cross-sections to create a 3D volumes of different anatomical structures under visualization. A CT scanner comprises of a single or multiple X-ray source/s and detector/s. CT can be used for visualization of both the hard and soft tissues and organs.

For thyroid imaging, it is generally used in four different scenarios which are detection of the incidental thyroid nodule, evaluation of the thyroid metastases, presurgical imaging for invasive disease and evaluation for recurrence in the post-treatment neck. [27]. CT are also used in conjunction with PET and SPECT for detecting thyroid nodules and cancers.

Magnetic Resonance Imaging (MRI): It is a medical imaging technique that uses strong magnetic fields, magnetic field gradients and radio waves to generate images of

different anatomical structures in the body. Specifically, the hydrogen atoms inside the human body are aligned to the direction of the static magnetic field inside the MRI machine. A radio frequency (RF) pulse is then passed to the human body causing the protons to be aligned perpendicular or anti-parallel to the static magnetic field. When the RF pulse is turned off, the protons flip back to their original spin by releasing electromagnetic energy. The MRI image is produced by analysing the different energy released by different tissues during the relaxation process.

MRI is used for visualization of the soft tissues in human body (for example, the abdomen, brain, thyroid etc.) Similar to CT, MRI can also be used for four different scenarios as mentioned above. Unlike CT and scintigraphy, MRI does not involve the usage of X-rays or the ionizing radiations. This makes it safer compared to the aforementioned techniques.

Ultrasound (US): US images are produced by the reflection of US waves from different body structures. The superficial location of the thyroid in the neck makes it easily accessible using the high-frequency US waves (1-15 MHz).

Two different modes of US are used in the medical imaging: A-mode (1-D amplitude information) that uses a single transducer to scan a line through the body and plots the echoes as a function of depth and B-mode or 2D mode, which uses a linear array of transducers simultaneously to scan a plane through the body and computes a 2D image using the acoustic properties of the US images reflected from various body parts.

Recent technological advancements have made it possible to visualize 3D US images. For the 3D visualization, a special 3D probe containing curvilinear transducers inside a casing which automatically swivel during the image acquisition are used. Similarly, a freehand 2D US scans can be obtained using a tracking system (for example, electromagnetic (EM), optical, inertial etc.) and they can be computed volumetrically to

compute a 3D volume. US is usually used for determining the size and number of thyroid nodules, to access the volume of the thyroid gland and to differentiate the thyroid tissues from the non-thyroid tissues. Doppler imaging in US even allows screening of the thyroid nodule for malignancy [18].

Compared to the imaging modalities above, US is safer to use for the patients since it uses only US waves unlike CT which uses the harmful X-rays and scintigraphy which exposes the patients to ionizing radiations. Similarly, it is portable as compared to MRI which is very heavy because of the big magnets used to create magnetic fields. Additionally, US is very cost effective compared to other imaging modalities. Due to these advantages, we decided to use US for imaging the thyroid.

1.3   Data Acquisition

In this work, LogiqE9 US machine from General Electric (GE) was used to acquire freehand 2D thyroid US images. The machine was equipped with Ascension driveBay EM tracking system. Hence, each US was acquired along with a tracking matrix. The tracking matrix provided the transformation from the origin of the tracking system to the center of each image. These tracking matrices are used for 3D reconstruction of the segmented images to compute a 3D volume of thyroid. 6L and 12L transducers were used for acquiring the linear 2D thyroid US images. A total of two datasets were acquired using the Logiq E9 system. The first dataset (Dataset 1 - D1) consisted of six subjects with each subject containing between 53 and 189 2D thyroid US images. A total of 675 thyroid images with an image size of 760 x 500 pixels were acquired. Similarly, the second dataset (Dataset 2 - D2) consisted of sixteen subjects with each subject containing between 156 and 289 2D thyroid images. A total of 3,370 thyroid US images with an image size of 760 x 1020 pixels were acquired. The two datasets were

acquired by medical experts from SurgircEye GmbH and University Clinic Magdeburg,
Germany [3].

CHAPTER 2

Classic Image Based or Data-Driven Segmentation Methods for Thyroid Segmentation

2.1 Definition of Segmentation in Medical Imaging Context

The term *segmentation* refers to partitioning of an image into multiple segments. During this partitioning, each pixel in the image is assigned to one of the segments depending on certain characteristics. The segments can be from 2 to many. An example of segmentation of brain MRI image is shown in Fig. 2.1. In the figure, the red segment refers to the white matter, green refers to the gray matter and blue represents the cerebrospinal fluid in the brain.
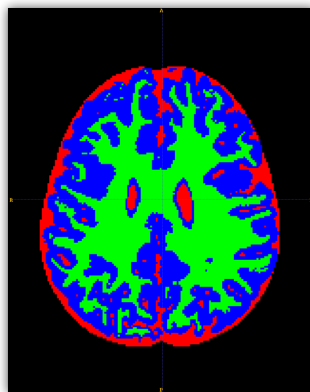
Fig. 2.1: GE LogiqE9 US machine

Each pixels in a segment share similar characteristics such as color, intensity, texture, etc. However, there is a difference of characteristics between the segments. Segmentation is an essential tool in image analysis, object recognition, visualization and many different image processing tasks. One of the application areas of segmentation is medical imaging as it can be used to compute region of interests, area and volume of different anatomical structures for locating tumors and other pathologies, planning and simulation of surgery scenarios and many more [54].

When it comes to image based segmentation, the images are partitioned into different segments by directly operating on the pixel information (i.e. data-driven approaches) in the images. For example, edges, colors, gradients, histograms, etc. are computed using the pixel intensities in the images and each pixel is assigned a segment depending on these characteristics. There are several state-of-the arts techniques that use this information for thyroid segmentation. We will discuss about them in the section below.

2.2 Related Work

There have been may research works that use the pixel information in the 2D and 3D US images for segmentation of the thyroid. Segmentation of 2D thyroid US images using edge detection, thresholding, region splitting and merging, watershed segmentation, active contour, graph theory and segmentation using normalized cut (Ncut) were performed by Zhao *et al.* [70]. Active contour without edges (ACWE), localized region based active contour and distance regularized level set were used by Kaur *et al.* [30] to segment the 2D thyroid US and Scintigraphy images. Similarly, a local region-based active contour was proposed in [41] for segmentation and area computation of the thyroid in 2D US images. Augustin *et al.* [62] used fuzzy c-means algorithm, histogram clustering, QUAD tree, region growing and random walk [64] to segment and test thyroid US

images. Mylona *et al.* used the local geometry information (i.e. orientation coherence in the edges of the regions to be segmented) to control the evolution of the contour in region based active-contour implementation for segmentation of medical images [47,48]. Level set active contour models based on variable background active contour and joint echogenicity texture were used in [43, 60] to segment thyroid US images.

Compared to the 2D segmentation of thyroid US images, there have been very limited research works to segment a full 3D thyroid image using the classic image based segmentation approaches. Kollorz *et al.* have proposed a semi-automated thyroid segmentation approach for volumetric quantification using geodesic active contour [32]. Poudel *et al.* used ACWE, graph cut and pixel based classifier approaches to segment the 2D thyroid US images and reconstructed them to get a 3D segmented thyroid [55]. Osman [52] worked on his PhD thesis to perform a complete segmentation and analysis of 3D thyroid US images by thresholding the voxel intensities and then connecting the similar voxels to predict the segmenting region. Apart from US images, there has been a work on segmentation of thyroid CT images in 3D. Dornheim *et al.* used 3D mass-spring models for thyroid segmentation by creating 3D deformable shape models [23].

2.3 Contributions

The approaches discussed above are used for segmentation of a single US image at a time. These non or semi-automatic approaches have to be initialized by the user every time they have to segment an image in a freehand US dataset. Hence, to get rid of this problem, three widely used classic image based segmentation approaches were chosen and an extension was made so that they could segment a sequence of thyroid US images in a dataset without the need of the user to initialize or operate on individual US images. ACWE, GC and PBC were chosen as these approaches usually work on a 2D image but

they can be extended to segment a sequence of freehand US images by making use of the spatial relationships between the corresponding thyroid US image frames. A brief explanation on how the extension of segmentation from a single US image frame to corresponding frames was done using ACWE, GC and PBC are explained below.

The segmentation results using these approaches showed that the segmentation was not highly accurate since these approaches failed to segment the smaller parts in the thyroid (eg: isthmus). Hence, two different machine learning based approaches were proposed for segmenting the thyroid images in 3D. Random Forest Classifier (RFC) and a U-Net convolutional neural network were used for this task. These approaches could segment the thyroid images with better accuracy compared to the three image based approaches.

Active Contour Without Edges (ACWE): The segmentation of thyroid US image was performed using the level set approach from Chan and Vese [15]. ACWE starts with the initialization of a square/rectangular mask around the thyroid region. The initialized contour then evolves over certain iterations towards object contours using the principle of energy minimization [29]. After segmenting an image using ACWE, a center of mass of the segmented object was computed which was used to find a probable center of mass of the thyroid in next image. Tracking matrices acquired during the image acquisition phase were used to compute this center of mass of consecutive images around which an initialization of a contour is done automatically. The details of the computation of the center of mass, probable center of mass of consecutive image frames and initialization of the initial contour are presented in Chapter A.

Graph Cut (GC): The approach of GrabCut algorithm from Rother *et al.* [16] is used to segment the thyroid US images using GC. It starts with the creation of an initial trimap by marking inside and outside of thyroid region as foreground and background regions. Two Gaussian Mixture Models (GMM) are computed from the initial marking and a graph is build at the final stage that separates the foreground pixels from the

background pixels. This distinguishes the thyroid region from the non-thyroid region. The initial user initializations are interpolated in the corresponding image frames to mark different regions, create corresponding GMMs and finally segment the consecutive images automatically. The user can make multiple initializations to improve the accuracy of segmentation. However, this increases the computation of the algorithm. Thus, an optimization between the number of user interactions and computation time has to be made to obtain the most accurate segmentation results using GC. A detailed explanation of this approach is presented in Chapter A.

Pixel Based Classifier (PBC): Three different image based features on a 4 and 8-neighbourhoods of each pixel were calculated to train the decision trees for thyroid segmentation. The selection of the features for this work was based on the work of Chang *et al.* [16]. PBC starts with the user clicking on the inside and outside of the thyroid region in multiple thyroid images. The features are computed from these areas which are passed for training the decision trees. The trained decision trees later segment the thyroid region from the non-thyroid regions. The user can click in different regions after the segmentation process if the segmentation results are not highly accurate. The user should be very careful in selecting the thyroid and non-thyroid regions as the wrong initialization might lead to wrong segmentation. A post-processing step was followed after the segmentation to get rid of over segmentation if there were any. This step involved some morphological operations to find the largest connected component in the segmented thyroid.

The segmentation results from these three approaches were later used for 3D reconstruction and volume computation of the segmented thyroid using Imfusion [1] and MeVisLab [2] imaging tools. The segmented images were saved as binary images to make a video file. The video file along with the tracking matrices acquired during the image acquisition phase could be passed to the ImFusion and MeVisLab for 3D recon-

struction of the thyroid as well as volume computation. An interpolation between the corresponding image frames of the US sweep is done to compute the 3D thyroid.

A comparison of segmentation accuracy, computation time, robustness of the algorithm and number of user interactions required were performed. ACWE performed the best in terms of accuracy with an average Dice Coefficient (DC) of 0.80. Similarly, PBC required the lowest number of user interactions and was the fastest of all the three methods.

To tackle the problems faced using these three approaches, two machine learning based approaches were proposed. RFC and U-Net could segment the thyroid images directly in 3D and produced better segmentation results. RFC and U-Net could segment the thyroid with an accuracy of 0.862 and 0.876 respectively in terms of Dice Coefficient (DC). Similarly, these approaches were automatic and could directly obtain the volume of the thyroid. Still, the accuracy of around 0.87 with these approaches were not sufficient for our task. Hence, we decided to explore more techniques using the feature extraction combined with machine learning and deep learning approaches.

The details of the implementation of all the five algorithms, their performance analysis and comparison with the works from literature are presented in Chapter A.

CHAPTER 3

Thyroid Segmentation with Machine Learning Classifiers Trained using Features from Different Feature Extraction Techniques

3.1   Introduction and Definition

US images have speckle noise [25] which are produced by the interference of the returning US waves at the transducer aperture as the US images are produced when the reflected sound waves from different surfaces inside the body are picked up by the transducer. The presence of speckle noise along with shadow artifacts and low signal to noise ratio makes it very difficult to segment US images. A preprocessing step can be applied to improve the US image quality, however this changes the textural properties of the structures in US images. Hence, a robust approach has to be applied to extract novel features even in presence of the speckle noise, shadow artifacts and low SNR and without changing the natural textural characteristics of different anatomical structures in US images.

Texture Based Features: are used to quantify the image texture inherent in the images by measuring the smoothness, coarseness and regularity of pixels which form an image [8, 45]. These features can be used for various applications such as segmentation, registration and lesion classification [14]. Generally, statistical features like histograms,

entropy, homogeneity, mean and variance, geometrical features like perimeter, area, convexity and extent of the structure under investigation, morphological features such as formation factor, roundness, long and short axis, etc, and frequency or spectral based features are used for analysing the textures in US images [45]. These approaches are however data-driven, meaning that the texture features are computed directly using the pixel intensities in the images that are affected by the presence of speckle noise.

3.2 Related Work

Thyroid segmentation and nodule characterization and classification have been major research topics in the literature. Many approaches have been presented in the literature for extracting features in thyroid US images. China *et al.* [20] have used apriori information based on the physics of the US imaging process and segmented the thyroid in US images by applying Iterative Random Walks and Random Forest. Similar to thyroid segmentation, several feature extraction techniques have been proposed for classifying thyroid nodules. Statistical features [28, 33, 61], spectral based features [7, 10], higher order statistical based features [11, 57], wavelet based features [8, 9, 12] and fractal based features [10, 56]. Similarly, first and second order statistical features (like gray-level histogram, entropy, correlation, mean, etc.) were used by Muller *et al.* [46] to classify thyroid nodules as benign or malignant.

Additionally, a novel boundary detection method and local binary patterns for texture analysis was proposed in a work by Keramidas *et al.* [31]. Apart from the feature discussed above, several machine learning based approaches that train on using the extracted features have been used for thyroid segmentation and nodule characterization in US images. A polynomial Support Vector Machine (SVM) was used in [61] to segment thyroid gland in US images. Chang *et al.* [17] proposed a radial basis function (RBF)

neural network to segment the blocks of thyroid gland. Speckle related pixels and imaging artefacts were used as a source of information or features by Narayan *et al.* [49] to perform a multi-organ (thyroid, carotid, artery, muscles and trachea) segmentation in US images. Nikita *et al.* [63] used grey level co-occurence matrix based features and classified them using SVM, k-nearest neighbour and bayesian classifiers to classify benign and malignant thyroid nodules. Similarly, SVM and probabilistic neural networks were used by Tsantis *et al.* [68] to classify thyroid nodules using morphological and wavelet features.

3.3 Contributions

The feature extraction techniques discussed above were used for extracting image based features from thyroid US images and these features were used for training the ML classifiers. This data-driven approach of feature extraction does not capture the true dynamical differences of the textures inside the thyroid due to the presence of speckle noise, low contract and high SNR in the US image. Hence a robust feature extraction method had to be used that could capture the dynamical texture difference inside of the thyroid and on the same time, it should not be affected by the presence of speckle noise, low contrast and SNR of the US images.

Hence, to tackle these problems, two different feature extraction techniques were used separately and the extracted features were used for training different ML classifiers to classify and finally segment the thyroid in US images. A brief explanation of these two feature computation techniques and eight machine learning approaches that were used are discussed below. Additionally, a statistical mechanical model based on the physics of the US imaging process along with iterative random walks (IRW) solver, random forest (RF) and a gradient vector flow (GVF) was used for volumetric segmentation of

thyroid US images. A detailed explanation of this approach is presented in Chapter B.

3.3.1 Feature Extraction

Autoregressive Modelling (AR): In this approach of feature extraction, a thyroid US image is first divided into smaller texture patches. The texture patches are converted to four different types of signals by traversing the texture patch matrices in ZigZag, Spiral, ZigZag Transpose and Spiral Transpose (see Figure C.3). These signals are then decomposed into four narrowband signals depending on different frequency components (low, high, middle and total frequency bands) by applying Continuous Wavelet Transformation (CWT). This transformation produces a total of 16 different narrowband signals which are modelled usign a parametrical AR model [42]. A total of 222 different features were computed using the parameters of the AR model which were undergone a series of optimization processes. Finally, 30 different energy based features were computed to train different ML classifiers.

This parametrical approach of modelling the thyroid texture models the texture dynamics inside and outside of the thyroid region by not operating directly on the pixel intensities but on the parameters of the modelled texture. The predictive characteristics of such a model representation provides with a good estimations of the characteristics of the texture in thyroid US images. Hence, it is not affected by the presence of speckle noise and low contrast and SNR issues in the US images. A complete explanation of the feature extraction and the AR modelling process is explained in Chapter C.

Higher Order Statistical Analysis (HOSA): This approach of feature extraction was also worked on the texture patches as in AR modelling. However, the patches are not converted to signals. The extracted texture patches were then used to compute the bispectrum using the bispectral analysis. Bispectral analysis is a higher order spectral

analysis technique that measures the assymmetry of a random process about its mean [65, 67]. For the detailed explanation of how these features were computed, a detailed explanation is provided in Chapter D. The bispectrum matrix is used for computation of the features. A total of ten different energy, frequency and entropy based features were computed by modelling the texture patches using bispectral analysis.

US images in general contain speckle noise which is known to have a non-Gaussian distribution and the formation of these noise results from a non-linear process. Hence, the speckle noise is dependent on the behavior of the sound propagation in the different structures inside the body depending on the texture [26, 50]. Since HOSA deals with the processing of the non-Gaussian data (i.e. speckle noise in our case), there is no need for any kind of pre-processing. This makes HOSA a well suited approach for extracting features from the thyroid US images.

3.3.2   Machine Learning for Thyroid Segmentation

Naive Bayesian: A naive bayesian classifier is a probabilistic model that represents the joint probability distribution of a set of random variables, which in our case are the AR features. These random variables were classified as thyroid and non-thyroid to segment the thyroid. It was tested using Weka [6] on D1 and obtained an accuracy of 86.26% with a sensitivity of 0.85 and specificity of 0.59.

Adaptive Boosting (Adaboost): Similar to Naive Bayesian, it was also tested using Weka on D1. Adaptive Boosting is a machine learning based algorithm which is used in conjunction with many other types of learning algorithms for performance improvement. Hence, it is a kind of ensemble learning. In our case we have used the decision stump classifier [44] as weak classifiers. This approach obtained an accuracy of 87.04 with a sensitivity of 0.79 and specificity of 0.604.

K-Means: k-means is an unsupervised machine learning method that is used for clustering of different observations into a predefined number of clusters. In our case, we clustered the 30 dimensional feature vectors into two different clusters (thyroid and non-thyroid). The approach obtained an accuracy of 89.66% with a sensitivity of 0.95 and specificity of 0.70 in D1 and an accuracy of 86.89% with a sensitivity of 0.891 and a specificity of 0.623 in D2.

Linear Discriminant Analysis (LDA): LDA is a form of dimensionality reduction method which can be used as a classification algorithm for predictive modelling. It is divided into two steps, 1) Discrimination where, the information is used in a learning set of labelled observations to construct a classifier (or classification rule) that will separate the predifined classes as much as possible and 2) Classification where a set of given measurements on a new unlabelled observation, use the classifier to predict the class of that observation. An accuracy of 90.6% with a sensitivity of 0.827 and specificity of 0.811 were obtained in D1 and an accuracy of 87.23% with a sensitivity of 0.771 and specificity of 0.575 in D2.

Bootstrap Aggregating (Bagging): Bagging is an ensemble machine learning approach that can be used for classification and regression problems. It is very similar to adaptive boosting with the only difference being on how the selection of the data is made for training the individual decision trees. A total of 50 different decision trees were used for the training purpose. The output of each of the decision trees are averaged to produce the final classification results. An accuracy of 93.34% with a sensitivity of 0.94 and specificity of 0.89 were obtained in D1 and an accuracy of 91.87% with a sensitivity of 0.8462 and a specificity of 0.5830 in D2.

Support Vector Machine (SVM): A SVM is a discriminative classifier that separates a labelled training dataset into different categories. A line is used as a hyperplane to categories data in a 2D space. However, in our case, the 30 features from each texture

patch make it a 30 dimensional data. Hence, a kernel has to be defined that to divide this large dimensional data into two classes (thyroid and non-thyroid). A radial basis function (rbf) kernel was used to achieve this categorization. An accuracy of 89.5% with a sensitivity of 0.896 and specificity of 0.818 were obtained in D1 and an accuracy of 88.7% with a sensitivity of 0.887 and specificity of 0.556 in D2.

Artificial Neural Network (ANN): ANN is an interconnected web of input, hidden and output nodes (or layers) called artificial neurons. The input layer receives the input data which in our case are the 30 different features and these features are passed onto hidden layers which compute several high-level features. These high-level features are passed onto output layer which classifies these features as either thyroid or non-thyroid class. A back-propagation step is used to correct the mistakes made during the prediction phase at the output layer. The correction improves the accuracy of classification of the features. This process is repeated until the best predictions are made. An accuracy of 93.0% with a sensitivity of 0.928 and a specificity of 0.970 was obtained in D1 and an accuracy of 89.4% with a sensitivity of 0.935 and a specificity of 0.535 in D2.

Random Forest Classifier (RFC): RFC is a ensemble supervised learning algorithm that is comprised of a forest of several decision trees. RFC creates a set of decision trees from randomly selected subset of training data and aggregates the votes from each decision tree to decide the final class of the test object. An accuracy of 92.5% with a sensitivity of 0.925 and specificity of 0.866 was obtained in D1 and an accuracy of 0.891 with a sensitivity of 0.935 and a specificity of 0.517 in D2.

The detailed explanation of the implementation of k-means is presented in Chapter C and SVM, RFC and ANN are presented in Chapter E. Naive bayesian, adaptive boosting, LDA and bagging were used as a feasibility analysis for testing the robustness of the extracted features for thyroid texture classification. Hence, the results of Naive Bayesian

and Adaptive Boosting were only presented as a conference paper in Comptuer Assisted Radiology and Surgery (CARS) conference. Similarly, the results of LDA and Bagging were presented in IEEE Engineering in Biology and Medicine (EMB) conference. These papers have been listed in the selected publications section in the beginning of this thesis.

CHAPTER 4

Deep Learning for Thyroid Segmentation

4.1 Introduction to Deep Learning

Deep learning (DL) is a part of machine learning that works mainly based on neural networks. DL can be either of supervised, semi-supervised or unsupervised learning. DL has been used in different fields such as automatic speech recognition, image recognition, visual art pricessing, natural language processing, bioinformatics, medical image analysis, etc. Apart from its usability in numerous applications, segmentation of various anatomical structures has been an important field of research in medical imaging community.

Convolutional Neural Network (CNN) is one of the deep learning approaches that has been used for image segmentation. CNN was originally developed by LeCun *et al.* [35] to recognize hand-written numbers on bank checks. A typical CNN consists of an input layer, multiple hidden layers and an output layer. The hidden layers of a CNN consist of a series of convolutional layers that perform the convolution of input image with different filters. Apart from the convolutional layers, it may contain pooling layers, fully connected layers and normalization layers and activation functions that compute the activations of each neuron/node. The final convolution layers involves back-propagation step which updates the weights of each nodes to achieve a better functioning network.

Computation of hand-crafted features in US images for segmentation tasks is often challenging because of the complexity of different organs as well as the presence of speckle noise and low contrast and SNR. Similarly, it is a very time consuming task to compute robust features. DL allows the computation of these robust and high level features by passing the images through different layers. For example, the lower layers might identify edges, curvatures, etc in the images, while the deeper layers may recognize the objects or anatomical structures as a whole (eg: thyroid, liver, kidney or something else).

4.2   Related Work

There have been many works on segmentation of thyroid and nodules characterization using deep learning methods. Garg *et al.* [24] used a feed-forward neural network to segment thyroid gland in US images. A cascaded CNN involving a U-shaped CNN was used for segmenting nodules in thyroid US images by Xiang *et al.* [69]. Similarly, a cascaded CNN with two deep fully connected CNN was used by Ma *et al.* [37] for automatic detection of thyroid nodules in US images. Liu *et al.* [36] used CNN to compute deep semantic features and combined these features with conventional features such as histogram of oriented gradient and local binary patterns to form a hybrid feature space. These features were later used for classification of thyroid nodules in US images.

On another work by Ma *et al.* [38], a fusion of two CNNs pretrained on natural images from ImageNet database [22] was used for diagnosis of thyroid nodules. A pre-trained GoogLeNet [66] was used for feature extraction in thyroid US images and the extracted features were used later for classification using RFC by Chi *et al.* [19]. Apart from segmenting the thyroid in US images, some works have been done in segmenting thyroid images from other imaging modalities. A work from Ma *et al.* [40] involved distin-

guishing of four different thyroid diseases (hyperthyroidism, hypothyroidism, methylene inflammation and Hashimoto disease) using CNN trained on SPECT images. Another work from Ma *et al.* [39] involved usage of CNN with optimization-based computer-aided diagnosis to diagnose three different thyroid diseases (Graves' disease, Hashimoto disease and subacute thyroiditis) in SPECT images.

4.3   Contributions

As seen in the literature, most of the works have made in classification of thyroid nodules and diagnosis of different types of thyroid diseases. Very limited works have been made in segmenting the thyroid gland as a whole. Hence, in this work, we focussed mainly in segmenting the thyroid gland using two different architectures of CNN. The three network architectures are the U-Net CNN by Ronnberger *et al.* [59] and a fully connected convolutional network (FCNN).

U-Net: U-Net architecture was first used by Ronneberger for segmentation of neuronal structures in electron microscopic tasks and for cell tracking in miscroscopy images. This network can work even with few training images. Hence, it was decided to use this network for segmenting thyroid US images. A U-Net architecture consists of the down-sampling and up-sampling layers that analyse the images by contracting in each successive layers and then expanding in order to produce a full-resolution segmentation respectively.

The down-sampling path consists of 3 x 3 convolution followed by a rectifier linear unit (ReLU) in each layer and then a 2 x 2 max pooling with no stride. The output feature space is doubled in each layer in the down-sampling path and the up-sampling path remaps the lower resolution feature maps to a higher resolution space of the input images. This is done by up-sampling the feature maps followed by a 2 x 2 convolution (up-

convolution) which halves the number of feature channels in each up-sampling step, a concatenation with the corresponding feature map from the down-sampling path and two 3 x 3 convolutions each followed by a ReLU activation. A total of 23 convolutional layers were used. The final activation function predicts the output label for all the pixels in the image (i.e. thyroid or non-thyroid as the output of the network). 'Binary Crossentropy' was set as the loss function and 'adam' optimizer was used for the minimization of the loss function and the network was trained for 20 epochs. The learning rate of the network was set to be 1 x 10^{-5} with a batch size of 32. Data augmentation was carried out by flipping, scaling and varying the lighting condition to the original image datasets (D1 and D2). A dropout of 0.25 was used in both the down-sampling and up-sampling layers. Additionally, all the images were normalized using the Z-score method (i.e. subtracting the mean and then dividing by the standard deviation of the pixel intensities of each patch). The U-net architecture that was implemented in shown in Fig. 4.1.
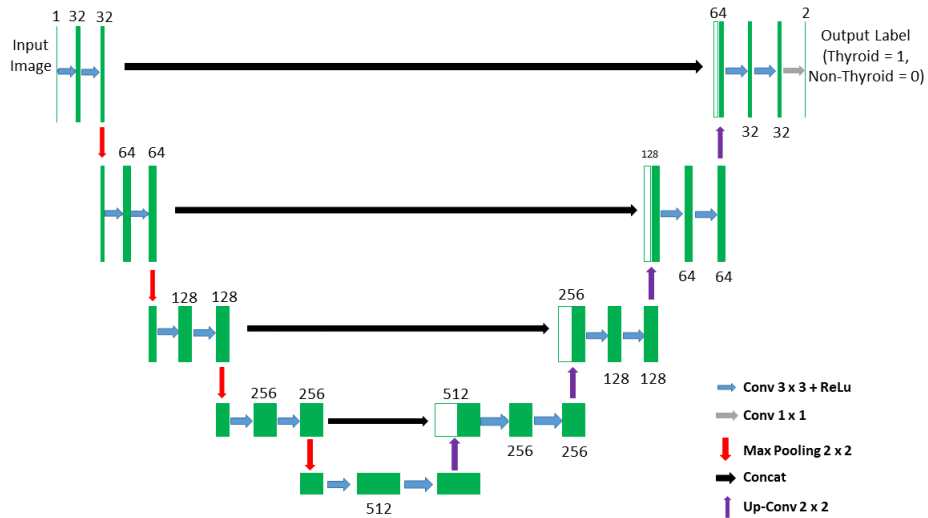
Fig. 4.1: Architecture of the U-Net CNN

The trained CNN classified each pixel in an image as either thyroid (=1) or non-thyroid(=1). The segmented images were undergone a post-processing stage where a largest connected component was chosen and the smaller components were disregarded.

A train-test split of 80% (out of which 80% were used for training and 20% for validation) and 20% was used for the training and testing of the CNN on both the datasets (D1 and D2). An example of thyroid segmentation using U-Net on D1 and D2 is shown in Fig. 4.2. In the figure, each of the first and the second rows represent four different thyroid images with different shapes and sizes from dataset D1 and D2 respectively. With this technique, a test accuracy of 0.872 and 0.896 were obtained in D1 and D2 respectively.
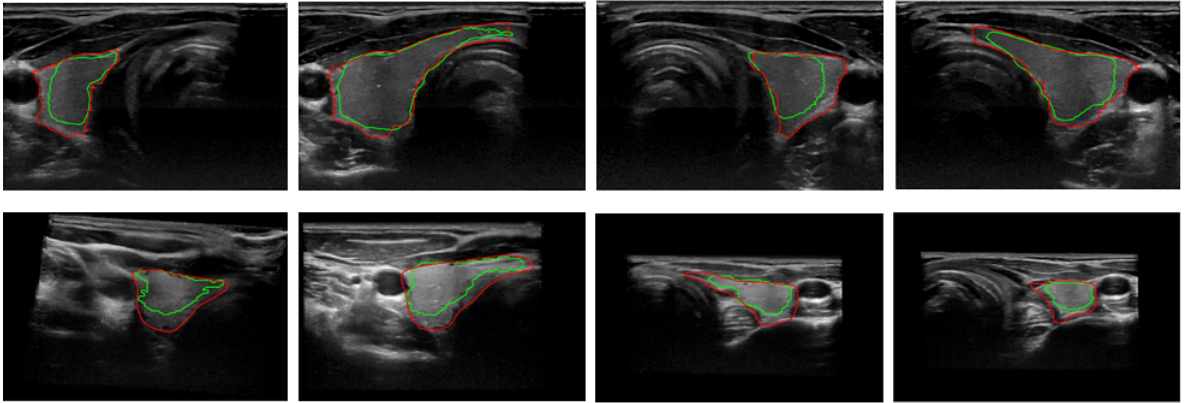
Fig. 4.2: Segmentation of thyroid in US images using U-Net. Green: Ground Truth, Red: U-Net Segmented Thyroid region

Fully Connected Neural Network (FCNN): A deep fully connected neural network was implemented for the classification of thyroid. A different approach was used to train this network. The thyroid US image was divided into smaller texture patches and the patches were assigned a label as thyroid (=1) or non-thyroid (=0). This division of an image into smaller texture patches possess some advantages as compared to training the image as a whole because it allows the network to learn more localized features in a smaller region.

The network consisted of three convolution layers, where the first convolution layer extracted very generalized features like edges, curvatures, pixel intensities, etc. and the deeper layers could extract more sophisticated and high level features that could identify

36

the whole texture patch as a thyroid or non-thyroid. A total of 32 filters with 3 x 3 size were used in the first layer. These filters were doubled in each consecutive layers with the last layer having 128 different filters. The convolution layers were followed by a Leaky Rectified Linear Unit (Leaky ReLU) and max-pooling layer of size 2 x 2 with 0 stride. 'Categorical Crossentropy' was set as the loss function and 'adam' optimizer was used for the minimization of the loss function. The last layer consisted of a two nodes with softmax activation to predict the probability of each pixel in the texture patch being either a thyroid or non-thyroid. Data augmentation was not carried out as we had a big texture patch database (660,250). Similar to U-Net, in order to avoid the problem of over-fitting, a dropout of 0.25 and normalization using Z-score method were used. A detailed explanation of the architecture, training and hyper-parameters selection have been presented in Chapter F.

CHAPTER 5
Discussion and Conclusion

In this thesis, three main approaches for thyroid segmentation were proposed. First, three different image based (data-driven) classic segmentation approaches were used for segmenting the freehand thyroid US datasets. ACWE, GC and PBC methods were used for segmenting the thyroid in a single 2D US image. The segmentation results of one image were used for segmenting the subsequent images in the dataset. The spatial relationship between the corresponding image frames was investigated to extend the segmentation from one image to another without the need of any human interaction. The approaches were compared based on accuracy of segmentation, ease of use (i.e. number of human interaction required), robustness and computation time. Similarly, the approaches were compared with state-of-art methods and it was found that they performed better than two of the approaches in the literature. DC and Hausdorff Distance were used as the performance measure. ACWE performed the best out of the three approaches achieving an average DC of 0.80 and HD of 8.1 mm. Similarly, PBC required the least number of user interaction and also the least computation time. 3D reconstruction of the segmented images was performed using Imfusion and MeVisLab tool and the volume comparison was done with the ground truth. A closest approximation of the volume of thyroid was obtained with ACWE achieving 11.29 cm^3 compared to the ground truth volume 13.56 cm^3. These approaches however failed in segmenting the smaller structures in thyroid (such as isthmus) and were not fully automatic. Hence, two different fully automatic and robust methods (i.e. RFC and U-Net) for segmenting

the 3D thyroid images were also proposed and these methods could segment the thyroid with more accuracy compared to the three approaches. The result obtained with these approaches encouraged us to explore more with the feature extraction combined with machine learning and deep learning approaches.

In the second phase, the problems identified in the first phase were investigated. Hence, machine learning was used for classification and segmentation of thyroid texture using the robust features extracted using two different methods as well as using the physics behind the formation of US images. The first feature extraction technique used the parametrized signal version of thyroid US images to extract 30 different energy based features using AR modelling. The extracted features were used for training eight different machine learning classifiers. In the second study, HOSA was used for extracting 10 different energy, frequency and entropy based features. The extracted features were then used for training a SVM for classifying thyroid texture. In another study, a weak estimate of the thyroid texture is obtained using a statistical mechanics model of US texture formation and a combination of RFC, IRW and GVF was used for segmenting a 3D volume of thyroid by processing individual image frames. The approaches were compared with different approaches in the literature and performed better in terms of DC, sensitivity and specificity. ANN achieved the best DC of 0.930 and specificity of 0.970 and k-means achieved the best sensitivity with 0.950.

Feature extraction from US images is a challenging task since US images suffer from speckle noise and low contrast and SNR. Even though a successful feature extraction was possible using AR modelling and HOSA, it was a very tough task. Hence, a possibility of using deep learning approaches that computes the features by itself was explored in the final phase of the thesis. U-Net architecture and a fully connected CNN were used for segmenting the thyroid in US images. Both the approaches obtained similar accuracy as obtained by the machine learning approaches in Chapter 3. This proves the fact that

the features extracted using AR and HOSA are highly robust to accurately segment the thyroid in US images.

All the work in this thesis revolved around analysis of the healthy thyroid US images. So, a small study was done to work with pathogenic images using U-Net since it does not require the computation of highly sophisticated features. The hyper-parameters like number of epochs, batch size, dropout, learning rate, momentum, etc. were changed to segment the nodules in thyroid images. 17 epochs were used for training the network using the early early stoppage criteria on loss function where the network was stopped from training if the loss did not decrease in five consecutive iterations. The batch size of 16 was used with dropout of 0.2 after each max-pooling layer. Adam optimizer with learning rate of 0.001 and the momentum of 0.9 was used as the loss function. The network was trained on only 453 unhealthy thyroid images and tested on 117 images. Due to a smaller number of images for training the U-Net, an accuracy of 0.612 in terms of DC was obtained. This accuracy could be easily improved by training the network on more images (also including data augmentation) as well as with better optimization of the network. Some examples of nodule segmentation using U-Net have been shown in Fig. 5.1.

Similarly, the machine and deep learning approaches usually require a large amount of data for training and usually it is difficult to acquire these images in medical imaging community because of ethical issues. Hence, to solve the problem of data availability, an approach of computing synthetic thyroid US images was employed using Generative Adversarial Networks (GAN). GAN is a deep learning approach that is based on training of two different neural networks alternatively. A GAN comprises of a generator and discriminator network where the generator tries to generate new or 'fake' images from random input noise as the input and the discriminator tries to identify whether the input image it gets is a real or fake. Hence, the two networks are learning from each
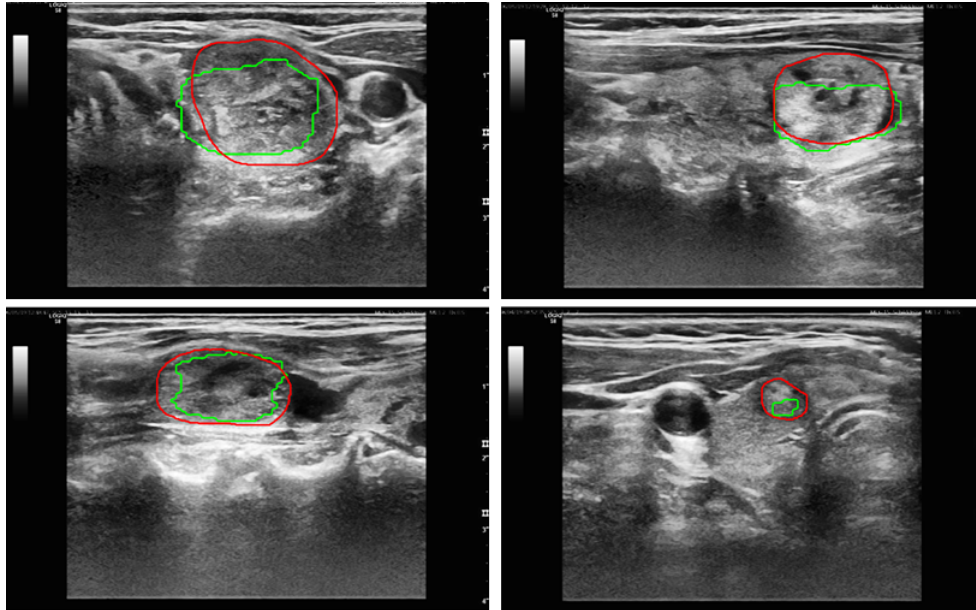
Fig. 5.1: Example of thyroid nodule segmentation using U-Net in four different thyroid Us images. Green: Ground Truth, Red: Segmented Nodules

other. The training goes on until the discriminator is no longer able to distinguish the real and fake images successfully. A comparison between real thyroid US images and images produced using GAN are shown in Fig. 5.2.

The discriminator network consisted of 6 convolution layers with 3 x 3 convolution filters doubling in each subsequent layers starting from 8 and upto 256. Each convolution layer was followed by a batch normalization, dropout of 0.25 and an average pooling. The activation function used was LeakyReLU. Two dense layers were in the output layer with a sigmoid activation. The size of the input images to the discriminator was 256 x 256 x 3. The input to the generator was a noise vector of size 4096 which was upsampled at each convolution layer to create the final US image of size 256 x 256. The filters were halved in each layer starting from 512 and upto 4. A total of eight convolution layers were used with a ReLU activation.

As future works, the features extracted using AR modelling and HOSA could be tested to classify thyroid nodules as benign or malignant. Similarly, more feature extraction

41
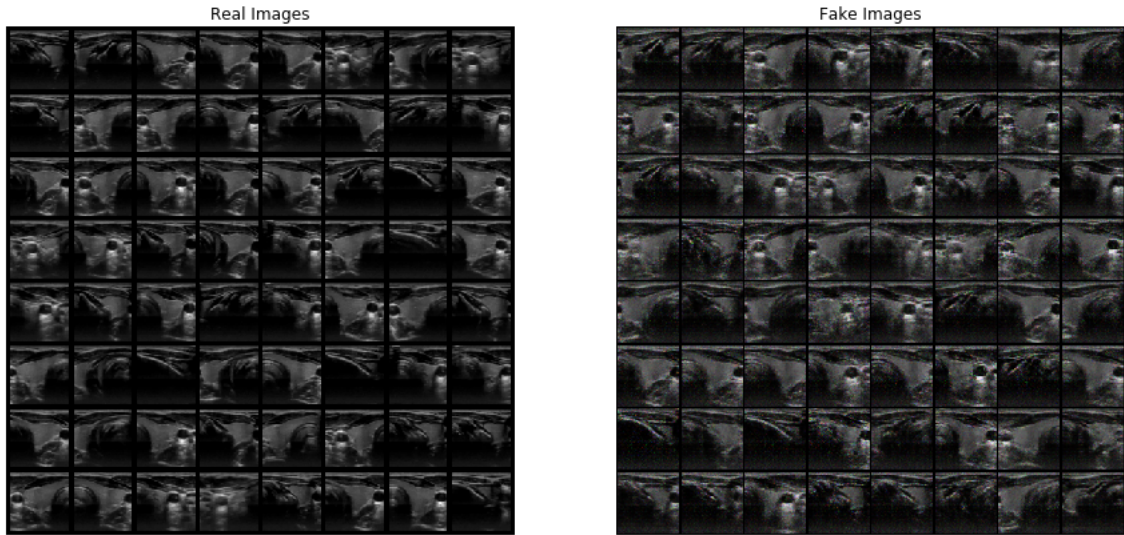
Fig. 5.2: Left: Real Thyroid US Images, Right: Fake Thyroid Images generated using GAN

techniques such as discrete wavelet transform and AR, empirical model decomposition [34], gray level co-occurrence matrix, etc could be used to extract more features. The AR and HOSA features themselves could be combined for thyroid segmentation. Segmentation using DL approaches could be investigated furthermore using the popular networks like AlexNet, GoogLeNet, CaffeNet, etc. Additionally, DL approaches could be used for feature extraction and these features along with the AR and HOSA features could be used for training different ML classifiers for thyroid segmentation. These works would increase the clinical relevance of thyroid segmentation even more.

Bibliography

[1] Imfusion gmbh. "https://www.mevislab.de/". 22

[2] Mevislab. "https://www.imfusion.com/". 22

[3] Surgiceye gmbh. "https://www.surgiceye.com/". 17

[4] Thyroid. https://www.webmd.com/women/picture-of-the-thyroid. xii, 9

[5] Thyroid. http://www.endocrineweb.com/conditions/thyroid/how-your-thyroid-works. 9, 10

[6] Weka. "https://www.cs.waikato.ac.nz/ml/weka/". 28

[7] U. R. Acharya, P. Chowrippa, H. Fujita, S. Bhat, S. Dua, j. E. Koh, L. Eugene, P. Kongmebhol, and K.-H. Ng. Thyroid lesion classification in 242 patient population using gabor transform features from high resolution ultrasound images. *Knowledge-Based Systems*, 107:235–245, 2016. 25

[8] U. R. Acharya, O. Faust, S. V. Sree, F. Molinari, R. Garberoglio, and J. Suri. Cost-effective and non-invasive automated benign & malignant thyroid lesion classification in 3d contrast-enhanced ultrasound using combination of wavelets and textures: a class of thyroscan algorithms. *Technology in cancer research & treatment*, 10(4):371–380, 2011. 24, 25

[9] U. R. Acharya, O. Faust, S. V. Sree, F. Molinari, and J. S. Suri. Thyroscreen system: high resolution ultrasound thyroid image characterization into benign and

malignant classic using novel combination of texture and discrete wavelet transform. *Computer methods and programs in biomedicine*, 107(2):233–241, 2012. 25

[10] U. R. Acharya, S. V. Sree, M. M. R. Krishnan, F. Molinari, R. Garberoglio, and J. S. Suri. Non-invasive automated 3d thyroid lesion classification in ultrasound: a classic of thyroscan systems. *Ultrasonics*, 52(4):508–520, 2012. 25

[11] U. R. Acharya, S. V. Sree, G. Swapna, S. Gupta, F. Molinari, R. Garberoglio, A. Witkowska, and K. S. Suri. Effect of complex wavelet transform filter on thyroid tumor classification in three-dimensional ultrasound. *Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine*, 227(3):284–292, 2013. 25

[12] A. A. Ardakani, A. Gharbali, and A. Mohammadi. Classification of benign and malignant thyroid nodules using wavelet texture analysis of sonograms. *Journal of Ultrasound in Medicine*, 34(11):1983–1989, 2015. 25

[13] W. G. Bradley. History of medical imaging. *Proceedings of the American Philosophical Society*, 152(3):349–361, 2008. 8

[14] G. Castellano, L. Bonilha, L. Li, and F. Cendes. Texture analysis of medical images. *Clinical radiology*, 59(12):1061–1069, 2004. 24

[15] T. F. Chan and L. A. Vese. Active contours without edges. *IEEE Transactions on image processing*, 10(2):266–277, 2001. 21

[16] C.-Y. Chang, Y.-F. Lei, C.-H. Tseng, and S.-R. Shih. Thyroid segmentation and volume estimation in ultrasound images. *IEEE transactions on biomedical engineering*, 57(6):1348–1357, 2010. 21, 22

[17] C.-Y. Chang, Y.-F. Lei, C.-H. Tseng, and S.-R. Shih. Thyroid segmentation and volume estimation in ultrasound images. *IEEE transactions on biomedical engineering*, 57(6):1348–1357, 2010. 25

[18] V. Chaudhary and S. Bano. Imaging of the thyroid: Recent advances. *Indian journal of endocrinology and metabolism*, 16(3):371, 2012. 13, 16

[19] J. Chi, E. Walia, P. Babyn, J. Wang, G. Groot, and M. Eramian. Thyroid nodule classification in ultrasound images by fine-tuning convolutional neural network. *Journal of digital imaging*, 30(4):477–486, 2017. 33

[20] D. China, A. Illanes, P. Poudel, M. Friebe, P. Mitra, and D. Sheet. Anatomical structure segmentation in ultrasound volumes using cross frame belief propagatng iterative random walks. *IEEE journal of biomedical and health informatics*, 23(3):1110–1118, 2018. 25

[21] R. Chung and D. Kim. Imaging of thyroid nodules. *Applied Radiology*, 48(1):16–26, 2019.

[22] J. Deng, W. Dong, R. Socher, L. J. Li, K. Li, and L. Fei-Fei. Imagenet: A large-scale hierarchical image database. In *CVPR09*, 2009. 33

[23] J. Dornheim, L. Dornheim, B. Preim, K. D. Tönnies, I. Hertel, and G. Strauss. Stable 3d mass-spring models for the segmentation of the thyroid cartilage. *Otto-von-Guericke-Universität Magdeburg. Germany*, 2005. 20

[24] H. Garg and A. Jindal. Segmentation of thyroid gland in ultrasound image using neural network. In *2013 Fourth International Conference on Computing, Communications and Networking Technologies (ICCCNT)*, pages 1–5, 2013. 33

[25] R. Gir, L. Jain, and R. Rai. Speckle reduction of synthetic aperture radar images using median filter and savitzky-golay filter. *International Journal of Computer Applications*, 113(11), 2015. 11, 24

[26] T. E. Hall and G. B. Giannakis. Bispectral analysis and model validation of texture images. *IEEE Transactions on Image Processing*, 4(7):996–1009, 1995. 11, 28

[27] J. K. Hoang, B. F. Branstetter IV, A. R. Gafton, W. K. Lee, and C. M. Glastonbury. Imaging of thyroid carcinoma with ct and mri: approaches to common scenarios. *Cancer Imaging*, 13(1):128, 2013. 14

[28] D. K. Iakovidis, E. G. Keramidas, and D. Maroulis. Fusion of fuzzy statistical distributions for classification of thyroid ultrasound patterns. *Artificial Intelligence in Medicine*, 50(1):33–41, 2010. 25

[29] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. *International journal of computer vision*, 1(4):321–331, 1988. 21

[30] J. Kaur and A. Jindal. Comparison of thyroid segmentation algorithms in ultrasound and scintigraphy images. *International Journal of Computer Applications*, 50(23), 2012. 19

[31] E. G. Keramidas, D. K. Iakovidis, D. Maroulis, and S. Karkanis. Efficient and effective ultrasound image analysis scheme for thyroid nodule detection. In *International Conference Image Analysis and Recognition*, pages 1052–1060. Springer, 2007. 25

[32] E. N. Kollorz, D. A. Hahn, R. Linke, T. W. Goecke, J. Hornegger, and T. Kuwert. Quantification of thyroid volume using 3-d ultrasound imaging. *IEEE Transactions on medical imaging*, 27(4):457–466, 2008. 20

[33] R. Koprowski, A. Korzyńska, Z. Wróbel, W. Zieleźnik, A. Witkowska, J. Małyszek, and W. Wójcik. Influence of the measurement method of features in ultrasound images of the thyroid in the diagnosis of hashimoto's disease. *Biomedical engineering online*, 11(1):91, 2012. 25

[34] a. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012. 42

[35] Y. LeCun, B. E. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. E. Hubbard, and L. D. Jackel. Handwritten digit recognition with a back-propagation network. In *Advances in neural information processing systems*, pages 396–404, 1990. 32

[36] T. Liu, S. Xie, J. Yu, L. Niu, and W. Sun. Classification of thyroid nodules in ultrasound images using deep model based transfer learning and hybrid features. In *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 919–923, 2017. 33

[37] J. Ma, F. Wu, J. Zhu, D. Kong, et al. Cascade convolutional neural networks for automatic detection of thyroid nodules in ultrasound images. *Medical Physics*, 44(5):1678–1691, 2017. 33

[38] J. Ma, F. Wu, J. Zhu, D. Xu, and D. Kong. A pre-trained convolutional neural network based method for thyroid nodule diagnosis. *Ultrasonics*, 73:221–230, 2017. 33

[39] L. Ma, C. Ma, Y. Liu, and X. Wang. Thyroid diagnosis from spect images using convolutional neural network with optimization. *Computational intelligence and neuroscience*, 2019. 34

[40] L. Ma, C. Ma, Y. Liu, X. Wang, and W. Xie. Diagnosis of thyroid diseases using spect images based on convolutional neural network. *Journal of Medical Imaging and Health Informatics*, 8(8):1684–1689, 2018. 33

[41] N. H. Mahmood and A. H. Rusli. Segmentation and area measurement for thyroid ultrasound image. *International Journal of Scientific & Engineering Research*, 2(12):1–8, 2011. 19

[42] D. G. Manolakis, V. K. Ingle, S. M. Kogon, et al. Statistical and adaptive signal processing: spectral estimation, signal modeling, adaptive filtering, and array processing. 2000. 27

[43] D. E. Maroulis, M. A. Savelonas, D. K. Iakovidis, S. A. Karkanis, and N. Dimitropoulos. Variable background active contour model for computer-aided delineation of nodules in thyroid ultrasound images. *IEEE Transactions on Information Technology in Biomedicine*, 11(5):537–543, 2007. 20

[44] A. Materka, M. Strzelecki, et al. Texture analysis methods – a review. *Technical university of lodz, institute of electronics, COST B11 report, Brussels*, pages 9–11, 1998. 28

[45] M. Mirmehdi. *Handbook of texture analysis*. Imperial College Press, 2008. 24, 25

[46] M. Müller, D. Lorenz, I. Zuna, W. Lorenz, et al. The value of computer-assisted sonographic tissue characterization in focal lesions of the thyroid. *Der Radiologe*, 29(3):132–136, 1989. 25

[47] E. A. Mylona, M. A. Savelonas, and D. Maroulis. Automated adjustment of region-based active contour parameters using local image geometry. *IEEE transactions on cybernetics*, 44(12):2757–2770, 2014. 20

[48] E. A. Mylona, M. A. Savelonas, and D. Maroulis. Self-parameterized active contours based on regional edge structure for medical image segmentation. *SpringerPlus*, 3(1):424, 2014. 20

[49] N. S. Narayan, P. Marziliano, J. Kanagalingam, and C. G. Hobbs. Speckle patch similarity for echogenicity-based multiorgan segmentation in ultrasound images of the thyroid gland. *IEEE journal of biomedical and health informatics*, 21(1):172–183, 2015. 26

[50] M. Nasrolahzadeh, Z. Mohammadpoory, and J. Haddadnia. Higher-order spectral analysis of spontaneous speech signals in alzeimer's disease. *Cognitive neurodynamics*, 12(6):583–596, 2018. 11, 28

[51] M. Nusynowitz. Thyroid imaging. *Lippincott's primary care practice*, 3(6):546–55, 1999. 13, 14

[52] A. Osman. *Automated evaluation of three dimensional ultrasonic datasets*. PhD thesis, INSA de Lyon, 2013. 20

[53] D. L. Pham, C. Xu, and J. L. Prince. A survey of current methods in medical image segmentation. *Annual Review of Biomedical Engineering*, January 1998. 10

[54] D. L. Pham, C. Xu, and J. L. Prince. A survey of current methods in medical image segmentation. *Annual Review of Biomedical Engineering*, January 1998. 19

[55] P. Poudel, A. Illanes, D. Sheet, and M. Friebe. Evaluation of commonly used algorithms for thyroid ultrasound images segmentation and improvement using machine learning approaches. *Journal of healthcare engineering*, 2018, 2018. 20

[56] U. Raghavendra, U. R. Acharya, A. Gudigar, J. H. Tan, H. Fujita, Y. Hagiwara, F. Molinari, P. Kongmebhol, and K. H. Ng. Fusion of spatial gray level dependency

and fractal texture features for the characterization of thyroid lesions. *Ultrasonics*, 77:110–120, 2017. 25

[57] U. Raghavendra, A. Gudigar, M. Maithri, A. Gertych, K. M. Meiburger, C. H. Yeong, C. Madla, P. Kongmebhol, F. Molinari, K. H. Ng, et al. Optimized multi-level elongated quinary patterns for the assessment of thyroid nodules in ultrasound images. *Computers in biology and medicine*, 95:55–62, 2018. 25

[58] C. C. Reading and C. A. Gorman. Thyroid imaging techniques. *Clinics in laboratory medicine*, 13(3):711–724, 1993. 13

[59] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomediacl image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015. 34

[60] M. A. Savelonas, D. K. Iakovidis, I. Legakis, and D. Maroulis. Active contours guided by echogenicity and texture for delineation of thyroid nodules in ultrasound images. *IEEE Transactions on Information Technology in Biomedicine*, 13(4):519–527, 2008. 20

[61] D. Selvathi and V. Sharnitha. Thyroid classification and segmentation in ultrasound images using machine learning algorithms. pages 836–841. IEEE, 2011. 25

[62] A. Sheeja Agustin, S. S. Babu, and K. T. Nadu. Thyroid segmentation on us medical images: An overview. 2012. 19

[63] N. Singh and A. Jindal. A segmentation method and comparison of classification methods for thyroid ultrasound images. *International Journal of Computer Applications*, 50(11):43–49, 2012. 26

[64] A. Stefano, S. Vitabile, G. Russo, M. Ippolito, M. G. Sabini, D. Sardina, O. Gambino, R. Pirrone, E. Ardizzone, and M. C. Gilardi. An enhanced random walk

algorithm for delineation of head and neck cancers in pet studies. *Medical & biological engineering & computing*, 55(6):897–908, 2017. 19

[65] A. Swami, G. B. Giannakis, and J. M. Mendel. Linear modeling of multidimensional non-gaussian processes using cumulants. *Multidimensional systems and signal processing*, 1(1):11–37, 1990. 28

[66] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1–9, 2015. 33

[67] A. M. Tekalp and A. T. Erdem. Two-dimensional higher-order spectrum factorization with application in nongaussian image modeling. In *Workshop on Higher-Order Spectral Analysis*, pages 186–190. IEEE, 1989. 28

[68] S. Tsantis, N. Dimitropoulos, D. Cavouras, and G. Nikiforidis. Morphological and wavelet features towards sonographic thyroid nodules evaluation. *Computerized Medical Imaging and Graphics*, 33(2):91–99, 2009. 26

[69] X. Ying, Z. Yu, R. Yu, X. Li, M. Yu, M. Zhao, and K. Liu. Thyroid nodule segmentation in ultrasound images based on cascaded convolutional neural network. In *International Conference on Neural Information Processing*, pages 373–384, 2018. 33

[70] J. Zhao, W. Zheng, L. Zhang, and H. Tian. Segmentation of ultrasound images of thyroid nodule for assisting fine needle aspiration cytology. *Health information science and systems*, 1(1):5, 2013. 19

CHAPTER A

Evaluation of commonly used algorithms for thyroid ultrasound images segmentation and improvement using machine learning approaches

Research Article

Evaluation of Commonly Used Algorithms for Thyroid Ultrasound Images Segmentation and Improvement Using Machine Learning Approaches

Prabal Poudel [iD],[1] **Alfredo Illanes** [iD],[1] **Debdoot Sheet,**[2] **and Michael Friebe**[1]

[1]*Department of Medical Engineering, Otto-von-Guericke University Magdeburg, 39106 Magdeburg, Germany*
[2]*Department of Mechanical Engineering, Indian Institute of Technology Kharagpur, 822113 Kharagpur, India*

Correspondence should be addressed to Prabal Poudel; prabal.poudel@ovgu.de

The thyroid is one of the largest endocrine glands in the human body, which is involved in several body mechanisms like controlling protein synthesis and the body's sensitivity to other hormones and use of energy sources. Hence, it is of prime importance to track the shape and size of thyroid over time in order to evaluate its state. Thyroid segmentation and volume computation are important tools that can be used for thyroid state tracking assessment. Most of the proposed approaches are not automatic and require long time to correctly segment the thyroid. In this work, we compare three different nonautomatic segmentation algorithms (i.e., active contours without edges, graph cut, and pixel-based classifier) in freehand three-dimensional ultrasound imaging in terms of accuracy, robustness, ease of use, level of human interaction required, and computation time. We figured out that these methods lack automation and machine intelligence and are not highly accurate. Hence, we implemented two machine learning approaches (i.e., random forest and convolutional neural network) to improve the accuracy of segmentation as well as provide automation. This comparative study intends to discuss and analyse the advantages and disadvantages of different algorithms. In the last step, the volume of the thyroid is computed using the segmentation results, and the performance analysis of all the algorithms is carried out by comparing the segmentation results with the ground truth.

Abstract

The thyroid is one of the largest endocrine glands in the human body, which is involved in several body mechanisms like controlling protein synthesis and the body's sensitivity to other hormones and use of energy sources. Hence, it is of prime importance to track the shape and size of thyroid over time in order to evaluate its state. Thyroid segmentation and volume computation are important tools that can be used for thyroid state tracking assessment. Most of the proposed approaches are not automatic and require long time to correctly segment the thyroid. In this work, we compare three different non-automatic segmentation algorithms (i.e., active contours without edges, graph cut, and pixel-based classifier) in freehand three-dimensional ultrasound imaging in terms of accuracy, robustness, ease of use, level of human interaction required, and computation time. We figured out that these methods lack automation and machine intelligence and are not highly accurate. Hence, we implemented two machine learning approaches (i.e., random forest and convolutional neural network) to improve the accuracy of segmentation as well as provide automation. This comparative study intends to discuss and analyse the advantages and disadvantages of different algorithms. In the last step, the volume of the thyroid is computed using the segmentation results, and the performance analysis of all the algorithms is carried out by comparing the segmentation results with the ground truth.

A.1 Introduction

The segmentation and volume computation of thyroid are of prime importance when it comes to the diagnosis and treatment of thyroid diseases. Thyroid is a butterfly-shaped gland located below the Adam's apple on the front of the neck. Most of the thyroid diseases like Graves' (excessive production of thyroid hormones), subacute thyroiditis

53

(inflammation of thyroid), thyroid cancer, goitre (thyroid swelling), and thyroid nodule (small abnormal lump growths in thyroid) involve changes in the shape and size of thyroid [1]. Hence, it is essential to compute the volume of thyroid over time to identify whether the thyroid is healthy or not. We use ultrasound (US) imaging for data acquisition instead of other imaging modalities as it is much safer and painless when used on the patients.

Several research works have been proposed on how to segment the thyroid in individual 2D US images. Zhao *et al.* [2] proposed several approaches (edge detection, method of threshold value, region splitting and merging, watershed segmentation, active contour, graph theory, US image segmentation based on Ncut, and segmentation based on improved normalized cut) based on 2D segmentation of thyroid in US images. Kaur and Jindal [3] segmented thyroid from 2D US and scintigraphy images using active contour without edges, localized region-based active contour, and distance regularized level set. Augustin *et al.* [4] tested and segmented thyroid US images using fuzzy c-means algorithm, histogram clustering, QUAD tree, region growing, and random walk [5]. A polynomial Support Vector Machine (SVM) was used [6] to segment the thyroid gland in US images. A local region-based active contour was proposed [7] to segment and compute the area of segmented thyroid in a 2D US image. Another region-based active-contour implementation to segment medical images was carried out by Mylona *et al.* [8, 9] where they encoded the local geometry information (i.e., orientation coherence in the edges of the regions to be segmented) to control the evolution of the contour. Similarly, thyroid segmentation in US images using a novel boundary detection method and local binary patterns for texture analysis was proposed by Keramidas *et al.* [10]. Level-set active-contours models for thyroid segmentation in US images were used in [11, 12]. These methods were mainly based on variable background active contour and joint echogenicity texture. Garg and Jindal later worked on feedforward neural network

[13] to segment thyroid gland from US images. Recently, Narayan *et al.* [14] made use of the speckle-related pixels and imaging artefacts as source of information to perform multi-organ (thyroid, carotid artery, muscles, and trachea) segmentation in thyroid US images.

Similarly, several research works have been carried out to segment a full 3D thyroid image. Kollorz *et al.* [15] proposed a semi-automated approach to classify thyroid gland for volumetric quantification using geodesic active contour. Chang *et al.* [16] proposed radial basis function (RBF) neural network to segment the blocks of thyroid gland. 3D mass spring models for thyroid cartilage segmentation by creating a 3D deformable shape models were proposed by Dornheim *et al.* [17] but on computed tomography (CT) images. A complete segmentation and analysis of 3D thyroid images was carried out by Osman [18] by thresholding the voxel intensities and then connecting similar voxels to predict the segmenting regions.

The aforementioned approaches have limitations in the sense that they work either on a single 2D image or on a whole 3D image and they do not make use of the spatial relationship between the neighbouring US slices. Hence, we propose three widely used segmentation algorithms which usually work on a 2D image but can be extended to segment a sequence of freehand US images by making use of the spatial relationships between the corresponding image frames. These three approaches are based on active contours without edges (ACWE), graph cut (GC), and pixel-based classifier (PBC). In case of ACWE, the centroid of the segmented image is used as the priori information to find the location of contour initialization in the corresponding slices. GC allows the user to select the foreground and background areas in one image, and these information are transferred to the corresponding slices for further initialization. In PBC, the user clicks inside and outside of the thyroid region to extract the features for thyroid segmentation which are used to train the decision trees and later to classify thyroid and nonthyroid

regions in the corresponding images. Our approach of segmenting individual slices and then reconstructing them to a volume possess greater advantages than segmenting directly on the volume itself as segmenting on the 3D image is very complex and is difficult to control. Also, segmentation in 2D allows to analyse the shape of the thyroid in detail as compared to segmenting directly on the thyroid volume.

The main purpose of this work is to compare three nonautomatic segmentation techniques, which are based on ACWE [19], GC [20], and PBC [16] to perform the segmentation in the thyroid images. They are compared based on their accuracy, robustness, ease of use, and computation time. We also compare the results of these approaches to some of the existing methods [17] that use mass spring models. These algorithms were chosen over others as they can be used not only on one image but also on a sequence of US images in a dataset to produce a 3D segmented thyroid as the information from a segmented image could be transferred to the corresponding image slices to segment them. Additionally, when the segmentation is ongoing in different images in a dataset, the user could directly interact with the segmentation results and correct them if there are any under- or oversegmentation. After segmentation, the segmented images are later used for 3D reconstruction and volume computation using ImFusion [21] and MeVis-Lab [22]. We figured out that the nonautomatic methods pose several disadvantages and thus implemented two automatic machine learning based methods such as Random Forest Classifier (RFC) and Convolutional Neural Network (CNN) and compared their performance with the nonautomatic methods. We came to the conclusion that the commonly used algorithms could not segment a series of US images highly accurately as compared with these supervised learning techniques.

A.2   Materials and Methods

In this section, we will explain the three non-automatic as well as the two automatic methods that are compared in this work to segment the thyroid glands in US images. The automatic methods use 3D thyroid images while the non-automatic methods use 2D images. We will also present the 3D reconstruction (using segmented results from non-automatic methods) as well as volume computation technique.

A.2.1   Active Contour without Edges

Preprocessing

ACWE segmentation was followed by a preprocessing step as the algorithm mainly worked on the gradient information for contour evolution and the preprocessing step improved the gradient visualization. US images mainly contain speckle noise [23] and have low contrast [16]. The speckle noise is produced by the interference of the returning ultrasound waves at the transducer aperture as the ultrasound images are produced when the reflected sound waves from different surfaces inside the body are picked up by the transducer. To enhance the contrast and suppress the speckle noise, a preprocessing step is carried out. Contrast enhancement [24] is used to improve the visibility of the thyroid region. In this work, we make use of Histogram equalization technique which is one of the methods used in contrast enhancement. It helps in recovering the lost contrast in the image by remapping the brightness values such that they are distributed over all the pixels. After histogram equalization, median filter is applied to suppress the speckle noise. It not only reduces speckles but also preserves the gradient/edge information.

Segmentation

After preprocessing, the segmentation process was carried out using the level-set approach developed by Chan and Vese [19]. It is based on the minimization of the Mumford-Shah functional and involves four main steps.

In the first step, the user starts by initializing a rectangle/square mask around the region to be segmented. The initialization of the mask (Fig. A.1) is a very important step in this algorithm as a wrong initialization can lead to the segmentation of unnecessary segments inside the image. The initial mask separates the image into two regions: foreground (=1) which is inside of the mask and the background (=0) which is outside of the mask.
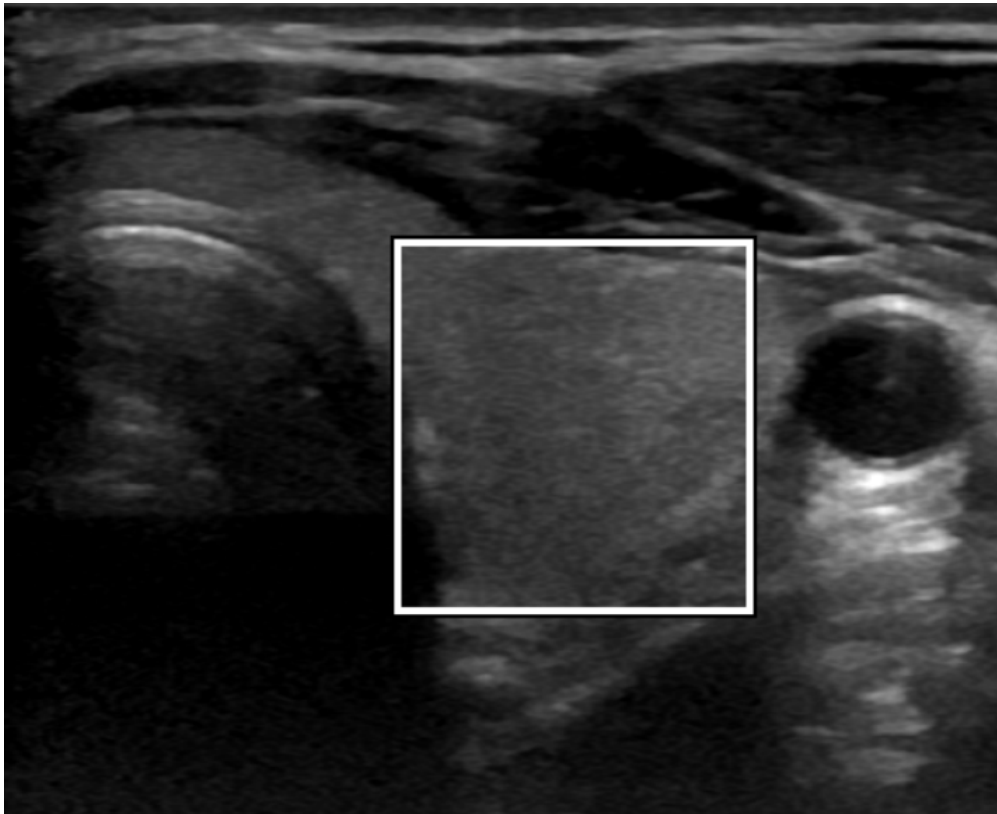
Fig. A.1: ACWE initialization of the mask by the user

In the second stage, a Signed Distance Function (SDF), ϕ is computed from the initial

mask (C) by using Euclidean Distance. SDF is one of the methods of representing the level sets which are used to keep track of evolving curve over time. Our goal is to evolve $\phi(x, y)$ when the evolving contour (C) is the zero level set of $\phi(x, y, t)$ at each time t.

Thirdly, the forces that control the evolution of the initial contour are computed. These forces are: force from the image and force from curvature. Hence they are calculated as following:

$$F_{image} = \int_{insideC} (I - \mu_{in})^2 + \int_{outsideC} (I - \mu_{out})^2 \tag{A.1}$$

where, I is the image, μ_{in} is the average inside the contour and μ_{out} is the average outside the contour.

$$F_{curvature} = \frac{\phi_x^2 * \phi_{yy} + \phi_y^2 * \phi_{xx} - 2\phi_x\phi_y\phi_{xy}}{(\phi_x^2 + \phi_y^2)^{\frac{3}{2}}} \tag{A.2}$$

All the derivatives of ϕ are computed using central difference method. Using these two forces, the evolution of the curve is computed using the Taylor expansion given by the following equation:

$$\phi((x, y), t + \triangle t) = \triangle t * \phi_t + \phi((x, y), t) \tag{A.3}$$

where,

$$\phi_t = \alpha * F_{curvature} + \frac{F_{image}}{max|F_{image}|} \tag{A.4}$$

$$\triangle t = \frac{1}{max(\phi_t) + \epsilon} \tag{A.5}$$

59

where, α represents the smoothing term and ϵ represents the coefficient to satisfy Courant, Friedrichs, Lewy (CFL) condition [14]. The evolution of the contour stops after the given number of iterations are complete.
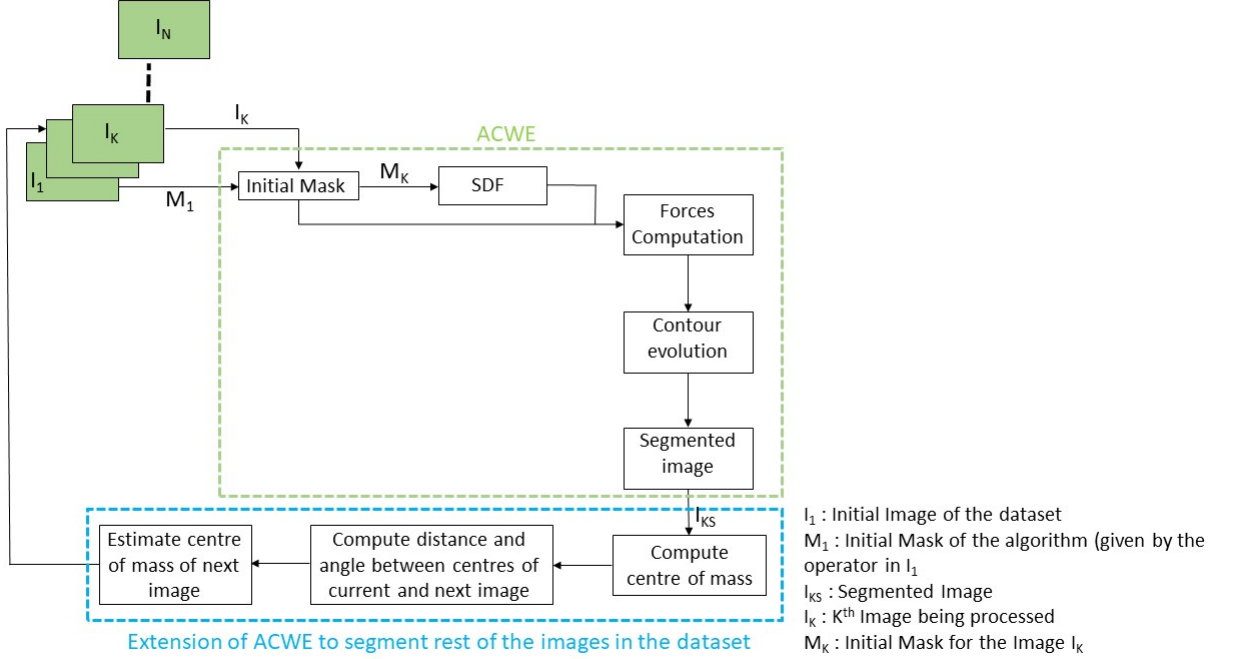
Fig. A.2: Schematic description of ACWE segmentation method

In the last stage, the result of the segmentation on the first image of the dataset is used to segment rest of the images in the dataset. After the segmented thyroid is obtained, its centre of mass is computed. This centre of mass is used to find probable centre of mass of the thyroid in next image slice.

It is computed by making use of the tracking matrices obtained during the data acquisition phase. Each image in the dataset has an associated tracking matrix which gives the transformation from the origin of electromagnetic (EM) tracking system to the centre of the image. Hence, the centre of each image can be computed using the transformation matrix which has the information about the centre of each image in the dataset.

Using this information, the Euclidean distance between the image centres of the current and the next image is computed and the angle between the centres is computed. After computing the distance and angle between the two image centres, a probable centre of mass of the thyroid in the next image is computed by traversing the same distance and angle from the centre of mass of the current segmented thyroid [26]. Centre of mass computed this way will serve to be the centre of rectangle in the next image frame around which the new mask will be initialized automatically. The size of the rectangle will be the same as it was drawn by the user in the first image. In this way, the automatic initialization of segmentation mask is done in the consecutive image frames which will undergo the ACWE algorithm to produce segmented thyroids. The schematic description of the approach is presented in Fig. A.2.

A fixed number of iterations is set by the user for the contour evolution. By increasing the number of iterations, the computation time will be higher. Hence, a trade-off between the accuracy and computation time has to be maintained while running this algorithm.

A.2.2   Graph Cut

This approach makes use of the GrabCut algorithm from Rother et al. [20]. It is also a semi-automatic 2D segmentation algorithm just like the ACWE as the user needs to mark the regions as being thyroid and nonthyroid in the initialization phase. It starts with the user creating an initial trimap by marking the thyroid region to be segmented by using yellow scribbles and the surrounding (i.e., nonthyroid) regions by using violet scribbles as seen in Fig. A.3. The pixels outside of the violet scribble are marked as known background, pixels inside of the violet scribble are marked as unknown, and the yellow scribble areas are marked as definite foreground. The schematic description of

the approach can be seen in Fig. A.4.
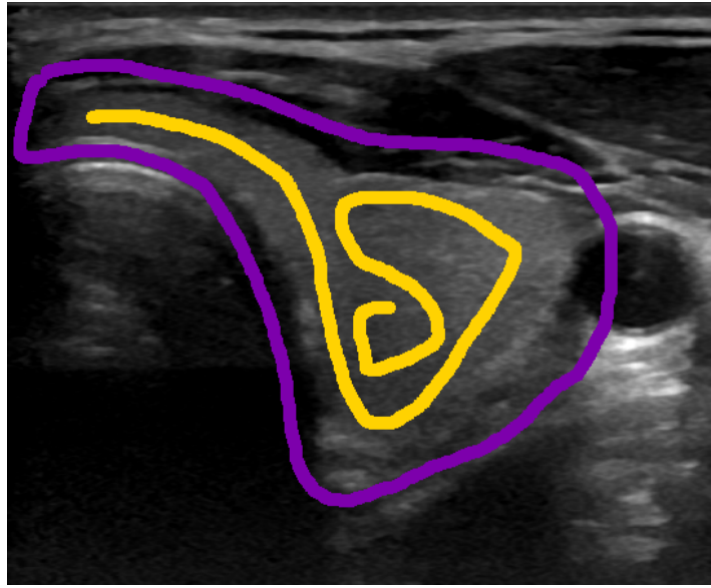
Fig. A.3: Graph cut initialization of the user

Then, an initial image segmentation is computed where all the unknown pixels are placed in the foreground class and all the known background pixels are placed in the background class. These initial foreground and background classes are used to construct foreground and background Gaussian Mixture Models (GMMs) using the Orchard-Bouman clustering algorithm [27]. Each pixel in the foreground class is assigned to the most likely Gaussian component in the foreground GMM, and similarly, each pixel in the background class is assigned to the most likely background Gaussian component. With the new distribution of the pixels, the initial GMMs are disregarded and new GMMs are learned from the pixel distributions in each of the two classes.

Finally, a graph is built which consists of each pixel as node and two special nodes (i.e., foreground and background). All of these nodes are connected by two types of edges (also called as links). The first link (i.e., N-link) connects a pixel to its 8-neighbourhood pixels. These links describe the penalty for placing a segmentation boundary between the neighbouring pixels. The second link (i.e., T-link) connects each pixel to the fore-

ground and background nodes. Each of these links has a weight which represents the probability of a pixel belonging to either a foreground or a background. These probabilities are computed in the GMM models and updated in each iteration until a convergence is reached to get a segmented thyroid. The weight of the N-links between pixel m and its 8-neighbourhood pixels, n , is computed as

$$N(m,n) = \frac{\gamma}{dist(m,n)} e^{-\beta||z_m - z_n||^2} \tag{A.6}$$

where z_m is the color of pixel m, $\gamma = 50$ as suggested by Blake *et al.* [28], and β is given as following by Boykov and Jolly [29]:

$$\beta = \frac{1}{2 < ||z_m - z_n||^2 >} \tag{A.7}$$
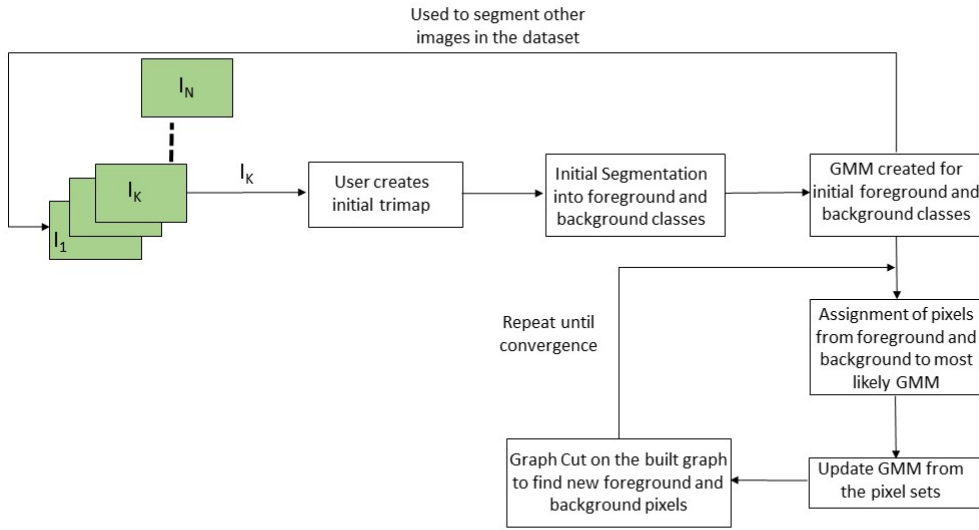
Fig. A.4: Schematic description of GC segmentation method

The initial user initializations are interpolated in the corresponding slices to mark the thyroid and nonthyroid regions (i.e., foreground and background) and create corresponding GMMs. The aforementioned processes are then repeated in each individual images to segment all the thyroid in the dataset. The advantage of this algorithm over ACWE is that it is much faster than ACWE and the user can interact with the result of the segmentation (i.e., postsegmentation) and correct if any errors are present. The results of the segmentation from GC in all the 2D images are used to reconstruct the 3D thyroid by using MeVisLab [22]. The 3D model is updated as soon as the user tries to improve the segmentation results by further interaction in the segmented images. Thus, the accuracy of the algorithm is directly proportional to the number of user interactions on the segmented images.

The increased number of user interactions adds to the computation time of the algorithm. Hence, an optimum number of user interaction should be chosen to obtain the best segmentation results with minimum user interaction. For this purpose, the user interaction in every 10 slices or every 2mm was proposed.

A.2.3 Pixel-Based Classifier

This approach is based on training the decision trees by using different features computed from the images. In this work, three image features are computed. The selection of the features is based on the work of Chang *et al.* [16]. The first feature that is computed is the coefficient of variation $C_v = \frac{\sigma}{\mu}$ where σ means the standard deviation and μ is the mean of the user selected region during the initialization process. This coefficient is computed in two different sized neighbourhoods (i.e., 4-neighbours and 8-neighbours) of every pixel, thus resulting in two features. The third feature that is computed is the mean of the smaller of the two neighbourhoods. So, the first two features are the coef-

ficient of variation at two different sized neighbourhoods of every pixel, and the third feature is the mean of the smaller of the two neighbourhoods.

The algorithm starts by the user clicking on the inside and outside of the thyroid in several thyroid images from where the features are computed which are then passed as training input for the decision trees. The trained trees later classify the different regions in the image as thyroid or non-thyroid. After segmentation, the user can click in more regions to improve the segmentation results. However, selection of wrong thyroid regions for training the decision trees might result in over-segmentation. So, the user should carefully select the thyroid regions. The presented approach is shown in the schematic diagram in Fig. A.5.
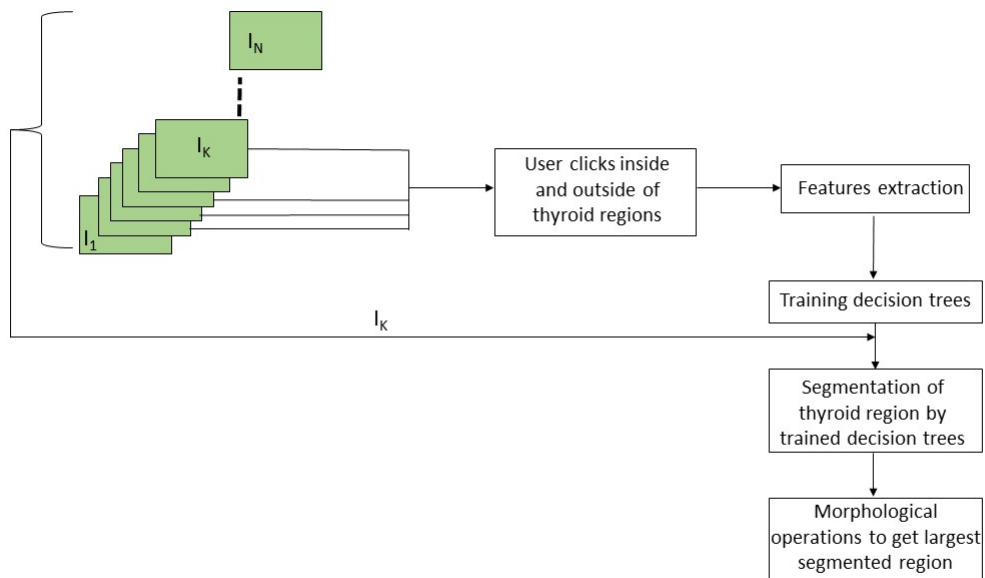
Fig. A.5: Schematic description of PBC segmentation method

After segmentation and correction of the segmentation errors, still some regions which are not part of thyroid might exist. In order to eliminate these regions, morphological

65

operations were carried out to find the largest connected component which is then considered to be the final segmented thyroid.

We also tried to test this approach by using additional features as presented in Chang *et al.*, but we found out that these features only slightly improved the segmentation accuracy, while the computation time increased significantly. Hence, we selected only the three aforementioned features and trained our decision trees with them. Similarly, we chose to work with decision trees instead of radial basis function (RBF) neural network as used in Chang et al. because of its faster computation time.

This approach is the most intuitive one and requires the least user interaction. However, the user can select more than one thyroid region during the initialization phase, and the features are computed accordingly. So, the user should wisely select the regions that are only the part of thyroid.

A.2.4 Random Forest Classifier (RFC)

This approach is based on training a random forest classifier for a binary classification problem, which classifies each of the voxel in the thyroid US images as thyroid or non-thyroid. RFC is basically a type of ensemble learning method which constructs a final classifier using a set of M individual weak classifiers. In our case, we created 12 binary decision trees of depth 10.

We trained our RFC using a 10-fold cross validation technique where 9 datasets were used for the training and 1 as validation data for testing the trained model. This was repeated until all the 10 datasets were used for testing at different iterations. The RFC approach uses some typical out of the box image processing features including gradients, Laplacian, Gaussian blur, and resampling at various resolutions, making it a total of 30 different features for training the decision trees. These features are computed in each

voxel with voxel size 15.

The input from the training data for each of the trees, $x \in \{1, ..., M\}$, in the ensemble is created by using bootstrapping of the samples (bagging) from the training dataset and randomly sampling the subset of the features supplied to the each tree. Each tree is a collection of nodes N and features F, which aid to final classification result. A decision tree is made up of a single parent node $N_{p,x}$, multiple splitting nodes $N_{s,x,i}, \forall i \in \{1, ..., k\}$, and leaf nodes $N_{l,x,j}, \forall j \in \{1, ..., p\}$. During splitting of the nodes, the best split is not chosen based on all the features but a random subset of features from the training dataset.

All the leaf nodes inside a decision tree will have a final probabilistic model $\phi_{x,j} \in [0, 1]$ associated with it. The final decision of a forest for each of the patches extracted from the US images is made by averaging the individual decisions $(\phi_{x,j}(p))$ from all the individual trees in the forest.

$$P^{RF}(y(p) = 1) = \frac{1}{M} \sum_{x=1}^{M} \phi_x(p) \tag{A.8}$$

We have used the most common and recognized method to train the classifier [31] and [32]. The implementation of RFC is carried out using IMFUSION [18].

A.2.5 Convolutional Neural Network (CNN)

This approach is based on training of the CNN using the U-net architecture (Fig. A.6) proposed by Ronneberger et al. [32] which consists of encoder and decoder parts that analyse the whole image by contracting in each successive layers and then expanding in order to produce a full-resolution segmentation, respectively. Just like RFC, the training and testing of CNN is performed using a 10-fold cross validation technique. The input

for the CNN consists of a 3D thyroid US images and its corresponding ground truth. The input can be represented as $D = (I_n, Gn)$, where I_n denotes one of the thyroid US image and G_n denotes its ground truth obtained from medical experts.
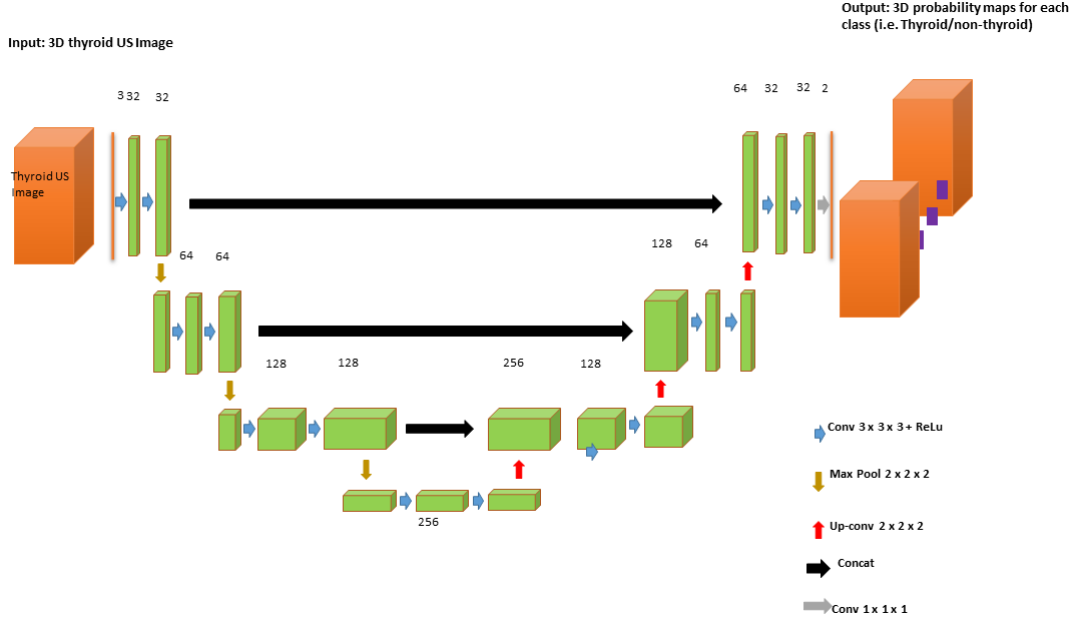
Fig. A.6: Architecture of our 3D U-Net CNN. Each green box represents the feature maps.

The network consists of two paths (i.e downsampling/encoder/left side and upsampling/decoder/right side). The downsampling path consists of two 3 x 3 x 3 convolution followed by a ectified linear unit (ReLu) in each layer and then a 2 x 2 x 2 max pooling with stride of 2 in each dimension. The number of feature channels are doubled in each downsampling step. The upsampling path remaps the lower resolution feature maps to the higher resolution space of the input images. It does this by upsampling the feature maps followed by a 2 x 2 x 2 convolution (up-convolution) which halves the number of feature channels in each upsampling step, a concatenation with the corresponding feature map from the downsampling path and two 3 x 3 x 3 convolutions, each followed by a ReLu activation. The final convolution layer uses a 1 x 1 x 1 convolution with a voxel-wise softmax activation function to compute a 3D probability map for each of the

target label (i.e. thyroid or non-thyroid) as the output of our network.

Since the available datasets were only with 10 datasets consisting of 1416 images, we had to make sure that the network was not overfitting. We performed data augmentation by rotating the images at random angles between -10°and +10°, translating between -20 and +20 voxels in each dimension, scaling between -1.5 to 1-5 times from the original size and since the thyroid are in elft and right side in human body, we also flipped the images. We added a dropout of 25% after each pooling layer so that the unnecessary neurons are discarded. Finally, we used Adam optimizer with relatively low learning rate to make sure that the network was not overfitting. During the training, we observed that the validation accuracy was very close to that of the training accuracy which proves that our network was not overfitting.

A.2.6   3D Reconstruction and Volume Computation

This step involves 3D reconstruction and volume computation of the segmented thyroid from ACWE and PBC using IMFUSION [18] and GC using MeVisLab imaging tools. The segmented 2D images are stored as binary images which are processed to make a video. The video file is passed along with the tracking data to IMFUSION and MeVisLab for 3D reconstruction of the thyroid as well as volume computation. The reconstruction is done by the interpolation between the corresponding image frames of the ultrasound sweep to fill the empty spaces between the image slices.

Volume computation of thyroid is particularly important for the medical doctors as this allows them to keep track of the size of the thyroid over time and diagnose whether the patients have any thyroid disorders or not.

A.3 Experimental Results

A.3.1 Data Collection

We acquired the thyroid datasets from different clinical university hospital-based sources. A total of 6 healthy human datasets were acquired using the General Electric (GE) Logiq E9 US system which was equipped with the Ascension driveBay EM tracking system. These dataset along with the ground truth are available at Open-CAS [33]. A ML6-15 linear probe was used to acquire the data. All the images were acquired along with a tracking matrix that gave the transformation from the origin of the EM tracking system to the centre of the image. These matrices are used for the 3D reconstruction of the segmented thyroid. The images for the evaluation of nonautomatic methods had a size of 760 x 500 pixels.

A total of 1416 2D images corresponding to 10 datasets were acquired and used for the evaluation of both the automatic and nonautomatic methods. The 3D models of all the 10 datasets were used for evaluating the automatic methods. All these datasets are stored in the DICOM format. To evaluate the accuracy of our segmentation approaches, we acquired the ground truth by manually tracing the thyroid contour with the help of two medical experts from Magdeburg university clinic using MeVisLab. The datasets are presented in Table A.1. The results and discussion may be presented separately, or in one combined section, and may optionally be divided into headed subsections.

A.3.2 Evaluation Procedure

For evaluation of the segmented images, we compare all the segmentation algorithms using two performance measures. We compute Dice's Coefficient (DC) to compare the segmentation accuracy between Active Contours, Graph Cut and Pixel Based Classifier.

Table A.1: Acquired Datasets for the evaluation of non-automatic and automatic methods.

Dataset	Number of Images
D1	96
D2	50
D3	94
D4	55
D5	135
D6	167
D7	216
D8	211
D9	201
D10	191
Total	**1416**

Similarly, we compute Hausdorff Distance (HD) to compare the accuracy of all the algorithms with the works of Dornheim *et al.* [14]. These measures are computed by comparing the segmentation results with the ground truth images.

Dice's Coefficient is a numerical estimate used for comparing the similarity of two samples. In our case, it is a measure to see how accurate our segmented results were by comparing the segmentation results with the ground truth obtained by manual segmentation of the thyroid by trained medical staff. It ranges from 0 to 1, 0 meaning the two datasets are completely different from each other and 1 meaning the datasets completely overlap with each other. It is computed by using the following formula:

$$Dice's \quad Coefficient = \frac{2\,|X \cap Y|}{|X| + |Y|} \tag{A.9}$$

where, X is the segmented image and Y is the ground truth.

Similarly, **Hausdorff Distance** measures how far two subsets of a metric space are

from each other. In other words, it is the greatest of all the distances from a point in one set to the closest point in the other set; so the less the distance, the closer are the sets. It is computed by using the following formula:

$$Hausdorff \quad Distance = \max_{a \in A}\{\min_{b \in B}\{d(a,b)\}\} \tag{A.10}$$

where, x are the pixels in the segmented image X and y are the pixels in the ground truth image Y.

The results of segmentation are later used for 3D reconstruction and volume computation. We compare the volumes of the segmented thyroid obtained from all the five algorithms. The accuracy of volume computation is computed by comparing the volume of the 3D reconstructed segmented thyroid to that of the ground truth.
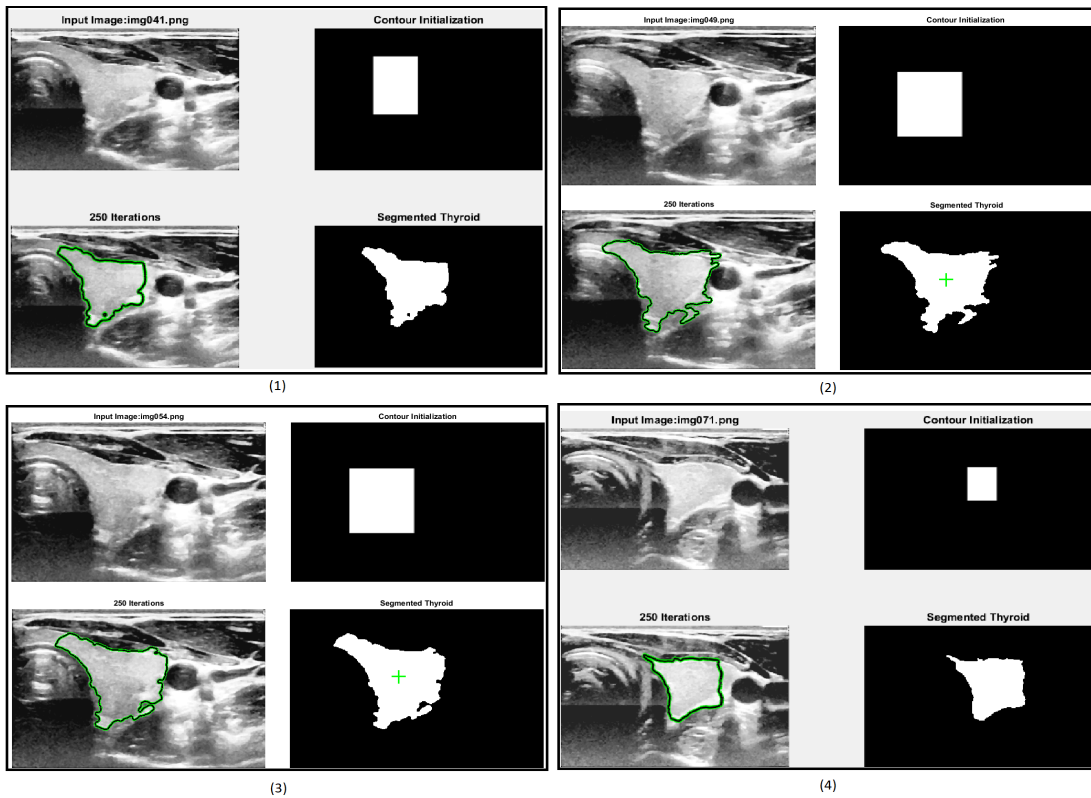
Fig. A.7: Segmentation of 4 different thyroid images using ACWE.

A.3.3 Analysis of Segmentation and 3D Reconstruction

This section is further divided into two sub-sections where the first sub-section will present the visual analysis of the segmented images and the second sub-section will present the quantitative comparison of accuracy, robustness, ease of use and computation time of all the segmentation algorithms that we have discussed.
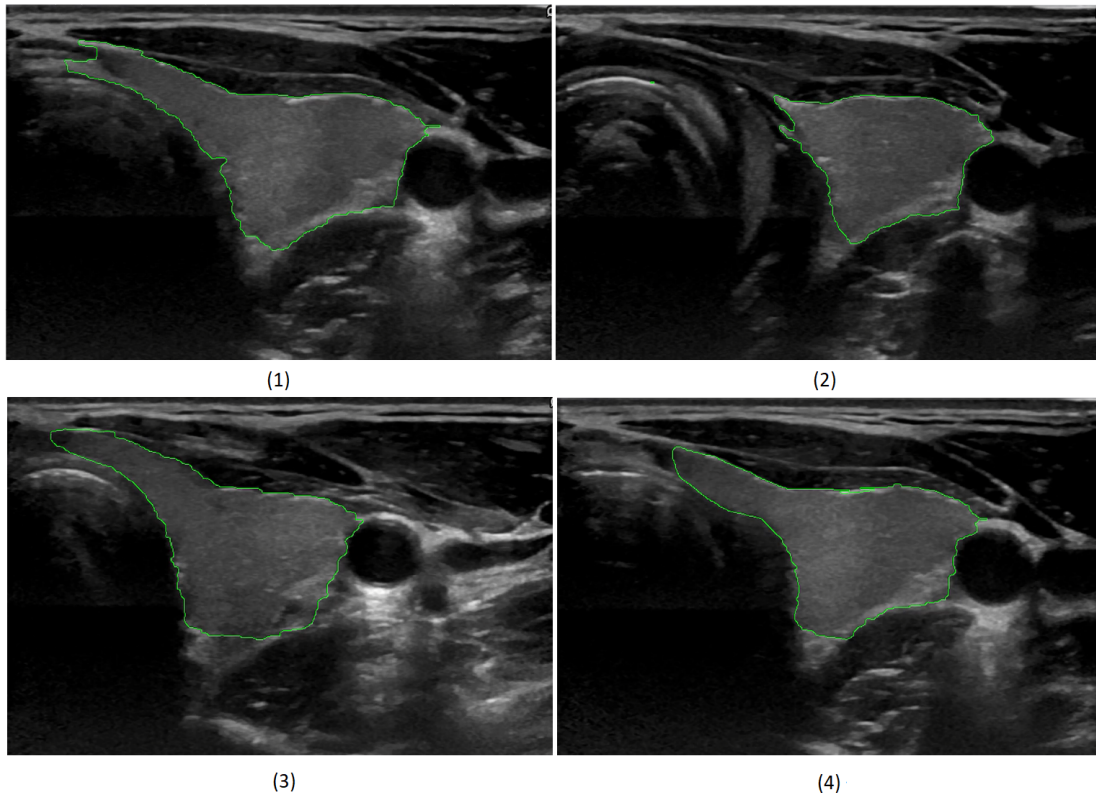
Fig. A.8: Segmentation of 4 different thyroid images using GC.

Visual Analysis

As mentioned earlier, a total of 1416 images in the ten datasets were taken for the evaluation procedure. An example of segmented thyroid US image from each of the proposed algorithms will be presented in this section.

The result of segmentation in four thyroid images using ACWE along with the user

initialized mask is shown in Fig. A.7, GC is shown in Fig. A.8, PBC is shown in Fig. A.9, RFC is shown in Fig. A.10 and CNN is shown in Fig. A.11. These segmentation results show that the automatic methods produce better segmented thyroid as compared to the non-automatic methods. Within the non-automatic methods, ACWE and GC give a better approximation of the segmented thyroid region compared to PBC as it has few over-segmented areas. In case of ACWE, the number of iterations of contour evolution is set by the user in order to optimize between the accuracy and the computation time and because of this, the contour does not reach the narrow areas like the isthmus as shown in Fig. A.1. We also allow the user to stop the segmentation process where GC is more user friendly as it allows the user to disregard the over-segmented areas in the post-segmentation stage. PBC works by computing the features from the areas the user select during the initialization process and because of this,only those areas that have very similar features to that of initialized areas are segmented as thyroid region. This results in under as well as over segmentation most of the times.

Similarly, for the visualization, we performed the 3D reconstruction of the segmented thyroid using the whole set of 2D segmented images. The 3D reconstructed thyroid using Imfusion is shown in Fig. A.12 and MeVisLab is shown in Fig. A.13. With MeVisLab, we could even segment the neighbouring artery (i.e. arteria carotis) using a Hessian-based cesselness filter [25].

Quantitative Analysis

In this section, the comparison of the accuracy of segmentation in the all the five algorithms (i.e., ACWE, GC, PBC, RFC, and CNN) in terms of DC is presented in Table A.2. Also, the comparison of segmentation accuracy of all the five algorithms with two of the standard algorithms [17] in terms of HD is presented in Table A.3.
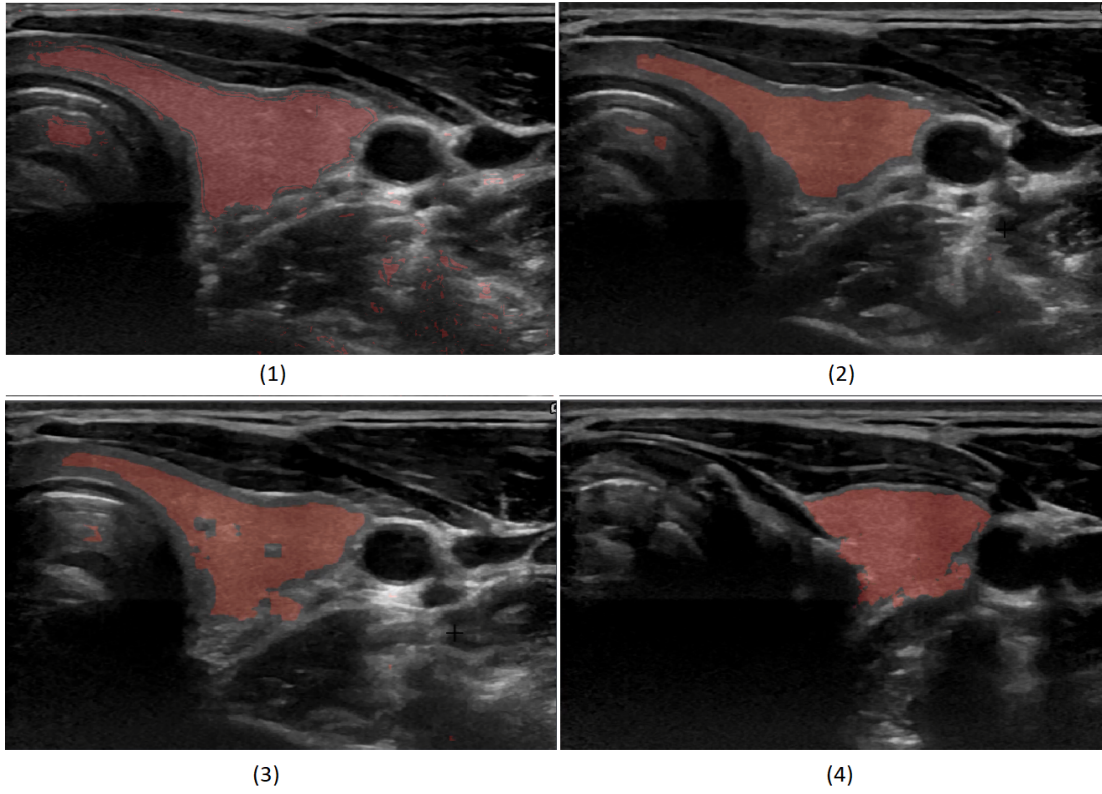
Fig. A.9: Segmentation of 4 different thyroid images using PBC.

Table A.2: Comparison of DC in five segmentation algorithms.

Dataset	ACWE	GC	PBC	RFC	CNN
D1	0.841	0.729	0.749	0.859	0.863
D2	0.819	0.636	0.666	0.864	0.876
D3	0.804	0.706	0.610	0.853	0.872
D4	0.816	0.841	0.680	0.872	0.869
D5	0.771	0.706	0.673	0.831	0.879
D6	0.781	0.853	0.623	0.853	0.874
D7	0.788	0.848	0.659	0.895	0.895
D8	0.746	0.746	0.547	0.877	0.877
D9	0.785	0.676	0.732	0.841	0.841
D10	0.852	0.912	0.761	0.875	0.862
Average	**0.800**	**0.765**	**0.670**	**0.862**	**0.876**
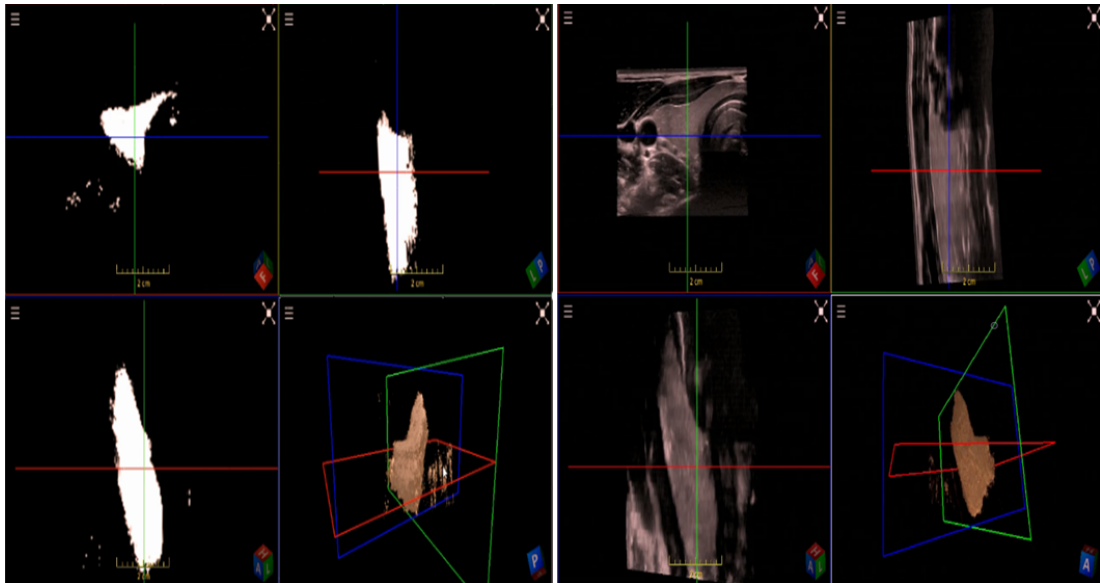
Fig. A.10: Segmentation of thyroid using RFC: Left: Segmented thyroid images as binary images with three different viewing angles (Top-Left, Top-right and Bottom-left and 3D thyroid in Bottom Right). Right: Original thyroid images with three different viewing angles (Top-left, Top-right and Bottom-left and Segmented 3D thyroid in Bottom Right
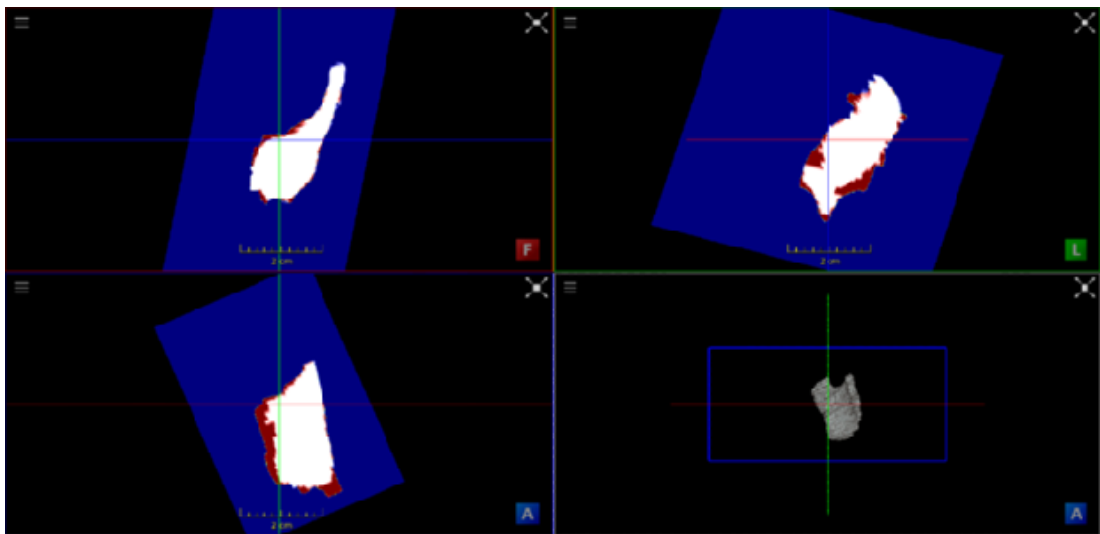
Fig. A.11: Comparison of segmentation of thyroid (red) with the ground truth (white) using CNN: Top-left, Top-right and Bottom-left: Three different vieweing angles of segmented thyroid and ground truth, Bottom-Right: Segmented thyroid in 3D.

The volume of the segmented thyroids from ACWE, GC and PBC were computed after the 3D reconstruction using Imfusion and MevisLab and are presented in Table
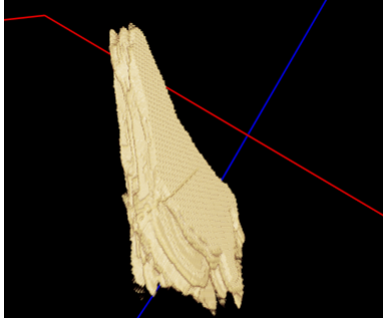
Fig. A.12: 3D reconstructed thyroid using Imfusion.
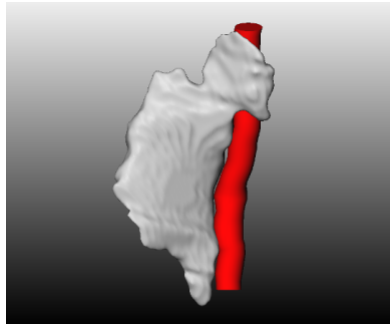
Fig. A.13: 3D reconstructed thyroid (white) along with the neighbouring artery (red) using MevVisLab

Table A.3: Comparison of our approaches with other segmentation algorithms.

Approach	Hausdorff Distance (mm)
ACWE	8.1
Graph Cut	8.3
Pixel-based	9.5
RFC	7.5
CNN	7.0
Volumetric Mass Spring Model	11.1
Surface Mass Spring Model	9.8

A.4. The results of the volume computation from segmentation results show a close correlation with the segmentation results as well as ground truth in terms of accuracy.

We compared the three nonautomatic algorithms not only based on their accuracy of segmentation but also on other factors like the computation time, robustness of the

Table A.4: Volume Comparison of 2D segmented and 3D reconstructed thyroid to ground truth in cm^3.

Dataset	ACWE	GC	PBC	Ground truth
D1	10.15	8.79	9.04	12.07
D2	11.46	8.90	9.30	13.99
D3	12.45	10.97	9.42	15.51
D4	11.91	13.82	9.86	14.64
D5	10.83	9.78	9.41	13.93
D6	12.18	13.95	10.90	9.86
D3	10.85	11.68	9.07	13.77
D4	10.66	10.66	7.82	14.29
D5	11.91	10.25	11.10	15.16
D6	10.52	11.26	9.40	12.35
Average	**11.29**	**11.01**	**9.53**	**13.56**

algorithm, number of user interactions required, etc. All the algorithms performed differently on average where ACWE performed the best with an average DC of 0.800, PBC performed the worst with an average DC of 0.670, and GC performed relatively well with an average DC of 0.765. Even though ACWE was found to be the best performer, it is not accurate enough to use for clinical practices as they require relatively higher accuracy.

ACWE produced undersegmented and oversegmented results in some cases as the contour evolution (set by the user) does not reach all the regions of thyroid (e.g., isthmus of thyroid) as well as due to the wrong initialization of the contour (this happens when the segmentation results from one image frame are used to segment the corresponding image frames). In order to address these problems, the user could stop the ongoing segmentation at any image frame and change the number of iterations as well as re-initialize the initial mask. 7.7 re-initializations were required on average per dataset. Similarly, the average computation time for ACWE was around 369 seconds in average making

it the slowest of all the algorithms, and the initialization determined the rest of the segmentation process. Hence, it is not very robust as compared to the other algorithms. All the methods were implemented in MATLAB in a Lenovo T430 ThinkPad notebook with Intel Core i5-3320M CPU, 2.60 GHz processor, and 8.00 GB RAM.

GC required the most number of user interactions (i.e., 36 scribbles on average) as the user could visualize the segmentation results instantly and improve it with more interactions. Hence, the quality of the results is directly proportional to the number of user interactions using this algorithm. The computation time was around 98 seconds on average per dataset. Graph cut is robust compared to the other two approaches as the user can control the results of the segmentation (i.e., during post-segmentation).

PBC required very few user interactions as the user had to click twice, one inside and one outside of the thyroid. However, the user could take more samples by additional clicks to improve the segmentation results. On average, 4.8 clicks were made while segmenting the images. In the same time, if wrong samples were taken, the user had to start the process from the beginning. This makes the algorithm less robust as compared to GC and ACWE. The computation time was around 10 seconds making it the fastest of all the algorithms. The comparison of the computation time and the number of user interactions required in all the three algorithms are shown in Table A.5.

Table A.5: Comparison of average computation time and number of interactions.

Algorithm	Computation Time	Number of User Interactions
ACWE	369	7.7
Graph Cut	98	36
Pixel Based	10	4.8
RFC	15.62	None
CNN	34.45	None

The RFC and CNN yielded an average DC of 0.862 and 0.876, respectively, in ten

datasets when tested using a 9-fold cross validation. The computation time for the predication of one volume was on average 15.62 seconds for the RFC and 34.45 seconds for the CNN. These approaches had higher accuracies of segmentation as compared with ACWE, GC, and PBC. Also, these methods are highly robust as the algorithm does not depend on user interaction. Both of these approaches were also implemented in the same workstation as mentioned before.

A.4   Discussions and Conclusions

As mentioned earlier, it is essential to keep track of thyroid shape and size over time as it helps to diagnose whether the thyroid is healthy or pathological. In this paper, we have worked on three thyroid segmentation techniques which attempted to extend the 2D segmentation algorithm to generate a 3D segmented thyroid. We have evaluated these algorithms on the basis of accuracy of segmentation, computation time, number of user interactions required, and the robustness. At the same time, a comparison analysis was carried out with the works of Dornheim et al. [17].

We found that all three nonautomatic algorithms performed to different levels. However, a specific approach can be chosen if faster results are required or the least human interaction is desired. The result of volume computation corresponds to the segmentation as well as to the ground truth results which shows that the volume-rendering process was correct. The accuracy of the discussed algorithms could be further improved.

The computation time of ACWE could be accelerated by reducing the image resolution and using different initialization shapes (e.g., ellipse as thyroid is elliptical in shape). Similarly, the highly echogenic areas near the thyroid could be detected by preprocessing and later the evolution of the contour could be restricted to these areas which would reduce the oversegmentation. The preprocessing step can be further worked on with

80

new contrast enhancement and filtering algorithms so that we generate a good quality ultrasound images before segmentation. In case of graph cut, a postprocessing step could be added which could take the shape prior information of the thyroid and remove the oversegmented areas automatically. Segmentation by pixel-based classifier could be improved with more image features. It can be made fully automatic using machine learning approaches and a postprocessing step to remove the oversegmented areas just like in graph cut. Also, advanced thresholding and connected component analysis could be performed to get the largest connected component and subsequently remove any elements outside that component to get a better segmented thyroid. Furthermore, all the acquired datasets were from healthy patients, so pathological datasets have to be acquired and tested on the discussed algorithms to evaluate their practical usefulness.

We figured out that the first three methods lacked automation and machine intelligence, were not highly accurate, and required long computation time. Hence, we implemented an RFC and a CNN that predict for each voxel the probability of belonging to the thyroid. Both approaches were trained for each voxel the probability of belonging to the thyroid in the available ten datasets, and they show better results as compared to the nonautomatic approaches.

As next steps, we will investigate several other thyroid segmentation approaches based on machine learning that operate directly on the volumetric three-dimensional ultrasound data instead of the 2D frames volumetrically compounded with isotropic spacing to form a 3D volume [35]. Similarly, future steps towards these automatic approaches must include more training data especially those with thyroid diseases as we have carried out our tests on healthy thyroid images only.

Data Availability

The data used for the evaluation purpose are uploaded in Open-CAS and available publicly. (http://opencas.webarchiv.kit.edu/data/thyroid.zip)

Conflicts of Interest

"The author(s) declare(s) that there is not conflict of interest regarding the publication of this paper."

Acknowledgments

References

1. Understanding Thyroid Problems - the Basics. "https://www.webmd.com/women/picture-of-the-thyroid".

2. Zhao, J., Zheng, W., Zhang, L., Tian, H.: Segmentation of ultrasound images of thyroid nodule for assisting fine needle aspiration cytology. Health information science and systems 1(1), 5 (2013).

3. Kaur, J., Jindal, A.: Comparison of thyroid segmentation algorithms in ultrasound and scintigraphy images. International Journal of Computer Applications 50(23) (2012).

4. Sheeja Agustin, A., Babu, S.S., Nadu, K.T.: Thyroid segmentation on us medical images: An overview (2012).

5. Selvathi, D., Sharnitha, V.: Thyroid classification and segmentation in ultrasound images using machine learning algorithms, 836-841 (2011). IEEE.

6. Mahmood, N.H., Rusli, A.H.: Segmentation and area measurement for thyroid ultrasound image. International Journal of Scientific & Engineering Research 2(12) (2011).

7. Keramidas, E.G., Iakovidis, D.K., Maroulis, D., Karkanis, S.: Efficient and effective ultrasound image analysis scheme for thyroid nodule detection, 1052-1060 (2007).

Springer.

8. Maroulis, D.E., Savelonas, M.A., Iakovidis, D.K., Karkanis, S.A., Dimitropoulos, N.: Variable background active contour model for computer-aided delineation of nodules in thyroid ultrasound images. IEEE Transactions on Information Technology in Biomedicine 11(5), 537-543 (2007).

9. Savelonas, M.A., Iakovidis, D.K., Legakis, I., Maroulis, D.: Active contours guided by echogenicity and texture for delineation of thyroid nodules in ultrasound images. IEEE Transactions on Information Technology in Biomedicine 13(4), 519-527 (2009).

10. Garg, H., Jindal, A.: Segmentation of thyroid gland in ultrasound image using neural network, 1-5 (2013). IEEE.

11. Narayan, N.S., Marziliano, P., Kanagalingam, J., Hobbs, C.G.: Speckle patch similarity for echogenicity-based multiorgan segmentation in ultrasound images of the thyroid gland. IEEE journal of biomedical and health informatics 21(1), 172-183 (2017).

12. Kollorz, E.N., Hahn, D.A., Linke, R., Goecke, T.W., Hornegger, J., Kuwert, T.: Quantification of thyroid volume using 3-d ultrasound imaging. IEEE Transactions on medical imaging 27(4), 457-466 (2008).

13. Chang, C.-Y., Lei, Y.-F., Tseng, C.-H., Shih, S.-R.: Thyroid segmentation and volume estimation in ultrasound images. IEEE transactions on biomedical engineering 57(6), 1348-1357 (2010).

14. Dornheim, J., Dornheim, L., Preim, B., Tonnies, K.D., Hertel, I., Strauss, G.: Stable 3d mass-spring models for the segmentation of the thyroid cartilage. Otto-von-Guericke-Universitat Magdeburg. Germany (2005).

15. Osman, A.: Automated evaluation of three dimensional ultrasonic datasets. PhD thesis, INSA de Lyon (2013).

16. Chan, T.F., Vese, L.A.: Active contours without edges. IEEE Transactions on image processing 10(2), 266-277 (2001).

17. Rother, C., Kolmogorov, V., Blake, A.: Grabcut: Interactive foreground extraction using iterated graph cuts 23(3), 309-314 (2004). ACM

18. ImFusion GmbH. "https://ww.imfusion.de".

19. MeVisLab. "https://www.mevislab.de/".

20. Gir, R., Jain, L., Rai, R.: Speckle reduction of synthetic aperture radar images using median filter and savitzky-golay filter. International Journal of Computer Applications 113(11) (2015).

21. Tay, P.C., Garson, C.D., Acton, S.T., Hossack, J.A.: Ultrasound despeckling for contrast enhancement. IEEE Transactions on Image Processing 19(7), 1847-1860 (2010).

22. Courant, R., Friedrichs, K., Lewy, H.: On the partial difference equations of mathematical physics. IBM journal of Research and Development 11(2), 215-234 (1967).

23. Poudel, P., Hansen, C., Sprung, J., Friebe, M.: 3d segmentation of thyroid ultrasound images using active contours. Current Directions in Biomedical Engineering 2(1), 467-470 (2016)

24. Open-CAS. http://opencas.webarchiv.kit.edu/data/thyroid.zip

25. Sato, Y., Nakajima, S., Atsumi, H., Koller, T., Gerig, G., Yoshida, S., Kikinis, R.: 3d multi-scale line filter for segmentation and visualization of curvilinear structures in medical images. In: CVRMed-MRCAS'97, pp. 213-222 (1997). Springer

26. Karamalis, A., Wein, W., Kutter, O., Navab, N.: Fast hybrid freehand ultrasound volume reconstruction. In:Medical Imaging 2009: Visualization, Image-Guided Procedures, and Modeling, vol. 7261, p. 726114 (2009). International Society for Optics and Photonics.

27. Cicek, O., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O.: 3d u-net: learning dense volumetric segmentation from sparse annotation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 424-432 (2016). Springer

28. Mylona, E. A., Savelonas, M. A., Maroulis, D.: Automated adjustment of region-based activecontour parameters using local image geometry. IEEE transactions on cybernetics, 44(12), 2757-2770 (2014).

29. Mylona, E. A., Savelonas, M.A., Maroulis, D.: Self-parameterized active contours based on regional edge structure for medical image segmentation. SpringerPlus, 3(1), 424 (2014).

30. Stefano, A., Vitabile, S., Russo, G., Ippolito, M., Sabini, M. G., Sardina, D., Gambino, O., Pirrone, R., Ardizzone, E., Gilardi, M. C.: An enhanced random walk algorithm for delineation of head and neck cancers in PET studies. Medical and Biological Engineering and Computing, Springer, 55(6) 897-908 (June 2017).

31. L. Breiman.: Random forests. Machine Learning, 45(1) 5-32 (2001).

32. Criminsi, A., Shotton, J.: Decision forests for Computer Vision and Medical Image Analysis. Springer Science and Business Media, (2013).

33. Orchard, M. T., Bouman, C. A.: Color Quantization of Images. IEEE Transactions on Signal Processing, 39(12), 2677-2458 (1991).

34. Blake, A., Rother, C. Brown, M., Perez, P. Torr, P.: Interactive image segmentation using an adaptive GMMRF model. In European conference on computer vision, Springer, 428-441, Berlin, Heidelberg, Germany (2004).

35. Boykov, Y. Y., Jolly, M. P.: Interactive graph cuts for optimal boundary and region segmentation of objects in ND images. In Computer Vision, ICCV, Proceedings. Eights IEEE International Conerence on, 1, 105-112, (2001).

CHAPTER B

Anatomical Structure Segmentation in Ultrasound Volumes using Cross Frame Belief Propagating Iterative Random Walks

Anatomical Structure Segmentation in Ultrasound Volumes Using Cross Frame Belief Propagating Iterative Random Walks

Debarghya China, *Graduate Student Member, IEEE*, Alfredo Illanes, Prabal Poudel, Michael Friebe, *Senior Member, IEEE*, Pabitra Mitra, *Senior Member, IEEE*, and Debdoot Sheet, *Member, IEEE*

Abstract—Ultrasound (US) is widely used as a low-cost alternative to computed tomography or magnetic resonance and primarily for preliminary imaging. Since speckle intensity in US images is inherently stochastic, readers are often challenged in their ability to identify the pathological regions in a volume of a large number of images. This paper introduces a generalized approach for volumetric segmentation of structures in US images and volumes. We employ an iterative random walks (IRW) solver, a random forest learning model, and a gradient vector flow (GVF) based interframe belief propagation technique for achieving cross-frame volumetric segmentation. At the start, a weak estimate of the tissue structure is obtained using estimates of parameters of a statistical mechanics model of US tissue interaction. Ensemble learning of these parameters further using a random forest is used to initialize the segmentation pipeline. IRW is used for correcting the contour in various steps of the algorithm. Subsequently, a GVF-based interframe belief propagation is applied to adjacent frames based on the initialization of contour using information in the current frame to segment the complete volume by frame-wise processing. We have experimentally evaluated our approach using two different datasets. Intravascular ultrasound (IVUS) segmentation was evaluated using 10 pullbacks acquired at 20 MHz and thyroid US segmentation is evaluated on 16 volumes acquired at 11–16 MHz. Our approach obtains a Jaccard score of 0.937 ± 0.022 for IVUS segmentation and 0.908 ± 0.028 for thyroid segmentation while processing each frame in 1.15 ± 0.05 s for the IVUS and in 1.23 ± 0.27 s for thyroid segmentation without the need of any computing accelerators such as GPUs.

I. INTRODUCTION

ULTRASOUND imaging has been employed for clinical use for half a century, and its salient advantages including non-ionizing nature of radiation, real time imaging capability are some of the factors which make it popular over other modalities like X-rays, computed tomography (CT) or magnetic resonance (MRI). Some of the commonly practiced clinical applications includes intravascular ultrasound (IVUS), thyroid, abdominal, cardiac, breast imaging, bone sonometry, prostate, fetal, ophthalmic imaging [1]. Current advances in signal acquisition and tracking systems, enable volumetric ultrasound acquisition using approaches such as 2D+t, 3D, and 3D+t [2]. The primary limitation in 2D ultrasound due to its planar appearance of structures, may lead to unavoidable occlusion of important 3D attributes characteristic of pathologies leading to inter-observer variability in reporting. Volumetric ultrasound can directly be acquired using a 3D imaging probe or by using a tracking device integrated with a 2D probe.

Ultrasound imaging primarily relies on sensing of backscattered echoes to image an object submerged within another medium that conceals it from direct external visibility [3]. A pulse of acoustic wave at a fixed frequency typically more than human audible range, is transmitted from the transducer and travels through the tissue. During this process, its energy is partially absorbed, attenuated and backscattered [4]. The backscattered signal moves back to the transducer and contributes to

Abstract

Ultrasound (US) is widely used as a low-cost alternative to computed tomography or magnetic resonance and primarily for preliminary imaging. Since speckle intensity in US images is inherently stochastic, readers are often challenged in their ability to identify the pathological regions in a volume of a large number of images. This paper introduces a generalized approach for volumetric segmentation of structures in US images and volumes. We employ an iterative random walks (IRW) solver, a random forest learning model, and a gradient vector flow (GVF) based interframe belief propagation technique for achieving cross-frame volumetric segmentation. At the start, a weak estimate of the tissue structure is obtained using estimates of parameters of a statistical mechanics model of US tissue interaction. Ensemble learning of these parameters further using a random forest is used to initialize the segmentation pipeline. IRW is used for correcting the contour in various steps of the algorithm. Subsequently, a GVFbased interframe belief propagation is applied to adjacent frames based on the initialization of contour using information in the current frame to segment the complete volume by frame-wise processing. We have experimentally evaluated our approach using two different datasets. Intravascular ultrasound (IVUS) segmentation was evaluated using 10 pullbacks acquired at 20 MHz and thyroid US segmentation is evaluated on 16 volumes acquired at 11-16 MHz. Our approach obtains a Jaccard score of 0.937 ± 0.022 for IVUS segmentation and $0.908 \, 0.028$ for thyroid segmentation while processing each frame in 1.15 ± 0.05 s for the IVUS and in 1.23 ± 0.27 s for thyroid segmentation without the need of any computing accelerators such as GPUs.

B.1 Introduction

Ultrasound imaging has been employed for chinical use for half a century, and its salient advantages including non-ionizing nature of radiation, real time imaging capability are some of the factors which make it popular over other modalities like X-rays, computed tomography (CT) or magnetic resonance (MRI). Some of the commonly practiced clinical applications includes intravascular ultrasound (IVUS), thyroid, abdominal, cardiac, breast imaging, bone sonometry, prostate, fetal, ophthalmic imaging [1]. Current advances in signal acquisition and tracking systems, enable volumetric ultrasound acquisition using approaches such as 2D+t, 3D, and 3D+t [2]. The primary limitation in 2D ultrasound due to its planar appearance of structures, may lead to unavoidable occlusion of important 3D attributes characteristic of pathologies leading to inter-observer variability in reporting. Volumetric ultrasound can directly be acquired using a 3D imaging probe or by using a tracking device integrated with a 2D probe.

Ultrasound imaging primarily relies on sensing of backscattered echoes to image an object submerged within another medium that conceals it from direct external visibility [3]. A pulse of acoustic wave at a fixed frequency typically more than human audible range, is transmitted from the transducer and travels through the tissue. During this process, its energy is partially absorbed, attenuated and backscattered [4]. The backscattered signal moves back to the transducer and contributes to the formation of the US signal. Backscattering is contributed by changes of acoustic index while traversing across different tissues. The received pulse is envelope detected and a log-compression for formation of ultrasound signal in B-mode per scan-line that is subsequently stacked to form an image and displayed on a monitor as a gray-scale image.

Ultrasound image segmentation is complicated and a challenging task due to the presence of certain characteristic artifacts viz. speckles, shadows, and signal dropout as well

as due to the orientation based variability that affects signal acquisition [1]. In this paper, primarily address the task of segmenting 3D US volume using frame-wise 2D segmentation approach following an iterative convergence approach per-frame and an transfer of belief on contour location for solving volumetric segmentation.

The brief prior art for both 2D and 3D segmentation in general and specifically for IVUS and thyroid are presented in Sec. II. Sec. III presents the problem statement and details the algorithm. The experiment description and results for both applications are stated in Sec. IV. The discussion of various characteristics of our approach as validated experimentally in Sec. V and finally we conclude our approach in Sec. VI.

B.2   Prior Art

In this section we present a comparison of prior art related to US image and volumetric segmentation in general and also specifically look into thyroid and IVUS segmentation problem.

B.2.1   Segmentation on 2D Ultrasound Images:

Approaches include genetic algorithm based optimization of active contour, maximum likelihood based region segmentation, adaptive region growing based segmentation, texture-based split-and-merge techniques, knowledge-based approach, watershed, and learning based approaches [1].

IVUS:

Prior art predominantly make use of active surface segmentation, Hopfield neural networks, fast-marching method, implicit mode of active contour and with anisotropic

89

contour closing or explicit form of active contour model using snakes, edge tracking and gradient-based techniques, shape-driven method and optimal graph search that have been summarized by Noble et al. [1]. Papadogiorgaki et al. [5] proposed a method for intensity based segmentation of lumen and media-adventitia boundaries which use radial basis function for refining initially recognized contours. Ciompi et al. [6] introduced a multi-class classifier using pixel-wise and contextual features to segment lumen from media. The final segmentation is achieved using the coarse definition of media and adventitia tissues. Balocco et al. [7] had proposed a standardized framework for evaluation of IVUS segmentation at the 2011 MICCAI workshop, where a total of eight algorithms including four automatic and four semi-automatic ones were evaluated. In an earlier work [8] we had proposed a graph theoretic random walk method for lumen wall segmentation. Subsequently also on ultrasound backscattering physics based learning model for segmentation of lumen and external elastic laminae were proposed in [9], [10].

The limitations in prior art is their primary dependence on shape of the edges and the region information of contours which is weakly discriminable in speckle images. Also most algorithms are device frequency specific and not generally semi-automated or fully automated whether cumulatively they are challenged with inability to handle guide-wire shadows and pitch shaft discontinuity arising due to helical scan.

Thyroid:

Keramidas et al. [11] had designed a neither for thyroid gland segmentation incorporating boundary detection along with local binary patterns based textures analysis. Maroulis et al. [12], [13] developed algorithms based on level-sets and active contours incorporating variable background modeling, active contours and joint echogenicity-texture. These models primarily reduce the effect of intensity inhomogeneity. Feedforward neural network [14] have also been proposed for segmentation of thyroid region.

90

Wunderling et al. had proposed approaches based on level set, graph cut, and feature classification for thyroid segmentation [15], and Narayan et al. [2] have introduced a multiorgan segmentation for thyroid gland, carotid artery, muscles and the trachea employing spatial location based search techniques.

The segmentation algorithms predominantly rely on using pattern of the region of interest and its structure are challenged when quality of the image degrades. Moreover they fail to work when small sized objects appear in frames.

B.2.2 Segmentation of structures in 3D Ultrasound Volumes:

Segmentation using 3D active surfaces operating on region-based external forces, neural network based approaches for detection of boundary and elastic surface model fitting are some of the popularly used approaches [1]. One common denominator is the predominance in use of speckle reduction filters like anisotropic diffusion to enable better localization of the boundary around high-intensity gradient. Active models viz. deformable active surface, intensity gradient based active surfaces and structural models have been used to segment lesions in breast tissue [1].

IVUS:

A semi-automated technique using discrete dynamic contour model and active surface segmentation has been proposed in [16] to segment lumen and adventitial border across a stack of IVUS frames in a pullback. In another approach [17], a fast active surface method using neighborhood-search technique was used. Cardinal et al. [18] had designed a 3D fast-marching method based on the gray level distribution modeled using mixture of Rayleigh on the whole IVUS pullback. A graph-based approach had been proposed for lumen and external elastic laminae segmentation [19] using gated IVUS

91

image sequences. Mendizabal-Ruiz et al. [20] proposed an approach based on minimization of a probabilistic cost function where the likelihoods were acquired from a support vector machine. Zakeri et al. has proposed a four-fold method [21], where an initial contour drawn from the classification in the sparse representation framework is used to initialize on active surface model.

These methods are highly specific to the application in hand and cannot be extend beyond IVUS for generalizable. Furthermore most of them are also US frequency specific and required modifications to model parameter for extension to different frequencies.

Thyroid:

A semi-automated approach with the help of active geodesic contour had been proposed in [22] to classify and analyze the thyroid. This framework had been used for volumetric quantification, and the algorithm had been extended to segmenting other parenchymatous organ as well. Chang et al. [23] had proposed a full solution to estimate the thyroid volume, as well as classify the blocks within thyroid.

The primary challenge posted by these methods are image size and size of the organ as visible on the image such that smaller sized contours predict high amount of false areas and process is also susceptible to the imaging frequency.

The existing limitations of these US segmentation methods in respect of being entirely application oriented, and highly dependent on the shape of the region of interest while being fully focused on a particular frequency of which imaging on they are designed. Furthermore they do not exploit the reduction in compute complexity which comes with using the inter-frame dependency arises in a 3D US volume acquired. The motivation of this paper is to address these limitations using a fully automated, robust and low time complexity algorithm. Our contributions has two distinct characteristics. First,

we employ the apriori information based on physics of the US imaging process using its statistical mechanics to obtain a rough estimate of the tissues. Second, we exploit the redundancy in compute arising due to the inter-frame dependency across adjacent 2D frames constituting a 3D volume which we also term as belief propagation.

B.3 Mathematical Model for Continuous Ultrasound Frame Segmentation

A basic observation on an ultrasonic acoustic pulse passing through tissues is that one part of the signal energy is backscattered, while another part is attenuated, and the remaining energy of the signal is absorbed by the tissues traverses through it. Imaging relies on the use of these returned echo signals. The leading cause of backscattering is the presence of scatterers in the tissue. Their nature varies with the tissue type and that leads to variation in statistical behavior of the envelope of the detected ultrasonic echo signal. The scatterers' contribution to an echo can be treated as a random walk due to its presence at arbitrary locations within the resolution limits of the backscattered ultrasonic echo pulse. The primary challenge associated with segmenting structures in ultrasonic images is the stochastic nature of speckles in images. Here we handle this challenges by relying on the statistical nature of envelope of the received ultrasonic echo. Here we employ (i) a machine learning based model to weakly predict the type of the tissue based on parametric models of ultrasound echo in order to provide an initial seed point to the iterative random walks solver, and subsequently (ii) calculate the gradient vector flow between neighboring frames for belief-propagation based segmentation solution across neighboring frames for volumetric analysis, and (iii) an iterative random walks solver for computing the segmentation across the serial frames in an ultrasound volume. The mathematical problem statement is defined as below.

Let us consider an US frame I, where $i(x)$ is the intensity at location x and k is the number of unique contours split that I into $k + 1$ disjoint set as $I_{L_1}, I_{L_2}, ..., I_{L_{k+1}}$ such that $I_{L_{k_1}} \cap I_{L_{k_2}} = \phi$ and $\forall L_1 \neq L_2, (k_1, k_2) \in \{1, 2, ..., k+1\}^2$ and $I_{L_1} \cup I_{L_2} \cup ... \cup I_{L_{k+1}} = I$. The image I can further also be represented as graph G where the nodes of the graph can be represented as $n \in I$ and the weights of edges connecting the nodes of graph G are modeled based on the physics of the acoustic energy propagation and attenuation within highly scattering biological tissues. The probability of each node $n \in G$ to belong to the regions in $\{I_{L_1}, I_{L_2}, ... I_{L_{k+1}}\}$ can be obtained using the random walks approach for the image segmentation [24]. The class posterior probability at $x \in I$ is the probability of the corresponding node $n \in G$. A pixel at location x is labeled as $y = argmax\{p(L_1|x, I), p(L_2|x, I), ... p(L_{k+1}|x, I)\}$.

A set of seeds S_M constitutes some of the marked nodes of graph G such that $S_M \subseteq \{(S_M \in I_{L_1}) \cup (S_M \in I_{L_2}) \cup ... \cup (S_M \in I_{L_{k+1}})\}$ and $(S_M \in I_{L_1}) \cap (S_M \in I_{L_2}) \cap ... \cap (S_M \in I_{L_{k+1}}) = \phi$ is defined for initialization of the random walker. An ultrasonic pulse backscattering physics based model is used to provide these initial seeds. Thus on solution, a class specific posterior probability is assigned to the unmarked nodes $S_u = G - S_M$ of the graph to obtain the segmented anatomical structure such that $G \subseteq \{S_M \cup S_u\}$ and $S_M \cap S_u = \phi$. The detail information flow of our proposed algorithm is presented in Fig. B.1. Subsequently, we discuss the various stages of the algorithm and their significance to the solution.

B.3.1 Statistical Mechanics of Ultrasound Backscattering

Let us consider the value of the signal received at a specific location as r and let the tissue type be ω, then the conditional likelihood $p(r|\omega)$ can be written as a function of the following tissue specific effects on ultrasonic signals: 1. likelihood of the received
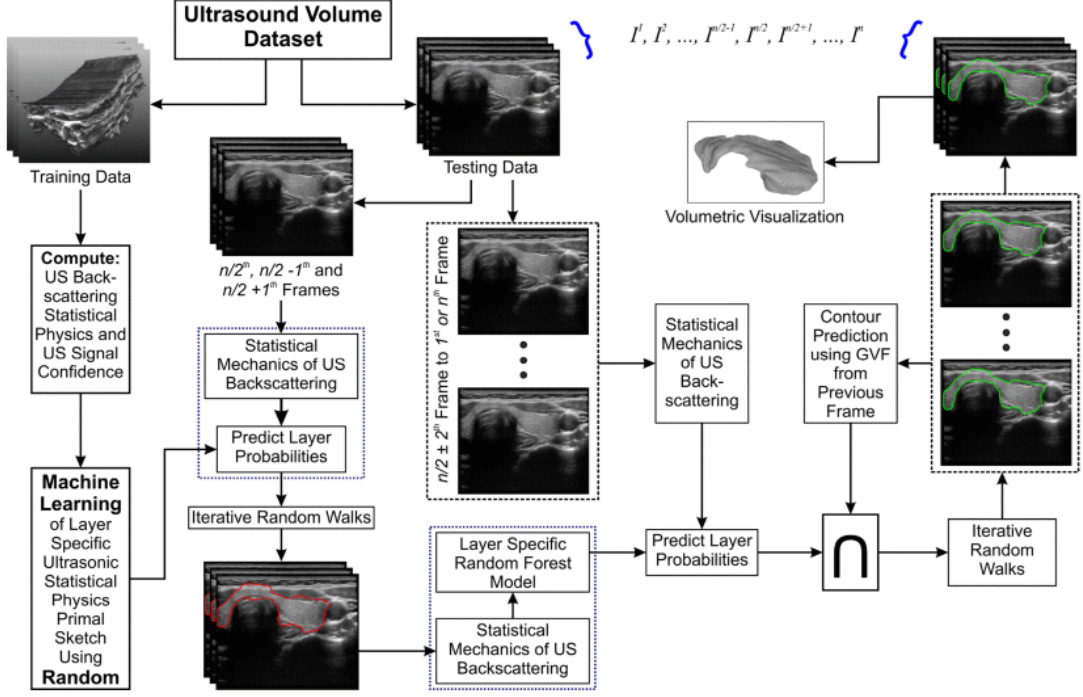
Fig. B.1: Segmentation framework of the proposed approach for ultrasound (US) volume. The US dataset has m number of volumes $(V_1, V_2, ..., V_m)$ where $m - 1$ number of volume has been used for Randm Forest (RF) training of initial segmentation and one volume is used for testing where the testing volume is consecutive of n number of individual frames i.e. $I^1, I^2, ..., I^{n/2-1}, I^{n/2}, I^{n/2+1}, ..., I^n$. Iterative random walks (IRW) is used for final segmentation in each different stages. Gradient vector flow (GVF) for each frame has been calculated from previous two frames. Finally the segmented volume of the US is visualized in volumetric visualization block.

ultrasonic signal: $f_1(r; \phi_1|\omega)$; 2. reliability of the received ultrasonic signal: $f_2(r; \phi_2|\omega)$ where ϕ_1 and ϕ_2 are the set of parameters defining $f_1(.)$ and $f_2(.)$ respectively which are dependent on the tissue specific properties. The likelihood $p(r; ...|\omega)$ can be parametrically represented taking into consideration these factors such that

$$p(r; \ldots |\omega) \propto \{ \underbrace{f_1(r; \phi_1|\omega)}_{\text{backscattering stats.}}, \underbrace{f_2(r; \phi_2|\omega)}_{\text{confidence}} \} \tag{1}$$

The received US signal r is Nakagami distributed [25] such that $p(r|\omega) \propto N(r|m, \Omega)$ [4],[10]. Since in a B-mode image, the image intensity i is a log-compressed version of

95

the signal r, the intensity $i \in I$ is accordingly Fisher-Tippett distributed [4],[10] such that $p(i|\omega) \propto F(i|\sigma)$ and

$$F(i|\sigma) \propto exp\left(\left[2i - ln\left(2\sigma^2\right)\right] - exp\left[2i - ln\left(2\sigma^2\right)\right]\right) \qquad (2)$$

where σ is the standard deviation of the intensity.

The parameters of i and σ are estimated through a nonlinear multiscale estimation. According to our proposition [26], these parameters are estimated at different scales $\tau = (\tau_{trans}, \tau_{axial})$ where τ_{trans} represents the span along the number of neighboring scan lines and τ_{axial} is the number of samples along each scan line.

The factor of reliability of an ultrasonic signal measured in terms of confidence in $f_2(.)$ is estimated using the method of random walks [10],[27]. The confidence of the ultrasonic signal is estimated as the probability of a random walker starting at a node of the graph equivalent of the US image located on the scan-line to reach the origin of each scan line where the virtual transducer element is placed akin to the physical location of the sensor element on the transducer. Thus the signal confidence is represented as

$$p(r; \dots |\omega) \propto f_2(r; \phi_2|\omega) \qquad (3)$$

where $\omega_2(.)$ is the ultrasonic signal confidence associated with the received echo r that was backscattered by a tissue type ω. This set of ultrasonic backscattering modeling statistical features enable on to learn a Random forest classifier [28] to classify different tissue layer from the US images.

B.3.2 Transfer Learning of Ultrasound Backscattering Statistics

A non-parametric machine learning framework using random forest [4] has been employed for the purpose of learning apriori information on speckles to seed. The parameters of $f_1(r; ...\omega)$ and $f_2(r; ...\omega)$ obtained earlier (Sec. B.3.1) constitutes the feature space to be modeled for prediction of the tissue-specific posterior probability. The prediction model of a random forest can be represented as

$$p(\omega|\Theta; r) = H(\omega|\Theta; r) \qquad (4)$$

where $H(\omega|\Theta; r)$ is the learnt RF model that is formally defined as a classifier consisting of a collection of tree-structured decsion makers $\{h(\omega|\Theta, \phi_z), z = 1, ...\}$, where ϕ_z are independent and identically distributed random vectors representing sample features, $\{\phi_z\} \subseteq \Theta$ and each tree $h(\omega|\Theta, \phi_z)$ casts a unit vote for the most popular class ω at input Θ [28]. During prediction, the vote casted by the forest is the class specific mean response of each of the trees such that $p(\omega|\Theta; i) = E[h(..., \phi_z)]$. The initial segmentation done using this random forest model is used to initialize seeds for each tissue type $\omega \in \{L_1, L_2, ..., L_{k+1}\}$.

B.3.3 Belief Propagation Across Neighboring Frames Using Gradient Vector Flow

Let us consider that a US volume consists of n number of frames organized as an ordered set $\{I^1, I^2, ..., I^{n/2-1}, I^{n/2}, I^{n/2+1}, ..., I^n\}$. The process starts with segmenting the $\{I^{n/2-1}, I^{n/2}, I^{n/2+1}\}$ frames in the volume using the method in Sec. B.3.2. The segments are used to initialize the seeds for the random walks based fine segmentation

detailed subsequently in Sec.3.4. Those three segmented frames are then used to further train a volume specific design random forest model presented in Sec. B.3.2. This volume specific learnt model helps to precisely predict the posterior probability of the contours and the background tissue in the current volume taking into account and learning patient specificity of ultrasonic tissue interaction with the contour in $\{I^{n/2}, I^{n/2+1}\}$, GVF is employed to estimate the contour in $I^{n/2+2}$ and this serves to initialize seeds for segmentation of the $I^{n/2+2}$ frame using IRW presented in Sec. B.3.4. This process repeats the other way to predict contour on $I^{n/2-2}$ using GVF over $\{I^{n/2-1}, I^{n/2}\}$. The final segmentation is done using interaction of the segmentation results obtained from the learnt model and the GVF for every frame process continues till it reaches the end frames on the volume. Since seed initialization with the RF on individual frames is not required, computation time is considerably reduced, and the method serves to accurately segment accounting for tissue specific variations on an auto-updating mechanism.

Table B.1: Performance Measure with the IVUS and Thyroid Dataset in the 3D Context.

Dataset	SE	SP	DSC	PAD
Phantom	0.992 ± 0.003	0.956 ± 0.031	0.950 ± 0.040	0.001 ± 0.001
IVUS	0.990 ± 0.003	0.944 ± 0.049	0.961 ± 0.023	0.002 ± 0.001
Thyroid	0.989 ± 0.006	0.938 ± 0.051	0.889 ± 0.043	0.002 ± 0.001

B.3.4 Solution to Iterative Random Walks for Final Segmentation

A predominant observation with RW is its mild variation in segmentation due to updates on seeds , and we follow this to a convergence using IRW. We iteratively employ random walk solver [24]. Here, we denote the probability of a random walker starting at a node v_q to reach a seeded point belonging to tissue type $\omega \in \{L_1, L_2, ..., L_{k+1}\}$ as x_q^w such that $\sum_\omega x_q^w = 1$.
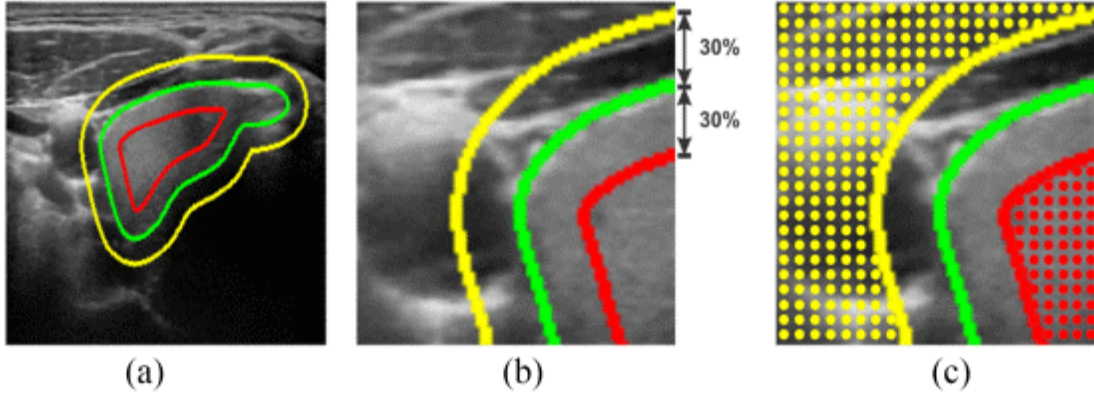
Fig. B.2: Illustration of the seed selection from foreground and background for solviing IRW. (a) Initial contour (green) with the seeds for background region (yellow) and foreground region (red). (b) Foreground (red) and background (yellow) region has been selected by a morphological regularization operation on the initial estimated region(green). (c) Neighborhood around the foreground (red dots) and background (yellow dots) that are selected as seeds for different regions.

Therefore, for label ω, the solution probabilities of a random walker could be generated at a node $q \in G$ where the G is the graphical representation of the image I and has been solved accordingly [24].

$$p(L_z|x, I) = x_q^w \forall \{q \in G \Leftrightarrow x \in I\}, Z = 1, 2, ..., k+1 \tag{5}$$

The initialization of the contour for random walks is obtained from the intersection between the prediction of the RF and GVF. The foreground and background seeds in the image were designed using morphological operation on the initial contour. First erosion operation is performed and the resultant region is used for foreground seeding while the inverse of that dilated region is used for background seeding. The structuring element in these morphological operations is chosen depending on the size of the initial contour and typically sized at 30% of the radius of the initial estimated region of the structure. This morphology based regularization is illustrated in Fig. 2. In case of iterative random walks (IRW), the solution of the current step acts as the initial contour

99

for the subsequent step until convergence is reached. This iterative process converges when the change in the posterior probability obtained in the current stage is smaller than a pre-defined tolerance (δ). The stepwise method for Iterative Random Walks is shown in Algorithm 1.

Input: Complete layer specific set of marked seeds $S_M \in \{S_{M1}, S_{M2}, \ldots, S_{Mk}\}$ with k number of layers and all the unmarked seeds S_U where present contour $\mathcal{G}_\mathcal{P} = S_M \cup S_U$

Input: Stopping criteria δ

Output: Optimal marking the unmarked seeds ($\mathcal{G}_\mathcal{P}$)

for $i \leftarrow 1$ **to** n **do**

    Compute layer specific prior probability $p(L_{\mathcal{Z}}|x, I)$ from (5);

    Generate new contour $\mathcal{G}_\mathcal{N}$;

    Count change of pixel $C_\mathcal{P} = count(\mathcal{G}_\mathcal{P} - \mathcal{G}_\mathcal{N})$;

    if $C_\mathcal{P} > \delta$ **then**

        $\mathcal{G}_\mathcal{P} = \mathcal{G}_\mathcal{N}$;

        Reinitialize S_M and S_U;

    else

        Return $\mathcal{G}_\mathcal{P}$;

    end

end

Algorithm 1: Iterative random walks (IRW) algorithm

Table B.2: Performance Evaluation Metrics of Results Obtained with Dataset and Comparison with rior Art. EEL Denotes external elastic luminae.

Methods	JCC		HD		PAD		DC	
	Lumen	EEL	Lumen	EEL	Lumen	EEL	Lumen	EEL
P1[7]	0.81 ± 0.12	0.76 ± 0.13	0.47 ± 0.39	0.64 ± 0.48	0.14 ± 0.13	0.21 ± 0.16	–	–
P2[7]	0.83 ± 0.08	–	0.51 ± 0.0.25	–	0.14 ± 0.12	–	–	–
P3[7]	0.88 ± 0.05	0.91 ± 0.04	0.34 ± 0.14	0.31 ± 0.12	0.06 ± 0.05	0.05 ± 0.04	–	–
P4[7]	0.77 ± 0.09	0.74 ± 0.17	0.47 ± 0.22	0.76 ± 0.48	0.15 ± 0.12	0.23 ± 0.19	–	–
P5[7]	0.79 ± 0.08	–	0.46 ± 0.30	–	0.16 ± 0.09	–	–	–
P6[7]	–	0.84 ± 0.10	–	0.57 ± 0.39	–	0.12 ± 0.12	–	–
P7[7]	0.84 ± 0.08	–	0.38 ± 0.26	–	0.11 ± 0.12	–	–	–
P8[7]	0.81 ± 0.09	0.79 ± 0.11	0.42 ± 0.22	0.60 ± 0.28	0.11 ± 0.11	0.19 ± 0.19	–	–
Intraobs[7]	0.88 ± 0.05	0.92 ± 0.03	0.28 ± 0.13	0.24 ± 0.12	0.11 ± 0.08	0.06 ± 0.04	–	–
Interobs[7]	0.93 ± 0.05	0.95 ± 0.03	0.17 ± 0.13	0.14 ± 0.10	0.04 ± 0.06	0.03 ± 0.03	–	–
Intermediate	0.911 ± 0.045	0.917 ± 0.049	0.239 ± 0.134	0.227 ± 0.116	0.092 ± 0.084	0.059 ± 0.050	0.876 ± 0.051	0.903 ± 0.073
Proposed	0.927 ± 0.017	0.947 ± 0.027	0.168 ± 0.081	0.173 ± 0.114	0.002 ± 0.001	0.002 ± 0.001	0.909 ± 0.025	0.0922 ± 0.035

B.4   Experiments and Results

The segmentation results at different stages of the algorithm is illustrated in Fig. B.1 for an example of thyroid segmentation. The first step in our work is to segment the three middle frames of the volume. We have employed RF and IRW to solve this problem. A RF model is learned (Sec. B.3.2) by training over a feature set (Sec. B.3.1) over 500 random samples, selected for each class in the training images and has been tested on these three middle frame in the test volume, which acts as initial segmentation of the layers for IRW (Sec. B.3.4). In the next step, a volume specific RF model is designed using the information in these three segmented frames. The model predicts the layers in the other frames of the particular volume. Subsequently a frame specific GVF is obtain for a frame using the previous two neighboring frames (Sec. B.3.3). Finally, we have taken the intersection of the two results from both RF and GVF model, which has acts as initial seeds for segmentation using IRW on a given frame. The algorithm has been experimentally evaluated on kidney segmentation using a multi-modal abdominal phantom (CIRS abdomen phantom), where four 3D US were acquired using a tomographic ultrasound device (Piur Imaging, Germany). The results of the segmentation are shown in Table B.1 together with the ground truth results are visually illustrated in Fig. B.3. Beyond the phantom based validation we also employ it for IVUS and thyroid segmentation using real patient data.

B.4.1   Intravascular Ultrasound Segmentation

Data Description:

The IVUS data used in this experiment is from the Lumen + External Elastic Laminae border detection of the IBUS Challenge dataset [7]. It consists 10 pullbacks each from

Table B.3: Hyper-parameters for both the experiments.

Hyper-parameters	IVUS	Thyroid	Explanation
$nClasses$	3	2	Number of classes
α	2	2	It affects the likelihood of vertical random walks
β	1250	5500	It effects on the robustness and accuracy of the segmentation
γ	0.05	0.05	It penalizes horizontal and diagonal random walks in the graph
$nTrees$	50	50	Number of trees
D	17	17	Number of features
$minLeaf$	50	50	Number of leafs for each trees
$treeDepth$	∞	∞	Level of the tree
$splitObj$	GDI	GDI	Gini's diversity index

a different patient acquired at 20 MHz. The data acquisition technique is detailed in [7]. All the frames are provided in DICOM format as grayscale images. The manually labeled ground truth of the lumen and external elastic laminae are also provided with this set of data. Each frame has a size of 384 x 384 pixels.

Experiments and Results:

At the time of random forest learning (Sec. B.3.2), a D (number of features) dimensional ordered vector Θ representing multiscale estimated Fisher-Tippett statistics parameter and ultrasonic signal confidence is computed and represented as $\{(\Theta; r) \forall r \in G\}$, where G is the IVUS image represented in polar domain. In this experiment, we have the tissue specific labels $y \in Y)$ with $Y = \{lumen, media, externa\}$ corresponding to the ultrasound echo measurements at grid points $r \in G$. Statistics of backscattered ultrasonic B-mode signals parameters are estimated at scales $\{(3x3),(5x3),(7x3), ...,(30x3)\}$ and this range of values are choosen according to [4]. This code is implemented on Python. The random forest classifier is learnt using hyper-parameter detailed in Table B.3. This

Table B.4: Quantitative Analysis with the Thyroid Dataset and Comparison with Previous Approaches.

Methods	SE	SP	DSC	PPV	JCC	HD	PAD
Narayan[2]	0.955 ± 0.024	0.889 ± 0.065	0.839 ± 0.047	0.806 ± 0.086	–	–	–
JCR[2]	0.564 ± 0.070	0.926 ± 0.070	0.479 ± 0.066	0.326 ± 0.057	–	–	–
Chang[23]	0.874 ± 0.117	0.560 ± 0.324	0.512 ± 0.288	0.531 ± 0.347	–	–	–
Garg[14]	0.473 ± 0.182	0.864 ± 0.224	0.400 ± 0.143	0.265 ± 0.111	–	–	–
Intermediate	0.937 ± 0.028	0.892 ± 0.068	0.828 ± 0.079	0.795 ± 0.157	0.889 ± 0.054	0.538 ± 0.488	0.011 ± 0.015
Proposed	0.989 ± 0.006	0.923 ± 0.059	0.854 ± 0.066	0.807 ± 0.110	0.908 ± 0.028	0.488 ± 0.430	0.002 ± 0.001

Table B.5: Performance Measure with the Thyroid Dataset and Comparison with Previous Approach in terms of PRI, GCE, VOI and BE.

Methods	PRI	GCE	VOI	BE
Narayan[2]	0.844 ± 0.037	0.192 ± 0.044	1.169 ± 0.206	12.221 ± 2.788
Proposed	0.972 pm 0.011	0.024 ± 0.009	0.172 pm 0.054	6.660 ± 1.962

trained model is finally tested on pre-selected test images (not used during training). This experiment is performed using a 10-fold cross-validation technique. The data was taken from 10 patient cumulatively yielding 2; 175 images used in this experiment and at each fold data from 9 patients were used for training and the remaining patient data was used for testing. Evaluation for segmentation of IVUS is performed using three measures (i) Jaccard index (JCC), (ii) Percentage of Area Difference (PAD) measure, and (iii) Hausdorff Distance (HD) following the approach presented in [7]. We compare our method with eight prior art i.e. shape-driven segmentation method [7](P1), geodesic

active contour [7](P2), fastmarching method based on gray level distributions [7](P3), graph search method [7](P4), multi-scale stacked sequential learning [7](P5), holistic approach [7](P6), support vector machine with a radial basis function kernel [7](P7), and angiocare [7](P8) respectively which has been reported by Balocco et. al. [7]. The results are quantitatively detailed in Table B.2.

B.4.2 Thyroid Segmentation

Data Description:

A total of sixteen healthy human subjects were imaged using a Logiq E9 US device (General Electric, USA) with a 11-16 MHz probe and equipped with an electromagnetic (EM) tracking system to form the free hand US volume dataset. All the volume are provided in DICOM format as RGB images. The datasets are available online [15].

Experimental Implementation and Results:

The experimental setup for thyroid segmentation is similar to IVUS as presented earlier. We have used tissue-specific labels $y \in Y$ and $Y = \{thyroid, background\}$ corresponding to the ultrasound echo measurements. The ultrasonic signal confidence is estimated using the parameters shown in Table III. The RF parameters (Table B.3) are similar to as used for earlier use case of IVUS pullback segmentation (Sec. B.4.1). This experiment is performed using a 10-fold cross-validation technique. The data was obtained from 16 patients for this experiment and at each fold 14-15 volumes were used for training and the remaining volumes were used for testing.

The segmentation performance of our algorithm applied to thyroid US has been evaluated in two different ways following the approach in [2]. The proposed algorithm is

qualitatively assessed using (i) Probabilistic rand index (PRI), (ii) Global consistency error (GCE), (iii) Variation of information (VOI) and (iv) Boundary distance error (BE). Also measures quantitatively based on the overlapping area are used, i.e. (i) Sensitivity (SE), (ii) Specificity (SP), (iii) Dice coefficient (DSC) and (iv) Positive predictive value (PPV). We have compared our approach with four different algorithms i.e. echogenicity-based quantization [2], joint classification-regression [2], RBF neural network [23], and feedforward neural network [14] which has been reported by Narayan *et. al.* [2], (Table B.4). The only limitation trying to compare commonality is that the dataset has used for this comparison is not same as the dataset used in prior art [2].
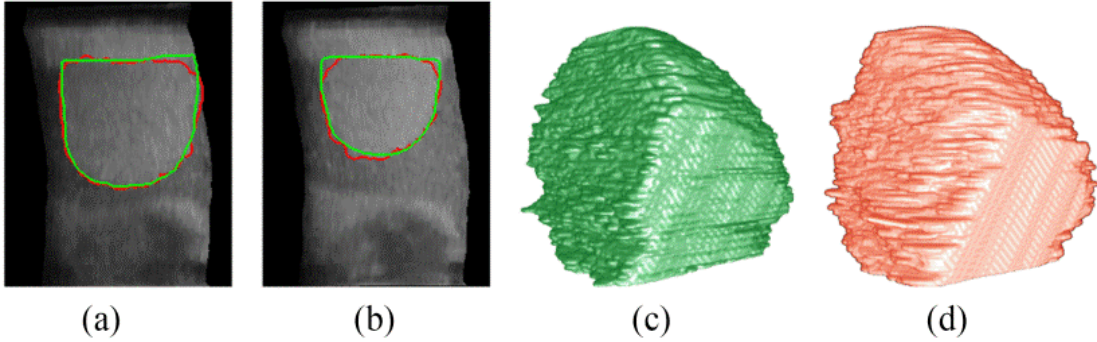
Fig. B.3: Contour segmentation for kidney phantom where (a) and (b) are two frames from different volumes, (c) is 3D visualization of the segmented phantom kedney and (d) is the ground truth volume of the same. GREEN - ground truth data and RED - result of our proposed method.

B.5 Discussion

B.5.1 Initializing seeds of Random Walks through Learning of Statistical Mechanics of Ultrasound

A key feature of this algorithm is its ability to predict the tissue layers in US images by learning of parametric model of speckle statistics. This prediction facilitates a reasonable

106
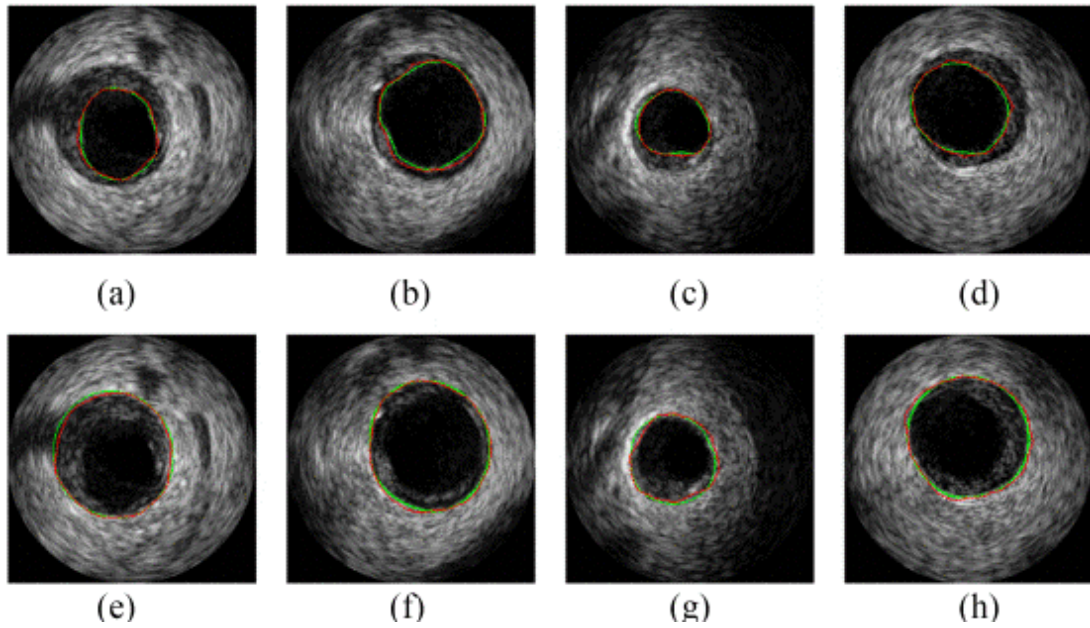
Fig. B.4: (a), (b), (c) and (d) are four segmented lumen contour and (e), (f), (g) and (h) are four segmented external elastic luminae (EEL) contour from four different IVUS pullback. Here GREEN - ground truth data and RED - result of our proposed method.
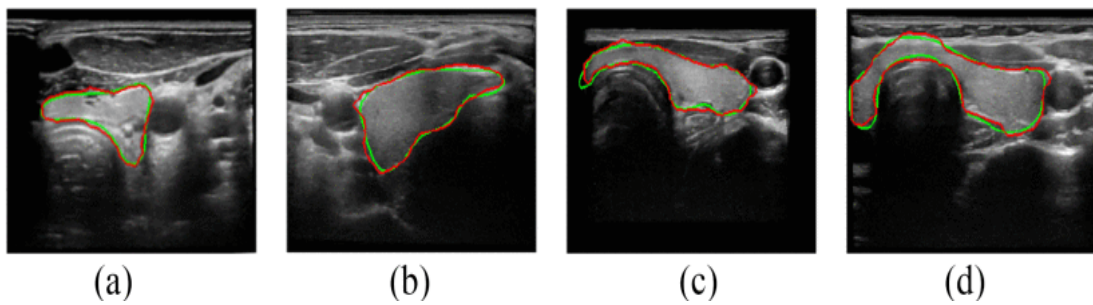
Fig. B.5: Contour segmentation for thyroid where (a), (b), (c) and (d) are four frames from four different volumes. GREEN - ground truth data and RED - result of our proposed method.

initial estimate for RW seed initialization. The RF model is trained using 500 samples from each tissue type per image. Except for a few limitations like guidewire artifact, necrotic core, muscles, trachea, and carotid, this algorithm yields perfect accurate results for segmenting the lumen, external elastic laminae as well as the thyroid. While the RF-based contour initialization on each frame acts as a good starting point, it has limitations in not being able to correctly steer the contour across neighboring frames in
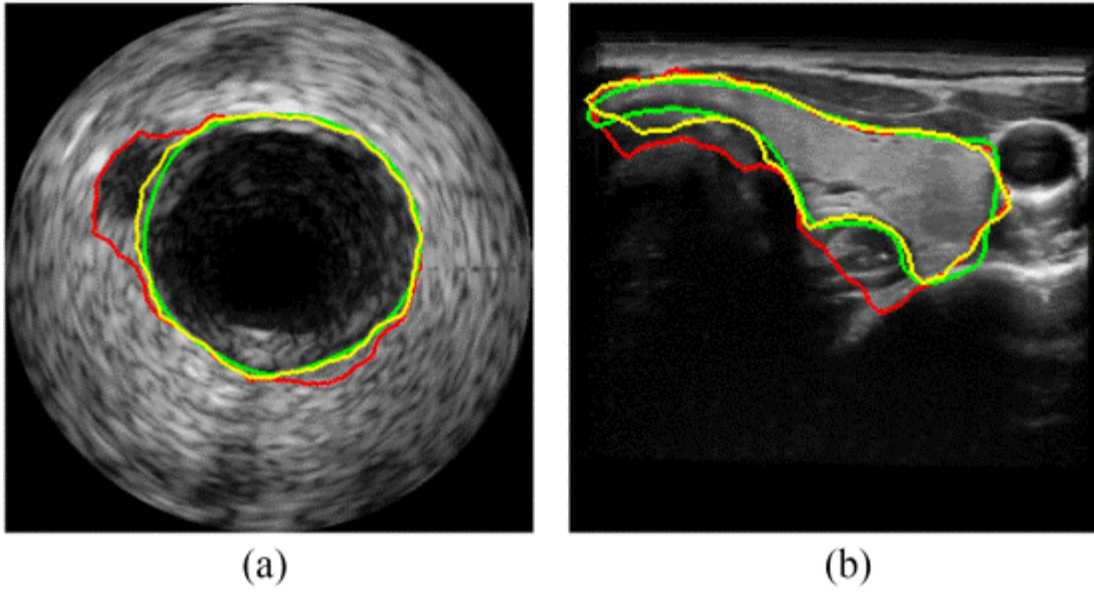
Fig. B.6: Contour segmentation in presence of artifacts for IVUS in (a) and thyroid in (b) where GREEN - ground truth of the contour, RED - result after applying only RF and YELLOW - result after applying IRW on the result of RF.
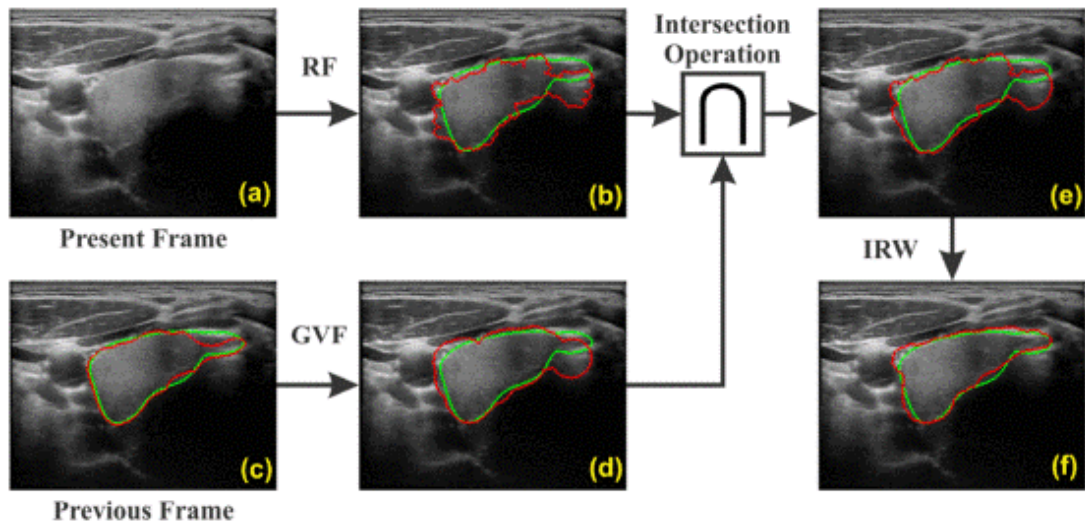
Fig. B.7: The graphical framework for the minimization of leaking and drifting from the RF model and GVF respectively.

the presence of necrotic core as is evident in the red contour on Fig. B.6 (a) and trachea as is evident in the red contour on the Fig. B.6 (b). To overcome this limitation, IRW has been employed iteratively in our formulation termed IRW as can be seen in the
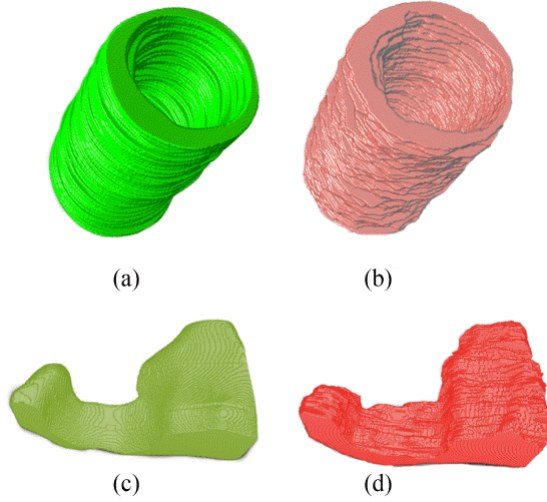
Fig. B.8: (a) and (b) are the media volume of IVUS where (a) is the ground truth and (b) is the segmented media volume. (c) and (d) are the thyroid volume where (c) is the ground truth and (d) is the segmented thyroid volume.

yellow contour in Fig. B.6.

B.5.2 Iterative Random Walker for Correcting Contour in Presence of Initialization Error

Due to the false prediction in the presence of the different artifacts, we have employed a subsequent refining stage using the RW for final segmentation using the seeds initialized by the RF model. In this experiment, 0.1% change of the contour is set as the convergence criteria for IRW. Fig. B.6 illustrates that IRW is well suited to correct the final contour. In Table II, the method Intermediate shows performance of the RF initialized IRW based segmentation obtained using method described in Sec. B.3.2 and Sec. B.3.4, which is comparably better than prior art [7] and the segmented contours are visually illustrated marked in yellow contours in Fig. B.6.

This approach while improving on the results, has the drawback when applied to a whole pullback for segmentation. When the seeds initialized for the contour are erroneous, the

computation time increases. This problem is accentuated for small size contour, i.e., it is difficult for the RF model to predict small area accurately, which sometimes leads to inaccurate segmentation results. In order to overcome this problem, we have suggested the use of belief propagation across neighboring frames as a possible solution, and its benefits are discussed subsequently.

B.5.3 Belief Propagation for Cross Frame Segmentation

The solution of the problem for segmentation on the whole volume using a frame wise approach has been discussed in Sec. V-B. We have employed belief propagation, where the predicted contour in the current frame is propagated to the next frame as an initial estimate of the contour, and then an IRW initialized with these seeds corrects the initial contour. The drawback of this method is that the error if accrued intermediate is propagated till the last frame. This helps in putting the belief propagation of the previous result with in a bound. In this method, we again employ another RF model, which is learnt to be the volume specific using the set of three segmented middle frames. This RF model predicts roughly the tissue presence for a frame, and for the same frame, the initial contour comes from the results of the previous frames as well. The intersection between these two contours is then used for the initial contour for IRW. If the RF model is predicted wrongly, the propagated results can also be used to correct the contour and vice versa. To get more accurate belief propagated contour, we introduce a gradient vector flow (GVF) over contour drifts across neighboring frames instead of conveying the exact result from the neighboring frame. The GVF is calculated using the last two frames cumulative neighbors (Sec. B.3.3). The propagation error has decreased through this combined method. The graphical framework for the minimization of leaking has been illustrated in Fig. B.7. Finally, all the segmented frames have been stacked into a volume and evaluated in comparison with the ground truth volume shown in Table B.1

110

and visualized using MeVisLab, as shown in Fig. B.8.

In this experiment we have obtained an average JCC value of 0.927 ± 0.017 and $0.947 \pm$ 0.027, HD of 0.168 ± 0.081 and 0.173 ± 0.114 and PAD value of 0.002 ± 0.001 and $0.002 \pm$ 0.001 for the segmentation of lumen and external elastic luminae boarder respectively in IVUS. Table B.2 shows the performance of our method obtained from Sec. B.3.3, which are better than as reported in the prior art [7]. Fig. B.4 (a) - Fig. B.4 (d) and Fig. B.4 (e) - Fig. B.4 (h) present the visualization of the segmented contours from lumen and external elastic luminae from four different IVUS pullbacks. The visual representation of the segmented media for whole pullback is shown in Fig. B.8 (b), which is visibly similar to the ground truth Fig. B.8 (a).

In the experiment of thyroid contour segmentation, the performance of the proposed method has been evaluated in two different ways. In Table B.4 and Table B.5, quantitative and qualitative results are summarized for our algorithm in comparison with approaches presented in prior art. In both cases it is clear that our approach outperforms the other compared algorithms [2]. Fig. B.5 (a) - Fig. B.5 (d) presents the visualization of segmented thyroid contours from four different thyroid volume and volumetric visual representation of thyroid has shown in Fig. B.8 (d) which is similar to the ground truth volume in Fig. B.4 (c) for thyroid segmentation.

B.5.4 Execution Time

The algorithm was implemented in Python 2:7 and executed on a PC with Intel Core i5 CPU operating at 3.20 GHz, 24.0 GB of RAM, and Ubuntu 14.04 LTS operating system. Per frame processing time was under 1.15 ± 0.05s and 1.23 ± 0.27s respectively for IVUS and thyroid segmentation respectively, without the need of deployment on any computing accelators like GPUs. In order to process an IVUS volume of 384 x 384 x

100 it takes around 115s and a thyroid US volume of size 372 x 252 x 100 it takes about 123s.

B.6   Conclusion

The approach to layer characterization and subsequent boundary detection using the ultrasonic backscattered signals operating on the freehand ultrasound images is a crucial task. In this paper, we have presented a computationally efficient and robust algorithm to automatically segment anatomical structures and without need for any user-interaction for initializing contours in US image sequences. The algorithm addresses the problem of freehand US volume segmentation using the formulation of a machine learning based model with the help of mathematical models of US backscattering statistical mechanics. Further to this with use of belief propagation across neighboring frames of the volume. The method shows high accuracy while consuming less time and yielding robust contours. Those abilities are in fact a significant advantage over existing tissue classifiers that are not adaptive to patient level variations. The proposed method can directly be leveraged to facilitate online learning or relearning on new cases for improved performance of segmentation without any manual interaction for seeding. The algorithm also demonstrates (1) reliable US layer segmentation from different tissue present in the frame and independent of US application, (2) interframe contour segmentation consistency in the volume, (3) independence to the size and shape of anatomical structure for detection of the contour, (4) ability to segment in the presence of various artifacts in the frame and (5) accurate prediction using a dominant learning model (RF) and less time complexity of per frame processing. These attributes make the method unique and better performing in comparison with stateof-art of US image and volume segmentation.

References

1. J. A. Noble and D. Boukerroui, "Ultrasound image segmentation: a survey," IEEE Trans. Med. Imaging, vol. 25, no. 8, pp. 987-1010, 2006.

2. N. S. Narayan, P. Marziliano, J. Kanagalingam, and C. G. Hobbs, "Speckle patch similarity for echogenicity-based multiorgan segmentation in ultrasound images of the thyroid gland," IEEE J. Biomed. Health Inform., vol. 21, no. 1, pp. 172-183, 2017.

3. B. Saleh, Introduction to subsurface imaging. Cambridge University Press, 2011.

4. D. Sheet, A. Karamalis, A. Eslami, P. Noel, J. Chatterjee, A. K. Ray, Â¨ A. F. Laine, S. G. Carlier, N. Navab, and A. Katouzian, "Joint learning of ultrasonic backscattering statistical physics and signal confidence primal for characterizing atherosclerotic plaques using intravascular ultrasound," Med. Image Anal., vol. 18, no. 1, pp. 103-117, 2014.

5. M. Papadogiorgaki, V. Mezaris, Y. S. Chatzizisis, G. D. Giannoglou, and I. Kompatsiaris, "Automated ivus contour detection using intesity features and radial basis function approximation," in Proc. Int. Symp. Comp. Based Med. Sys., pp. 183-188, 2007.

6. F. Ciompi, O. Pujol, C. Gatta, M. Alberti, S. Balocco, X. Carrillo, J. Mauri-Ferre, and P. Radeva, "Holimab: A holistic approach for media- adventitia border detection in intravascular ultrasound," Med. Image Anal., vol. 16, no. 6, pp. 1085-1100, 2012.

7. S. Balocco, C. Gatta, F. Ciompi, A. Wahle, P. Radeva, S. Carlier, G. Unal, E. Sanidas, J. Mauri, X. Carillo, et al., "Standardized evaluation methodology and reference database for evaluating ivus image segmentation," Comput. Med. Imaging Graph., vol. 38, no. 2, pp. 70- 90, 2014.

8. D. China, M. K. Nag, K. Mandana, A. K. Sadhu, P. Mitra, and C. Chakraborty, "Automated in vivo delineation of lumen wall using intravascular ultrasound imaging," in Proc. Int. Conf., Engg., Med., Biol. Soc., pp. 4125-4128, 2016.

9. D. China, P. Mitra, and D. Sheet, "Segmentation of lumen and external elastic laminae in intravascular ultrasound images using ultrasonic backscattering physics initialized

multiscale random walks," in Proc. Int. Conf. Comput. Vis. Graph. Image Process., pp. 393-403, 2016.

10. D. China, P. Mitra, and D. Sheet, "On the fly segmentation of intravascular ultrasound images powered by learning of backscattering physics," in Classification in BioApps, pp. 351-380, 2018.

11. E. G. Keramidas, D. K. Iakovidis, D. Maroulis, and S. Karkanis, "Efficient and effective ultrasound image analysis scheme for thyroid nodule detection," in Proc. Int. Conf. Image Anal. Recognition, pp. 1052-1060, 2007.

12. D. E. Maroulis, M. A. Savelonas, D. K. Iakovidis, S. A. Karkanis, and N. Dimitropoulos, "Variable background active contour model for computer-aided delineation of nodules in thyroid ultrasound images," IEEE Trans. Inf. Technol. Biomed., vol. 11, no. 5, pp. 537-543, 2007.

13. M. A. Savelonas, D. K. Iakovidis, I. Legakis, and D. Maroulis, "Active contours guided by echogenicity and texture for delineation of thyroid nodules in ultrasound images," IEEE Trans. Inf. Technol. Biomed., vol. 13, no. 4, pp. 519-527, 2009.

14. H. Garg and A. Jindal, "Segmentation of thyroid gland in ultrasound image using neural network,2 in Proc. Int. Conf. Comput. Commun. Netw. Technol., pp. 1-5, 2013.

15. T. Wunderling, B. Golla, P. Poudel, C. Arens, M. Friebe, and C. Hansen, "Comparison of thyroid segmentation techniques for 3d ultrasound," in Proc. Med. Imaging Image Process., vol. 10133, p. 1013317, 2017.

16. R. Shekhar, R. Cothren, D. Vince, S. Chandra, J. Thomas, and J. Cornhill, "Three-dimensional segmentation of luminal and adventitial borders in serial intravascular ultrasound images," Comput. Med. Imaging Graph., vol. 23, no. 6, pp. 299-309, 1999.

17. J. D. Klingensmith, R. Shekhar, and D. G. Vince, "Evaluation of threedimensional segmentation algorithms for the identification of luminal and medial-adventitial borders

in intravascular ultrasound images," IEEE Trans. Med. Imaging, vol. 19, no. 10, pp. 996-1011, 2000.

18. M. H. Cardinal, J. Meunier, G. Soulez, R. L. Maurice, E. Therasse, Â´ and G. Cloutier, "Intravascular ultrasound image segmentation: a threedimensional fast-marching method based on gray level distributions," IEEE Trans. Med. Imaging, vol. 25, no. 5, pp. 590-601, 2006.

19. S. Sun, M. Sonka, and R. R. Beichel, "Graph-based ivus segmentation with efficient computer-aided refinement," IEEE Trans. Med. Imaging, vol. 32, no. 8, pp. 1536-1549, 2013.

20. E. G. Mendizabal-Ruiz, M. Rivera, and I. A. Kakadiaris, "Segmentation of the luminal border in intravascular ultrasound b-mode images using a probabilistic approach," Med. Image Anal., vol. 17, no. 6, pp. 649-670, 2013.

21. F. S. Zakeri, S. K. Setarehdan, and S. Norouzi, "Automatic mediaadventitia ivus image segmentation based on sparse representation framework and dynamic directional active contour model," Comput. bio. med., vol. 89, pp. 561-572, 2017.

22. E. N. Kollorz, D. A. Hahn, R. Linke, T. W. Goecke, J. Hornegger, and T. Kuwert, "Quantification of thyroid volume using 3-d ultrasound imaging," IEEE Trans. Med. Imaging, vol. 27, no. 4, pp. 457-466, 2008.

23. C.Y. Chang, Y. F. Lei, C. H. Tseng, and S. R. Shih, "Thyroid segmentation and volume estimation in ultrasound images," IEEE Trans. Biomed. Engg., vol. 57, no. 6, pp. 1348-1357, 2010.

24. L. Grady, "Random walks for image segmentation," IEEE Trans. Pattern Anal. Mach. Intell., vol. 28, no. 11, pp. 1768-1783, 2006.

25. J. A. Noble, "Ultrasound image segmentation and tissue characterization," Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine, vol. 224, no. 2, pp. 307-316, 2010.

26. P. K. Sen and J. M. Singer, Large sample methods in statistics. Chapman and Hall, Inc., 1993.

27. A. Karamalis, W. Wein, T. Klein, and N. Navab, "Ultrasound confidence maps using random walks," Medical Image Analysis, vol. 16, no. 6, pp. 1101-1112, 2012.

28. L. Breiman, "Random forests," Machine Learning, vol. 45, no. 1, pp. 5- 32, 2001.

Chapter C

Parametrical modelling for texture characterization - a novel approach applied to Ultrasound thyroid segmentation

RESEARCH ARTICLE

Parametrical modelling for texture characterization—A novel approach applied to ultrasound thyroid segmentation

Alfredo Illanes *, Nazila Esmaeili, Prabal Poudel, Sathish Balakrishnan, Michael Friebe

INKA, Institute of Medical Technology, Otto-von-Guericke-Universität Magdeburg, Magdeburg, Germany

* alfredo.illanes@ovgu.de

Abstract

Texture analysis is an important topic in Ultrasound (US) image analysis for structure segmentation and tissue classification. In this work a novel approach for US image texture feature extraction is presented. It is mainly based on parametrical modelling of a signal version of the US image in order to process it as data resulting from a dynamical process. Because of the predictive characteristics of such a model representation, good estimations of texture features can be obtained with less data than generally used methods require, allowing higher robustness to low Signal-to-Noise ratio and a more localized US image analysis. The usability of the proposed approach was demonstrated by extracting texture features for segmenting the thyroid in US images. The obtained results showed that features corresponding to energy ratios between different modelled texture frequency bands allowed to clearly distinguish between thyroid and non-thyroid texture. A simple k-means clustering algorithm has been used for separating US image patches as belonging to thyroid or not. Segmentation of thyroid was performed in two different datasets obtaining Dice coefficients over 85%.

117

Abstract

Texture analysis is an important topic in Ultrasound (US) image analysis for structure segmentation and tissue classification. In this work a novel approach for US image texture feature extraction is presented. It is mainly based on parametrical modelling of a signal version of the US image in order to process it as data resulting from a dynamical process. Because of the predictive characteristics of such a model representation, good estimations of texture features can be obtained with less data than generally used methods require, allowing higher robustness to low Signal-to-Noise ratio and a more localized US image analysis. The usability of the proposed approach was demonstrated by extracting texture features for segmenting the thyroid in US images. The obtained results showed that features corresponding to energy ratios between different modelled texture frequency bands allowed to clearly distinguish between thyroid and non-thyroid texture. A simple k-means clustering algorithm has been used for separating US image patches as belonging to thyroid or not. Segmentation of thyroid was performed in two different datasets obtaining Dice coefficients over 85%.

C.1   Introduction

Texture analysis is the term used for methods developed to quantify image texture through description of image properties by textural features. In general, these features aim to measure smoothness, coarseness, and regularity of pixels, which form an image [1, 2]. Feature extraction methods are usually followed by classification or clustering and can be applied for image segmentation, image characterization, and for estimation of image similarity metrics [3].

Generally used approaches for computing texture features are based on statistical and frequency domain techniques. Statistical approaches compute histograms, entropy, ho-

mogeneity, mean and variance values for estimating features from the texture. Frequency domain techniques or spectral techniques collect a distribution of filter responses to extract different aspects from the texture [1]. Gabor filters and Wavelet decomposition are examples of this type of approach.

The main drawback of these approaches is that they are mainly data-driven, meaning that the computation of texture characteristics is performed directly from the pixel values. With that the estimation of texture characteristics is limited by the amount of data and the Signal to Noise Ratio (SNR) of the image.

In medical imaging, texture describes internal structures of human tissues or organs or pathological changes. Different modalities such as Magnetic Resonance, Computer Tomography and Ultrasound (US) require texture analysis and characterization for applications such as segmentation, registration ans lesion classification [3]. From all these modalities, US is known to be the most challenging because of the presence of characteristic artifacts such as speckles and shadows as well as due to the low SNR and resolution.

In this work, a novel approach for image texture feature extraction in US images is presented mainly based on parametrical modelling. The main idea behind this approach is to analyze the texture as data resulting from a dynamical process and to estimate the different dynamics involved in the texture in order to use mathematical operations between these dynamics as possible texture features. For that a signal version of the image is first computed, where the independent variable is the space, then the signal is decomposed in different frequency bands using Wavelet Transformation and finally an Autoregressive (AR) parametrical model of the decomposed signals provides spectral characteristics used for features computation.

The main advantage of this approach is that good estimations of texture characteristics

119

can be obtained with less data due to the predictive characteristics of the model representation. Moreover, using an appropriate model order noise can be optimally handled and a better estimation of the dynamical properties of the texture can be obtained, even under really complex SNR characteristics as usually seen in US data.

The usability of the proposed approach was demonstrated with US data for segmenting thyroid texture. The obtained results showed that features corresponding to energy ratios between different modelled texture frequency bands allow to clearly distinguish between thyroid and non-thyroid texture regions.

The thyroid is one of the largest endocrine glands in the human body and it is involved in several significant body mechanisms. Diseases of the thyroid gland are among the most frequent endocrine disorders and changes of the thyroid volume are often the symptom of these common pathological conditions. For this reason, it is essential to track and monitor changes on thyroid volume over time and segmentation of the thyroid is one of the main steps for this purpose.

Many approaches have been presented in the literature for extracting features in US thyroid image analysis, mainly for thyroid segmentation and nodule characterization and classification. Recent surveys demonstrate that these two topics for thyroid analysis are highly active research fields [4-7]. Following this trend, many new methods have been proposed in the last years. Concerning thyroid segmentation in [8] three semi-automatic algorithms based on general segmentation approaches such as active contours, graph cut and pixel based classifier were evaluated and compared with two machine learning approaches based on Convolutional Neural Networks and Random Forest (RF). In [9] the segmentation of the thyroid is made by taking into account apriori information based on the physics of the US imaging process and by applying Iterative Random Walks and RF based techniques. Furthermore, several type of features have been proposed for tissue characterization in order to classify nodules or lesions in thyroid US images. Among

the most used features are statistical features [10-12], Spectral-based features [13, 14], higher order statistics based features [15, 16], Wavelet-based features [2, 17, 18] and Fractal-based features [13, 19]. Additionally other works have proposed machine learning algorithms [20, 21] and neutrosophic clustering for thyroid tissue characterization.

As in the general literature for US feature extraction, the main drawback in thyroid US feature computation is that most of the proposed approaches are data driven operating directly from the pixel values of the image. We propose a completely different approach where the preprocessing or image aspects decomposition is made over a signal and not an image and where the features are computed not from the pixels values but from a parametrical model of each estimated image aspect. We believe that the predictive characteristics of such parametrical approach will be able to better deal with the low SNR of thyroid US images and will also allow to obtain better estimation of features with lower quantity of data than direct pixel feature computation.

The main purpose of this paper is not to propose a new thyroid segmentation algorithm but to show how features computed with a completely novel approach can be valuable for US texture characterization. However, to assess the performance of the proposed approach the algorithm was evaluated using two thyroid datasets obtaining Dice coefficients higher than 85% in both databases. Additionally, the results were compared with the ones obtained by other approaches proposed in the literature.

C.2   Methods

As mentioned above, one of the major issues with US imaging is the quality of the data, which particularly affects segmentation applications or texture characterization that are strongly influenced by the relatively low quality of clinical US images, causing that tissue echogenic characteristics and boundaries are often drowned in noise. Decomposition

and parametrization of each characteristic aspect of the data could reduce the noise and enhance the valuable information. These aspects could correspond to different type of noises or artifacts or to different levels of irregularity or granularity in the US image.

The core idea presented in this paper is to treat an US image as texture that can be represented as data resulting from a dynamical process, which depends on space as an independent variable and whose dynamical patterns can then characterize such a texture. These dynamics can be modelled using a parametrical approach and the estimated parameters can be taken as a mathematical representation of the texture that are used to compute valuable features that characterize the US texture at a given location.

Fig. C.1 illustrates the basic idea. A thyroid ultrasound image is shown and a sub-image or patch (red box) is selected in such a way that it contains thyroid (Texture 1 in the figure) and non-thyroid textures (Texture 2 in the figure). The boundary between the two textures (thyroid and non-thyroid) is not evident but in the mesh representation of the sub-image on the top right of Fig. C.1 it is possible to visualize the different texture characteristics of the two tissue types. If we extract a line profile passing through both textures (red dashed line in the US image) then it is possible to verify that the texture signals (inside the dashed rectangle over the line profile plot) involve different frequency components or more general, different signal dynamics that are characteristics of each texture.

The idea is now to model these texture dynamics using a parametrical approach to perform features computation not by operating the matrix data itself, but by operating the parameters of the modelled texture that represent the information that the image contains in terms of dynamical distribution. By using an optimal model order such an approach can be highly robust to the typical speckle noise of US images as well as to low trend intensity inhomogeneity. Additionally, because of the predictive characteristics of
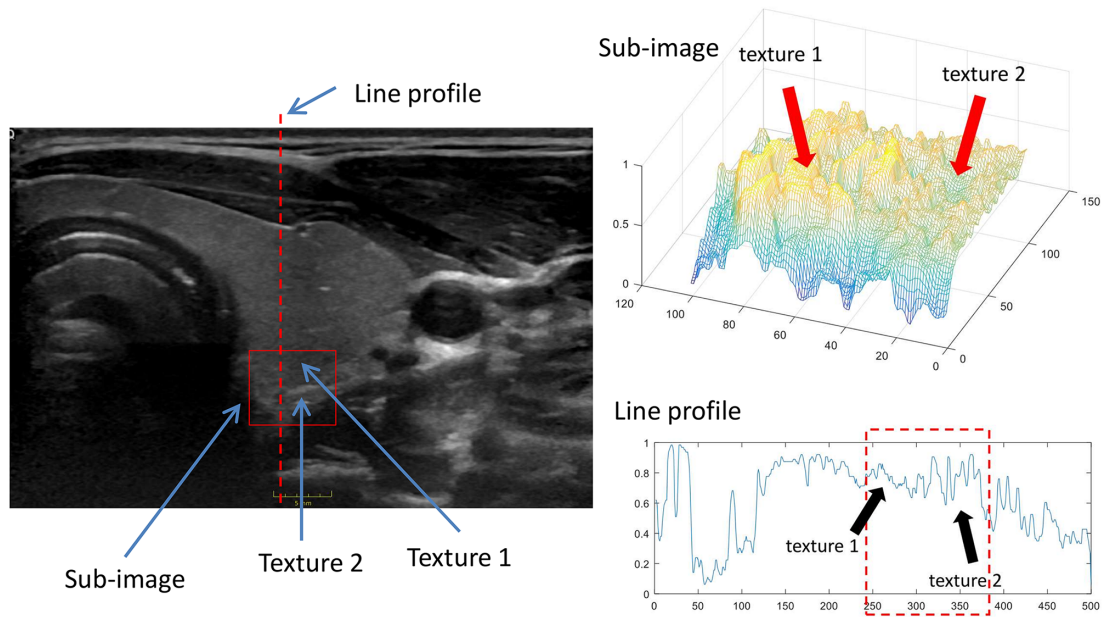
Fig. C.1: Illustration of the main principle behind the proposed US texture characterization approach.

such a model representation, good estimations of characteristics of a texture can be obtained with less data than with standard methods.

The block diagram of Fig. C.2 displays the main steps for feature computation of a US image patch in order to characterize its texture. First, the image patch gray-level matrix is converted into four texture signals using two different image to signal conversion procedures. Then, each of the texture signals is decomposed in four signal bands using Continuous Wavelet Transformation (CWT). The 16 resulting narrow-band texture signals are then modelled using an Autoregressive (AR) parametrical model to finally compute features from ratios between different energy bands of the decomposed signals. In the following, each step of the algorithm will be detailed.
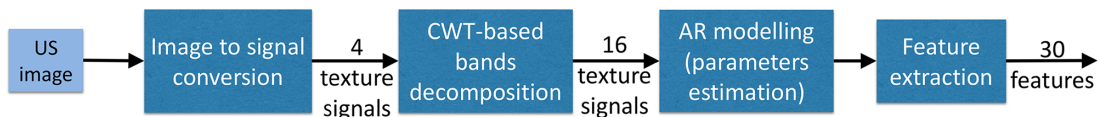
Fig. C.2: Main steps of the general concept of the signal processing algorithm for texture modelling and feature extraction in US images.

123

C.2.1   Image to Signal Conversion

In order to track dynamical texture characteristics of an US image resulting from a dynamical process, the matrix data is first converted into a signal. For that we use ZigZag (following the rows direction) and spiral conversion of the US matrix image and of their 90 degrees rotation matrix versions (see Fig. C.3). The output of this first step results in four texture signals, one per each conversion (ZigZag, spiral and their 90 degrees conversions).
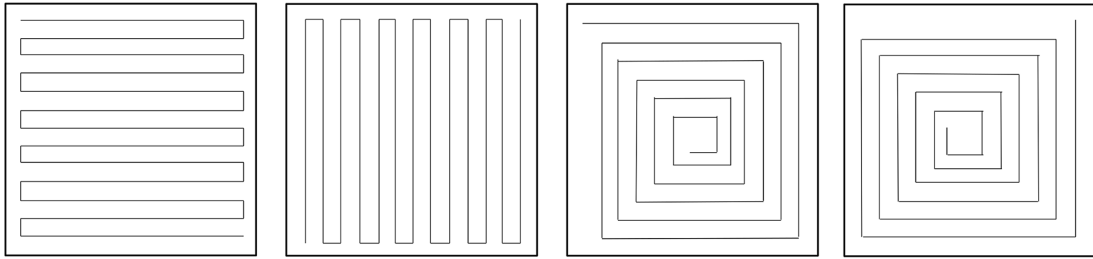
Fig. C.3: Conversion of a matrix by traversing the matrix and its transposed in ZigZag and in spiral.

C.2.2   Continuous wavelet texture frequency band decomposition

The second step decomposes each one of the four texture signals in several frequency bands, each containing one different aspect of the texture. We assume that an image texture is composed of several dynamics representing irregularity characteristics of the texture such as smoothness or roughness. Therefore the signals can be decomposed into several dynamics that can represent levels of irregularities presented in the image/texture. Each signal is separated in different frequency components or scales and we then reconstruct several narrow band signals that should contain information of different levels of texture irregularity.

Since the four signals resulting from the image to signal conversion step can contain components that are not necessarily oscillatory, they are decomposed using scale decomposition instead of frequency Fourier- based decomposition. For that the CWT was applied to decompose the signal in different scales and then reconstructing new signals using scales equivalent to three frequency bands representing low, middle and high frequency components (LF, MF and HF) using a Daubechies mother wavelet. Additionally a fourth frequency band called Total Band (TB) was computed using the full frequency band of the signals but erasing the Very Low Frequency components Wavelet scales, which correspond to low trend image intensity inhomogeneity.

In summary this step results in 16 texture narrowband signals, four per texture signal. They are denoted in the sequel as $y_i^B[n]$, where n represent the discrete index (independent variable), $i = 1, 2, 3, 4$ denotes the signal conversion type (1, 2 for ZigZag and its rotated version, 3, 4 for spiral and its rotated version) and B denotes the type narrowband signal LF, MF, HF or TB.

Fig. C.4 shows an example of CWT decomposition of three texture signals (ZigZag non rotated conversion versions) at different positions in the US image denoted I_{US1}, $IUS2$ and $IUS3$ in the figure. Two patches, I_{US1} and I_{US2}, were taken from inside the thyroid and the patch I_{US3} was taken outside the thyroid, but in a region with similar texture characteristics as thyroid. At the right side of Fig. C.4 the CWT spectrum concatenated for the three patches is displayed. It is possible to visualize in the spectrum the scale frequency bands that were taken for constructing the narrow-band texture signals HF, MF and LF.

The TB, HF, MF and LF bands for the ZigZag signal version for each patch are shown also at the right of Fig 4 by concatenating the resulting signals in order to observe the dynamical difference between textures at different image locations. It is possible to observe that for I_{US1} and I_{US2}, belonging both to patches located inside the thyroid,
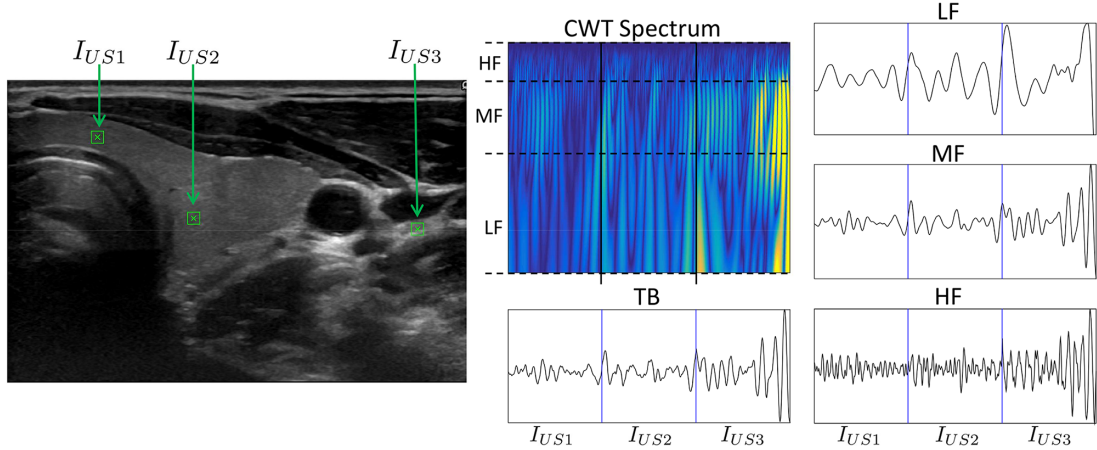
Fig. C.4: Example of a CWT decomposition of a thyroid US image when three image patches are taken from different locations of the US images.

the dynamics of the decomposed signals are similar in MF, HF and TB, while the dynamics resulting from I_{US3} are clearly different. Particularly in HF it is evident that the frequency components and the amplitude are different inside compared to outside the thyroid. In LF even if the difference is low we can observe tiny changes in amplitude when we compare inside and outside thyroid patches. Given this analysis, what we want is to quantify these dynamical differences between image textures inside and outside the thyroid, and as explained in the next section, this will be done using a parametrical model of the different extracted texture signals.

C.2.3   Ultrasound texture parametrical modelling

The size of a patch should be small enough in order to perform highly localized texture feature characterization. This requirement results in two characteristics of the data that a texture characterization method must deal with: the small quantity of dynamical changes (texture variability involving limited number of oscillations or damped oscillations) and the under-sampled characteristics of the data due not only to the size of the patch, but also to the resolution of the US image modality. Moreover, a texture

characterization method should also be able to deal with the low SNR characteristics of US imaging. Under these conditions, classical methods for feature extraction, such as spectral or statistical based ones and in general data-driven approaches, no longer can obtain a good estimation of texture characteristics. This is why we propose in this work to use parametrical modelling of the resulting 16 texture signals from the CWT decomposition. These signals present narrow-band characteristics that are well suited to be modelled with an autoregressive (AR) approach. Our approach consist of a parametric representation of signal dynamics, which can deal with the drawbacks of the generally used current methods.

AR modelling is a well-known and well published technique for parametrical spectral estimation that has shown advantages over non-parametrical based methods (for detailed information about AR modelling we suggest [22]). The advantages of the AR representation is that it is possible to obtain good estimation of the spectrum and higher spectral resolution using less data than classical methods and that it provides a parametric way to analyze the data.

The AR model for each one of the sixteen $y_i^B[n]$ texture signals consists of a linear combination of past samples of the respective signal and a white zero mean noise $e[n]$ of variance σ^2:

$$y_i^B[n] = -\sum_{k=1}^{p_B} a_{ik}^B y_i^B[n-k] + e[n] \tag{1}$$

where $a_{ik}^B\{k = 1, 2, ..., p_B\}$ are the estimated AR parameters for the narrowband signal $y_i^B[n]$. In this work the model order p_B is dependent on the band B and was set on 100, 50, 30 and 80 respectively for $B = $ LF, $B = $ MF, $B = $ HF and $B = $ TB. The AR parameters were estimated using the Yule-Walker method [22].

C.2.4 Feature extraction and selection procedure

From equation (1) power spectral densities can be computed for each one of the 16 narrowband signals. Since the textural dynamics should be different from one type of tissue to another, the main estimated components of the spectrum should vary between different tissues. Additionally, the different degrees of texture irregularities in a given tissue should be manifested by different contributions on energy of the frequency bands characterizing the texture signal. Therefore features related to spectral energy of the 16 texture signals (resulting after CWT decomposition) are used in this work.

To compute spectral energy features the Power Spectral Densities $S_i^B(f)$ for each one of the 16 narrowband signals have to be first computed from the AR parameters. For that the Z-transform can be applied to equation (1) and then AR spectrum can be computed from the resulting transfer function:

$$S_i^B(f) = \frac{1}{|1 + \sum_{k=1}^{p_B} a_{ik}^B e^{j2\pi fn}|^2} \tag{2}$$

Fig. C.5 shows an example of the information that the AR spectrum can provide for characterizing tissue. The AR spectra of patches I_{US1}, I_{US2} and I_{US3} for LF, MF, HF and TB signals of Fig. C.4 are displayed. It is possible to observe that for both patches located inside the thyroid (I_{US1} and I_{US2}) the AR spectra (in blue and red lines respectively) main components are similar for the four narrowband texture signals. For the patch located outside the thyroid (I_{US3}) the AR spectral characteristics are completely different in terms of main frequency components and spectral energy.

These parametrical characteristics can be exploited using the spectral energy of the estimated spectra. Therefore, the features computed in this novel approach are based on band energy ratios computed between the different frequency bands of the AR spectra.
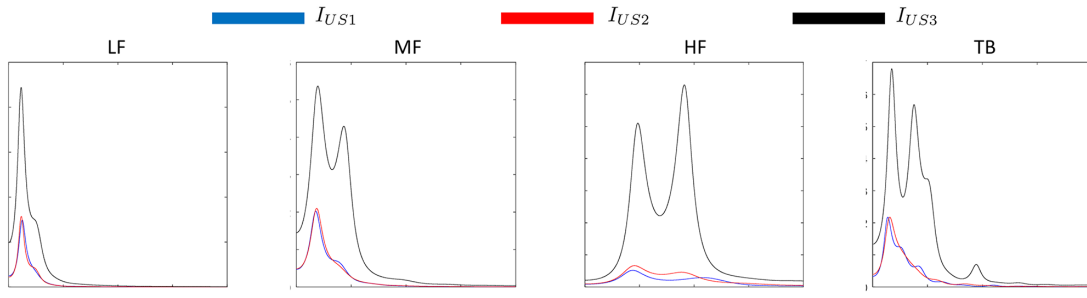
Fig. C.5: AR spectra for patches I_{US1}, I_{US2} and I_{US3} (in blue, red and black lines respectively) for the four narrowband signals belonging to the ZigZag matrix to signal conversion.

These type of AR features have already been used for Heart Rate Variability analysis for extraction of dynamical relationships between sympathetic and parasympathetic activities of the Autonomous Nervous System [23]. Analog to that, we assume that in one US texture the relationship between its different levels of irregularity can provide information for classifying tissue echogenicity.

Taking into account the number of spectra from the different texture signals belonging to the different conversions, the number of possible spectral energy ratios to be used as potential features is 256. Therefore a procedure of features selection has been performed using an analytic test. First, inconsistent energy ratios were eliminated from the analysis. Then different US images were selected from a US thyroid image dataset (Dataset 1, which will be introduced in the next section). For each image 100 patches (20 x 20 pixels) were selected manually, 50 patches belonging to thyroid and the other 50 belonging to non-thyroid regions. In order to evaluate the ability of a feature to distinguish between thyroid and non-thyroid texture, the patches belonging to regions outside the thyroid were divided in three classes according to the visual level of texture similarity that a non-thyroid region has with respect to thyroid texture: similar, semi-similar and dissimilar. The set of energy ratio features were computed for all the selected patches and the results were plotted using a color-map matrix as shown in Fig. C.6. Each col-

umn of this matrix correspond to an energy ratio feature and allows to visually analyze the ability of an energy ratio feature to distinguish between thyroid and non-thyroid tissues for different levels of similarity degrees.
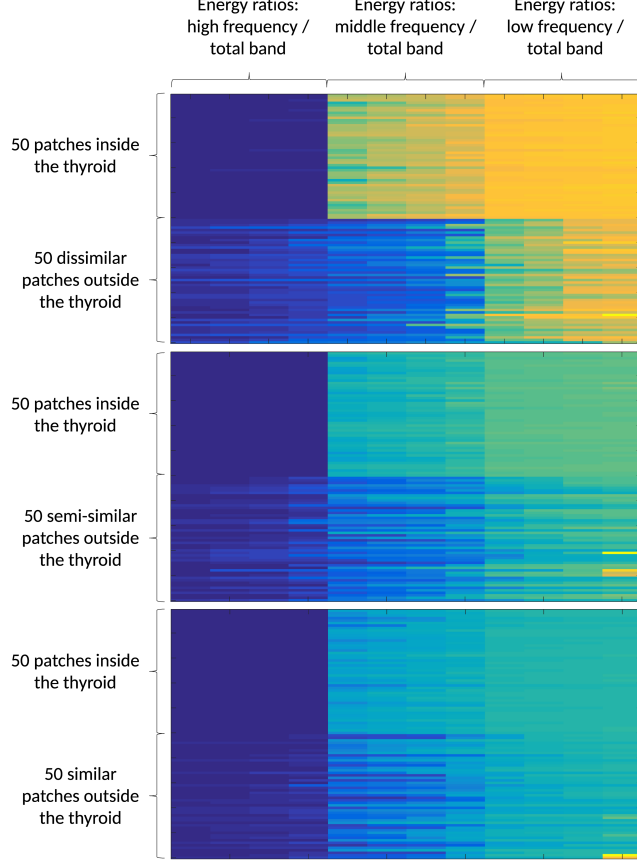
Fig. C.6: Color-map pf the computed features in patches belonging to thyroid and three classes of non-thyroid regions.

Following the analysis of the matrix of Fig. C.6, 30 energy ratios were selected from the previous analysis. According to their characteristics, they can be divided into two types of energy ratios (ER). 4 ERs are computed as the energy of the maximal spectral peak divided by the total spectral energy in a same frequency band and 26 ERs are computed as ratios between total energy of different frequency bands:

$$ER_{1-4} = \frac{\sum_{f=f1}^{f=f2} S_{NUM}(f)}{\sum_{f=0}^{f=\infty} S_{DEN}(f)} \qquad ER_{5-30} = \frac{\sum_{f=0}^{f=\infty} S_{NUM}(f)}{\sum_{f=0}^{f=\infty} S_{DEN}(f)} \tag{3}$$

where $f1$ and $f2$ are the frequency onset and offset respectively of the main peak and S_{NUM} and S_{DEN} are the AR spectra used in the numerator and denominator of equations (3) and are shown in Table C.1

Table C.1: Spectra used in the numerator (NUM) and denominator (DEN) of equation (3) for computing the 30 energy ratio features

$Features_{1-15}$	NUM	DEN	$Features_{16-30}$	NUM	DEN
ER_1	S_1^{HF}	S_1^{HF}	ER_{16}	S_2^{LF}	S_2^{HF}
ER_2	S_1^{MF}	S_1^{MF}	ER_{17}	S_2^{LF}	S_3^{HF}
ER_3	S_1^{LF}	S_1^{LF}	ER_{18}	S_2^{LF}	S_4^{HF}
ER_4	S_1^{TF}	S_1^{TB}	ER_{19}	S_3^{LF}	S_2^{HF}
ER_5	S_1^{HF}	S_1^{TB}	ER_{20}	S_3^{LF}	S_3^{HF}
ER_6	S_2^{HF}	S_1^{TB}	ER_{21}	S_3^{LF}	S_4^{HF}
ER_7	S_1^{MF}	S_1^{TB}	ER_{22}	S_4^{LF}	S_4^{HF}
ER_8	S_2^{MF}	S_2^{TB}	ER_{23}	S_4^{LF}	S_3^{HF}
ER_9	S_2^{LF}	S_2^{TB}	ER_{24}	S_4^{LF}	S_4^{HF}
ER_10	S_3^{LF}	S_3^{TB}	ER_{25}	S_1^{MF}	S_2^{HF}
ER_11	S_4^{LF}	S_4^{TB}	ER_{26}	S_1^{MF}	S_3^{HF}
ER_12	S_1^{LF}	S_2^{HF}	ER_{27}	S_1^{MF}	S_4^{HF}
ER_13	S_1^{LF}	S_3^{HF}	ER_{28}	S_2^{MF}	S_2^{HF}
ER_14	S_1^{LF}	S_4^{HF}	ER_{29}	S_2^{MF}	S_3^{HF}
ER_15	S_2^{LF}	S_1^{HF}	ER_{30}	S_2^{MF}	S_4^{HF}

The feature extraction algorith was fully implemented in Matlab R2015b and executed on a PC with a CPU operating at 2.60 GHz resulting in an execution time of 0.06 seconds for computing the 30 features in one patch.

C.3   Results

This section shows the usability of the proposed approach for US feature extraction. 2D US data from thyroid is used in order to analyze the capabilities of the 30 extracted ER

features to differentiate between thyroid and non-thyroid tissue in order to use them for segmenting thyroid.

C.3.1   Thyroid US Data Description

Two different real US image datasets have been used to evaluate the proposed approach. The first dataset (in the sequel Dataset 1) has been introduced in [24] and involves six healthy human subjects freehand US images acquired using a Logiq E9 US device with a linear probe and equipped with an electromagnetic tracking system. This database has a total of 675 2D US slices with a 760 x 500 pixels with between 53 and 189 US slices per subject. The second dataset (in the sequel Dataset 2) has been presented in [25] and can be downloaded in http://opencas.webarchiv.kit.edu/?q=node/29. It involves freehand US images of 16 healthy subjects, each acquired also with a GE Logiq E9 system but operated by a different clinician in a different hospital than in the Database 1 case. From this dataset, a total of 1600 slices belonging to the 16 subjects (100 slices per subject) were used with a size of 760 x 1020 pixels per 2D US slice.

For each 2D slice the thyroid was manually segmented by an expert clinician (ground truth) and was then divided into patches of 20 x 20 pixels labelled as thyroid or non-thyroid according to the ground truth. It is important to notice that in both datasets the number of patches belonging to thyroid are less that the ones belonging to non-thyroid. This is because the ground truth was used for the automatic patch labelling and usually in a US image the region of thyroid is smaller than the non-thyroid one.

C.3.2 Average value differences between thyroid and non-thyroid patches for the selected features

In order to observe the capacity and suitability of the selected features for distinguishing between thyroid and non-thyroid texture, the average and standard deviation (STD) of the feature values were computed for the six subjects belonging to Dataset 1 (see Fig. C.7 and Fig. C.8). It is possible to visualize that for the whole set of selected ER features, the average values are clearly different between thyroid (red) and non-thyroid (blue) tissues. Moreover, in most of the ER features the thyroid and non-thyroid average values do not strongly change from one subject to another one. Concerning the STD, it is possible to observe that some ER features works better than others. This is the case for example of features ER_3 and ER_6 where the STDs inside the thyroid are much smaller than outside the thyroid, what is consistent to the homogeneity of texture inside one healthy organ.

C.3.3 Features evaluation for thyroid segmentation

The proposed approach have been tested on the 359712 patches of Dataset 1 and on the 1791397 patches of Dataset 2. For each patch, the 30 ERs of Eq (3) were computed and analyzed to see their suitability to distinguish between thyroid and non-thyroid tissues.

In Fig. C.9 3D scatters are displayed for 12 ERs (in groups of three features) computed from Dataset 1 clearly showing the differences of these ratio values for thyroid (in red) and non-thyroid (in blue). This confirms that the AR characterization of US texture is well suited to obtain features that are able to be used for classification of thyroid tissue.

To evaluate the performances of our approach, the computed ER features were used to segment the thyroid. Because our goal is to show the usability of the extracted AR
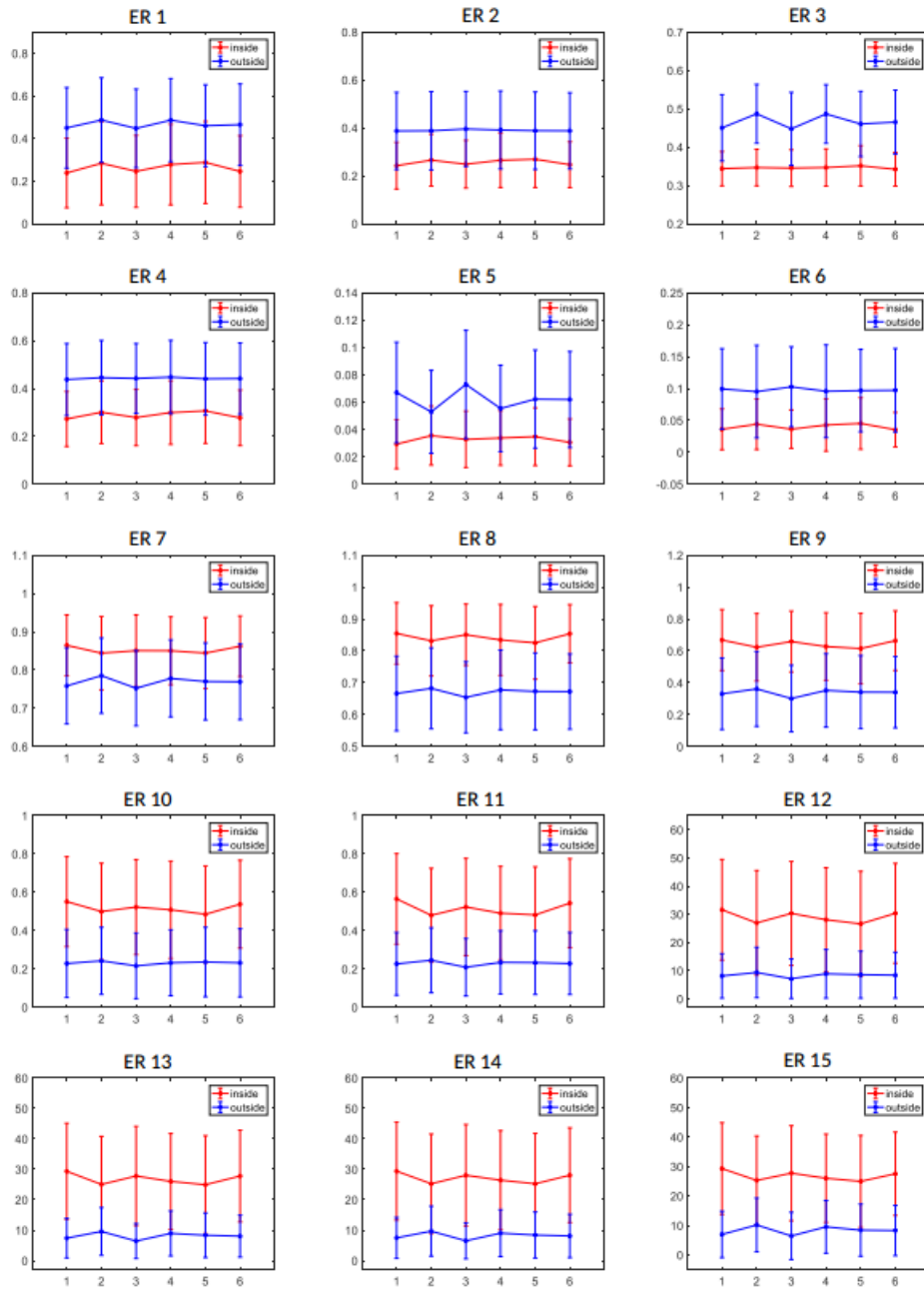
Fig. C.7: Mean and standard deviation of values of ERs features 1 to 15 of thyroid and non-thyroid patches for the 6 subjects of the Dataset 1.

features for texture characterization in US images, complex classification procedures were avoided. Therefore only a simple K-Means algorithm for clustering the patches as thyroid or non-thyroid using the 30 features computed in both datasets was used in
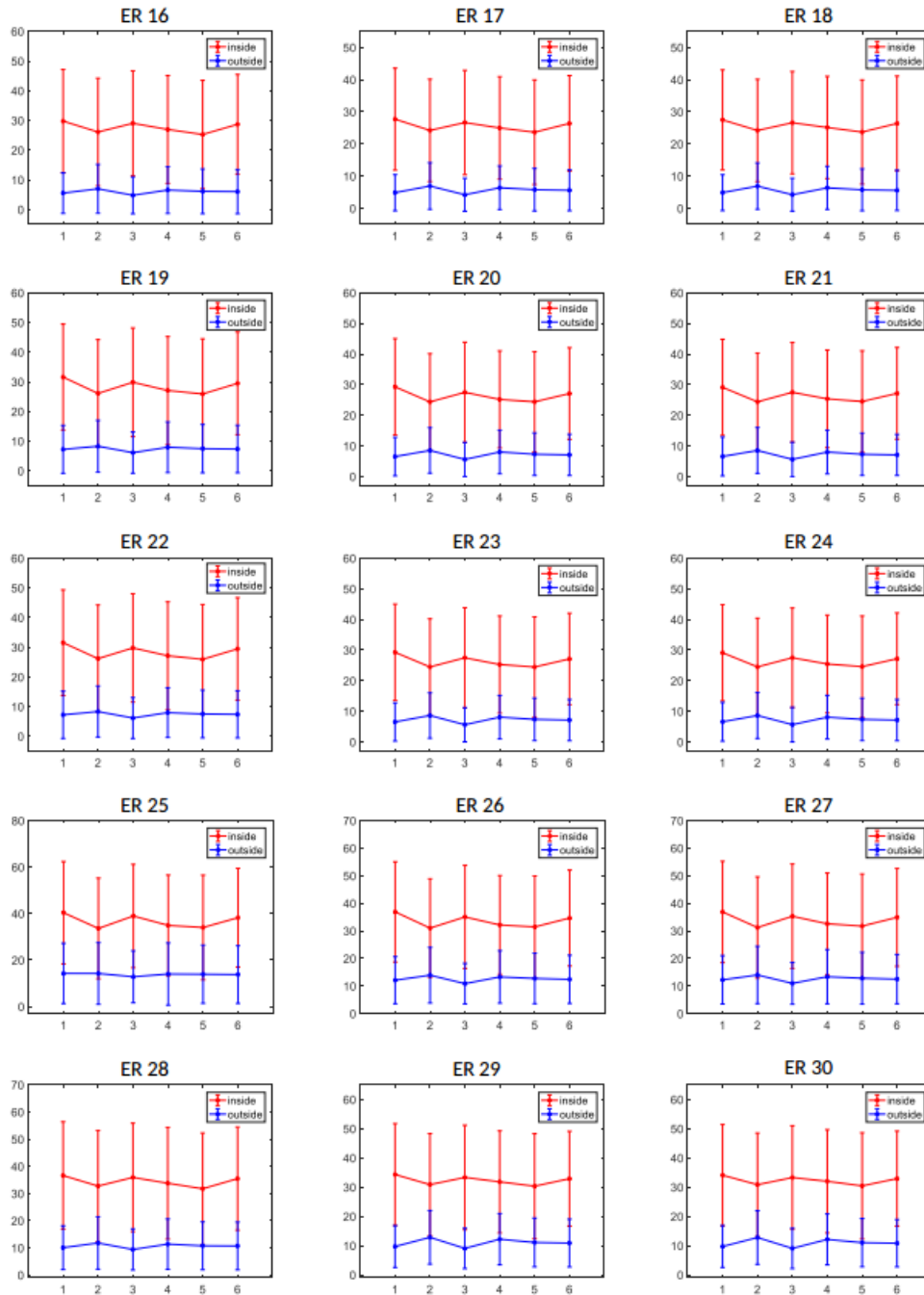
Fig. C.8: Mean and standard deviation of values of ERs features 16 to 30 of thyroid and non-thyroid patches for the 6 subjects of the Dataset 1.

this work. As an unsupervised classification for thyroid segmentation, this automatic labelling was then used in each US image to separate the patches belonging to that
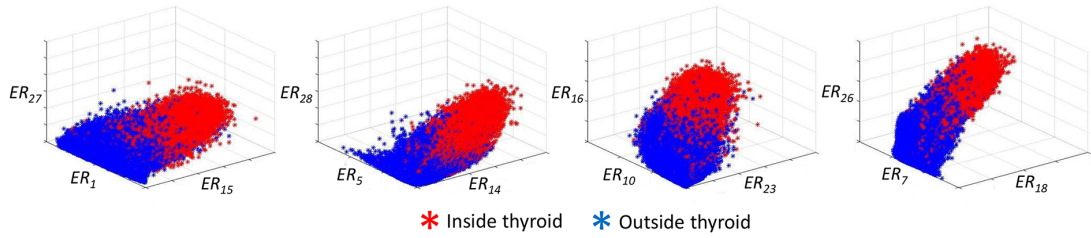
Fig. C.9: Example of obtained AR spectral energy ratios when the approach is applied to the full set of patches extracted from the thyroid US Dataset 1.

image as thyroid and non-thyroid.

Fig. C.10 shows some example results of the thyroid segmentation for eight slices: four examples for correct segmentation (first row) and four involving some false positives (second row). In solid red line the ground truth is displayed and the green squares correspond to the 20 x 20 patches that were classified as thyroid by applying our approach to each one of the US images. We displayed US images belonging to different subjects and also to different positions of the 2D slices with respect to the thyroid volume.
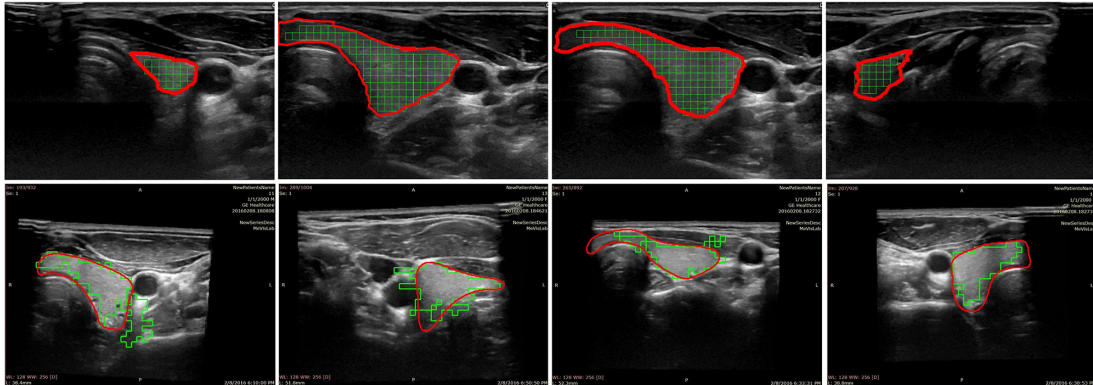
Fig. C.10: Examples of thyroid segmentation using the proposed approach and comparison with the ground truth.

In order to globally evaluate our approach with both datasets, the Dice coefficient (DC) is computed from all the 2D segmentations for each Dataset. Additionally, the sensitivity (SE) and specificity (SP) are also computed. The approach obtains a DC of 89.66% with a SE of 0.95 and a SP of 0.70 for the Dataset 1 and a DC of 86.89% with a SE of 0.89

and a SP of 0.62 for the Dataset 2.

In order to analyze the significance of our results, the proposed approach was compared with other thyroid segmentation methods proposed in the literature. For that the comparison results reported in [8] and in [9] are used.

In [8] five thyroid segmentation algorithms are compared using ten subjects, where six of them are taken from Dataset 1 used in this work. The algorithms that this work compares in terms of Dice Coefficient are Active Contour Without Edges (ACWE), Graph Cut (GC), Pixel-Based Classifier (PBC), Random Forest Classifier (RFC) and Convolutional Neural Network (CNN). The first three are semi-automatic requiring different levels of interaction with the operator and the other two are automatic. The results of this comparison are displayed in Table C.2 showing that our approach outperforms the other five algorithms.

Table C.2: Comparison of the proposed approach in terms of Dice Coefficient using the Dataset 1 with algorithms compared in [8].

ACWE	GC	PBC	RFC	CNN	this work
80.53%	74.52%	66.68%	85.53%	87.22%	**89.66%**

In [9] an algorithm based on Iterative Random Walks and Random Forest (IRWRF) was evaluated with the Dataset 2 used in this work. They have compared their approach with four other algorithms presented in the literature: Echogenicity-based Quantization (EBQ), Joint Classification-Regression (JCR), RBF Neural Network (RBF), and Feedforward Neural Network (FNN). However the reported results for the other algorithms do not use the same dataset. Despite this fact, we display the results of this comparison in Table C.3. The algorithm were compared in terms of DC, SE and SP.

Table C.3: Comparison of the proposed approach using Database 2 with five algorithm results reported in [9].

	IRWRF	EBQ	JCR	RBF	FNN	**this work**
DC	85.4%	83.9%	47.9%	51.2%	40.0%	**86.9%**
SE	98.9%	95.5%	56.4%	87.4%	47.3%	**89.0%**
SP	92.3%	88.9%	92.6%	56.0%	86.4%	**62.0%**

C.4 Conclusions and discussions

In this work a novel approach for ultrasound image feature extraction was presented. The approach is based on characterizing ultrasound texture through parametrical modelling. The image was transformed into a signal, which was decomposed in several dynamics representing different aspects of the texture. We showed that features consisting on frequency band based energy ratios between the different signal dynamics contain valuable information about texture and can be useful for US image texture classification.

The usability of the proposed approach was demonstrated in US thyroid segmentation. The 30 extracted AR features computed from energy ratios of the parametrical AR spectra obtain very good and reproducible results for differentiating thyroid and non-thyroid regions in US images. Using a simple K-Means procedure we demonstrated that thyroid patches were successfully clustered for thyroid segmentation. The approach was evaluated with two datasets and compared with ten other algorithms proposed in the literature, obtaining Dice Coefficients over 85

We strongly believe that this approach can be used in a variety of US applications, not only for segmentation, but also for data comparison, pattern recognition and possibly others. The presented research contribution and scientific innovation could lead to an objective characterization and differentiation of tissues in US, but likely also be used for other bio-medical imaging methods.

One of the drawbacks of the proposed approach is that edges between two tissues are prone to segmentation errors. This is due to the patch approach that we have used. In order to deal with this problem the next steps is to implement a *space-variant* AR modelling, analogue to the time-variant version generally used for non-stationary signal processing. This not only should deal with tissue border problems but also should allow to process *signal trajectories* in a volume in order to perform voxel characterization.

The main focus of this paper was not on thyroid segmentation but on analyze the usability of features that have never been used in the literature for US image analysis. These novel proposed features, even if they are linear, they have obtained interesting results in signal processing in application fields such as biosignal processing or tool wear monitoring. We wanted to analyze how this type of features could work for extracting characteristics from US images. However, in the near future many aspect of this research should be treated in order to think in clinical significant results for US texture characterization. First, a more exhaustive analysis and optimization needs to be performed for AR features selection. It is not only necessary to revise the spectral energy based features but also to see how other AR features (such as pole-based or space variant features) can be used for US texture characterization. Second, a next step should focus on analyze how AR features together with nonlinear features (such as higher order statistics or entropy-based features) and a deep learning procedure can work not only for thyroid segmentation but also for thyroid lesion classification. Finally, in order to show clinical significance, further research is required in order to see the behaviour of the proposed features in larger datasets involving unhealthy thyroids, different US acquisition parameters, different probes and US devices. This is the main next step for our approach.

References

1. Mirmehdi M. Handbook of texture analysis. Imperial College Press; 2008..

2. Acharya U, Faust O, Sree SV, Molinari F, Garberoglio R, Suri J. Cost-effective and non-invasive automated benign and malignant thyroid lesion classification in 3D contrast-enhanced ultrasound using combination of wavelets and textures: a class of ThyroScan algorithms. Technology in cancer research and treatment. 2011;10(4):371-380.

3. Castellano G, Bonilha L, Li L, Cendes F. Texture analysis of medical images. Clinical radiology. 2004;59(12):1061-1069. pmid:15556588

4. Sheeja A, Babu SS. Thyroid Segmentation on US Medical Images: An Overview. International Journal of Emerging Technology and Advanced Engineering. 2012;2(12).

5. Koundal D, Gupta S, Singh S. Survey of computer-aided diagnosis of thyroid nodules in medical ultrasound images. In: Advances in Computing and Information Technology. Springer; 2013. p. 459-467.

6. Acharya UR, Swapna G, Sree SV, Molinari F, Gupta S, Bardales RH, et al. A review on ultrasound-based thyroid cancer tissue characterization and automated classification. Technology in cancer research and treatment. 2014;13(4):289-301.

7. Sollini M, Cozzi L, Chiti A, Kirienko M. Texture analysis and machine learning to characterize suspected thyroid nodules and differentiated thyroid cancer: Where do we stand? European journal of radiology. 2018;99:1-8. pmid:29362138

8. Poudel P, Illanes A, Sheet D, Friebe M. Evaluation of Commonly Used Algorithms for Thyroid Ultrasound Images Segmentation and Improvement Using Machine Learning Approaches. Journal of healthcare engineering. 2018;2018. pmid:30344990.

9. Debarghya C, Illanes A, Poudel P, Friebe M, Mitra P, Sheet D. Anatomical Structure Segmentation in Ultrasound Volumes using Cross Frame Belief Propagating Iterative Random Walks. IEEE journal of biomedical and health informatics. 2018.

10. Iakovidis DK, Keramidas EG, Maroulis D. Fusion of fuzzy statistical distributions for classification of thyroid ultrasound patterns. Artificial Intelligence in Medicine.

2010;50(1):33-41. pmid:20427164.

11. Selvathi D, Sharnitha V. Thyroid classification and segmentation in ultrasound images using machine learning algorithms. In: Signal Processing, Communication, Computing and Networking Technologies (ICSCCN), 2011 International Conference on. IEEE; 2011. p. 836-841.

12. Koprowski R, Korzynska A, Wrobel Z, Zieleznik W, Witkowska A, Malyszek J, et al. Influence of the measurement method of features in ultrasound images of the thyroid in the diagnosis of Hashimoto's disease. Biomedical engineering online. 2012;11(1):91. pmid:23190930.

13. Acharya UR, Sree SV, Krishnan MMR, Molinari F, Garberoglio R, Suri JS. Non-invasive automated 3D thyroid lesion classification in ultrasound: a class of ThyroScan systems. Ultrasonics. 2012;52(4):508-520. pmid:22154208.

14. Acharya UR, Chowriappa P, Fujita H, Bhat S, Dua S, Koh JE, et al. Thyroid lesion classification in 242 patient population using Gabor transform features from high resolution ultrasound images. Knowledge-Based Systems. 2016;107:235-245.

15. Acharya UR, Sree SV, Swapna G, Gupta S, Molinari F, Garberoglio R, et al. Effect of complex wavelet transform filter on thyroid tumor classification in three-dimensional ultrasound. Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine. 2013;227(3):284-292.

16. Raghavendra U, Gudigar A, Maithri M, Gertych A, Meiburger KM, Yeong CH, et al. Optimized multi-level elongated quinary patterns for the assessment of thyroid nodules in ultrasound images. Computers in biology and medicine. 2018;95:55-62. pmid:29455080.

17. Acharya UR, Faust O, Sree SV, Molinari F, Suri JS. ThyroScreen system: high resolution ultrasound thyroid image characterization into benign and malignant classes using novel combination of texture and discrete wavelet transform. Computer methods

and programs in biomedicine. 2012;107(2):23-241. pmid:22054816.

18. Ardakani AA, Gharbali A, Mohammadi A. Classification of benign and malignant thyroid nodules using wavelet texture analysis of sonograms. Journal of Ultrasound in Medicine. 2015;34(11):1983-1989. pmid:26396168. 19. Raghavendra U, Acharya UR, Gudigar A, Tan JH, Fujita H, Hagiwara Y, et al. Fusion of spatial gray level dependency and fractal texture features for the characterization of thyroid lesions. Ultrasonics. 2017;77:110-120. pmid:28219805

20. Ozyilmaz L, Yildirim T. Diagnosis of thyroid disease using artificial neural network methods. In: Neural Information Processing, 2002. ICONIP'02. Proceedings of the 9th International Conference on. vol. 4. IEEE; 2002. p. 2033-2036.

21. Chi J, Walia E, Babyn P, Wang J, Groot G, Eramian M. Thyroid nodule classification in ultrasound images by fine-tuning deep convolutional neural network. Journal of digital imaging. 2017;30(4):477-486. pmid:28695342.

22. Manolakis DG, Ingle VK, Kogon SM. Statistical and adaptive signal processing: spectral estimation, signal modeling, adaptive filtering, and array processing. McGraw-Hill Boston; 2000.

23. Acharya UR, Joseph KP, Kannathal N, Min LC, Suri JS. Heart rate variability. In: Advances in cardiac signal processing. Springer; 2007. p. 121-165.

24. Poudel P, Ataide E, Illanes A, Friebe M. Linear Discriminant Analysis and K-Means Clustering for Classification of Thyroid Texture in Ultrasound Images. In: Proc IEEE Eng Med Biol Soc. Honolulu, USA; 2018.

25. Wunderling T, Golla B, Poudel P, Arens C, Friebe M, Hansen C. Comparison of thyroid segmentation techniques for 3D ultrasound. In: Medical Imaging 2017: Image Processing. vol. 10133. International Society for Optics and Photonics; 2017. p. 1013317.

CHAPTER D

Higher Order Statistical Analysis for Thyroid Texture Classification and Segmentation in 2D Ultrasound Images

Higher Order Statistical Analysis for Thyroid Texture Classification and Segmentation in 2D ultrasound Images

Naghmeh Mahmoodian[1], Prabal Poudel[1], Alfredo Illanes [1] and Michael Friebe[1]

Abstract—Ultrasound (US) imaging is one of the most cost-effective imaging modality that utilizes sound waves for generating medical images of anatomical structure. However, the presence of speckle noise and low contrast in the US images makes it difficult to use for proper classification of anatomical structures in clinical scenarios. Hence, it is important to devise a method that is robust and accurate even in the presence of speckle noise and is not affected by the low image contrast. In this work, a novel approach for thyroid texture characterization based on extracting features utilizing higher order spectral analysis (HOSA) was used. A Support Vector Machine (SVM) was applied on the extracted features to classify the thyroid texture. Since HOSA is a well suited technique for processing non-Gaussian data involving non-linear dynamics, good classification of thyroid texture can be obtained in US images as they also contain non-Gaussian Speckle noise and non-linear characteristics. A final accuracy of 93.27% was achieved using the proposed approach.

HOSA method has been used for distinguishing between benign and malignant nodules in thyroid US images. These approaches involved a pre-processing step where the US images were enhanced using complex wavelet transform and finally HOSA was used to extract the features [4], [5]. However, up to our knowledge, all these approaches were used for classifying the thyroid nodules only but not the thyroid texture itself from the non-thyroid textures. Similarly, either frequency or entropy based features were used in these works to classify the thyroid nodules.

In this work, we propose a novel feature extraction technique that combines the energy, frequency and entropy based features to classify the thyroid textures. US images contain speckle noise which is known to have a non-Gaussian distribution and moreover the formation of these

Abstract

Ultrasound (US) imaging is one of the most cost-effective imaging modality that utilizes sound waves for generating medical images of anatomical structure. However, the presence of speckle noise and low contrast in the US images makes it difficult to use for proper classification of anatomical structures in clinical scenarios. Hence, it is important to devise a method that is robust and accurate even in the presence of speckle

noise and is not affected by the low image contrast. In this work, a novel approach for thyroid texture characterization based on extracting features utilizing higher order spectral analysis (HOSA) was used. A Support Vector Machine (SVM) was applied on the extracted features to classify the thyroid texture. Since HOSA is a well suited technique for processing non-Gaussian data involving non-linear dynamics, good classification of thyroid texture can be obtained in US images as they also contain non-Gaussian Speckle noise and non-linear characteristics. A final accuracy of 93.27%, sensitivity of 0.92 and specificity of 0.62 were obtained using the proposed approach.

D.1   Introduction

Thyroid is an important endocrine gland that is involved in several important body mechanisms like metabolic regulation, iodine absorption and protein synthesis. Thyroid diseases often involve changes in the shape and size of thyroid which make it essential to monitor the state as well as shape and volume of thyroid over time. We use US imaging instead of other medical image modalities as it is much safer and painless for the patients as well as easy to use with a higher availability for medical practitioners.

Several approaches have been proposed to classify and segment thyroid texture in US images [1], [2], [3]. These methods perform the segmentation by edge detection, thresholding between different gray values, region splitting and merging, active contours without edges, graph theory, segmentation based on normalized cut, localized region based active contour, distance regularized level set, fuzzy c-means algorithm, histogram clustering, QUAD tree, region and random walk.

Most of these methods are not automatic and are highly affected by the presence of speckle noise and thus require a pre-processing step. In general, it is very difficult to get rid of the speckle noise completely. In order to better deal with these noise, machine

learning based approaches have been proposed, but they require a large amount of data to train the classifiers.

HOSA method has been used for distinguishing between benign and malignant nodules in thyroid US images. These approaches involved a pre-processing step where the US images were enhanced using complex wavelet transform and finally HOSA was used to extract the features [4], [5]. However, up to our knowledge, all these approaches were used for classifying the thyroid nodules only but not the thyroid texture itself from the non-thyroid textures. Similarly, either frequency or entropy based features were used in these works to classify the thyroid nodules.

In this work, we propose a novel feature extraction technique that combines the energy, frequency and entropy based features to classify the thyroid textures. US images contain speckle noise which is known to have a non-Gaussian distribution and moreover the formation of these noise results from a non-linear process. This means that the speckle noise is dependent on the behavior of the sound propagation in the different body structure depending on the texture [6], [7]. Thus, no pre-processing has to be carried out using this approach because this method is a well suited technique for processing non-Gaussian data (i.e. the speckle noise) involving nonlinear dynamics. For that, the images are first divided into smaller texture patches and then features are extracted using HOSA, specifically bispectral analysis. The bispectral features subsequently were used for thyroid texture classification using a SVM. The results show that the features extracted from bispectrum of thyroid US image are significant to distinguish between the thyroid and non-thyroid textures in a thyroid US image. The SVM classification obtained an average test accuracy of 91.6% when a 1-fold holdout test approach was used for testing of the thyroid US images. A post-processing method was applied to improve the thyroid texture classification accuracy and produce a segmented thyroid with an accuracy of 93.27%.

145

D.2   Materials and Methods

D.2.1   Texture Dataset Generation

We have used a publicly available dataset [8] for our work. The dataset consists of 703 thyroid US images each with an image resolution of 760×500 pixels from 6 different patients with each patient containing between 53 to 218 2D thyroid images. The images were acquired along with the ground truth manually annotated by clinical experts. The images in the dataset were divided into smaller texture patches of size 20×20 pixels to prepare a texture dataset that consisted of 667,850 texture patches in total with each image producing 950 patches. Each texture patch was labelled as either thyroid ($= 1$) or non-thyroid ($= 0$). This labelling was done by comparing the texture patches to the ground truth obtained from expert clinicians. Fig. D.1 shows an example of separation of a thyroid US images into smaller texture patches.

D.2.2   Bispectral Analysis

Bispectrum is statistical analysis that falls in the category of higher order spectral which measures the spatial distribution of gray value and their deviation from its Gaussianity. One of the main advantages of bispectrum is that it allows the extraction of non-linear characteristics in an US texture patch. Additionally, it can assess the non-Gaussian property of speckle noise in US images [9].

In this work, the bispectrum is computed for each texture patch in the thyroid US images by using the Fourier Transform of a third-order cumulants sequence. In order to compute the bispectrum, first of all, the third-order moment $m_3^x(k_1, k_2)$, of $X[k]$, where $X[k]$ denotes a $N \times N$ texture patches (N=20) in an image and k is a 2-D index e.g.,

146
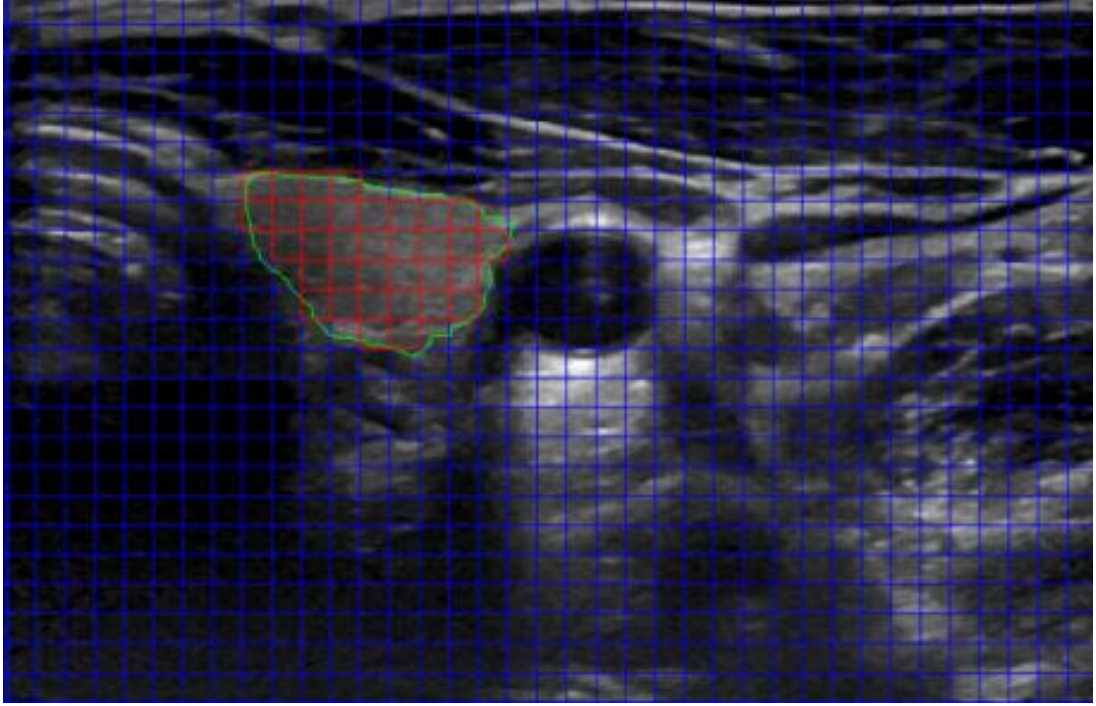
Fig. D.1: Dividing of a thyroid US image into texture patches (Green: Ground-Truth, Red: Thyroid Patches, Blue: Non-Thyroid Patches)

$k = [i_1, i_2]$ where $i_1, i_2 = 1, .., N$ and $X[k] \equiv X[i_1, i_2]$ is computed as following:

$$m_3^x(k_1, k_2) = EX(k)X(k + k_1)X(k + k_2) \tag{D.1}$$

where k_1 and k_2 are first and second-order correlation of $X[k]$ [?].

The third-order moment is then used to calculate the third-order cumulant of $X[k]$ using the following equation:

$$c_3^x(k_1, k_2) = m_3^x(k_1, k_2) - m_1^x[m_2^x(k_1) - m_2^x(k_2)$$
$$+ m_2^x(k_2 - k_1)] + 2(m_1^x)^3 \tag{D.2}$$

where, m_1^x corresponds to the first-order moment and m_2^x is the second-order moment.

Finally, all the three moments (i.e. first-order, second-order and third-order) are used to

compute the bispectrum that is the 2-D Fourier transform of the third-order cumulant function:

$$B(f_1, f_2) = \sum_{k_1} \sum_{k_2} c_3^x(k_1, k_2) \exp^{-j(k_1 f_1 + k_2 f_2)} \tag{D.3}$$

where f_1 and f_2 denote discrete index spatial frequencies.

The bispectrum of Eq. D.3 contains 12 symmetric regions which are repeated (redundant region), and knowing one bispectrum is sufficient to derive the others. To prevent this repetition, only $w \in \{f_1 = f_2\}$ is used for computing the bispectrum which is known as triangular region of bispectrum computation [11].

The bispectrum values were computed for each texture patches in thyroid US image. Then some features were extracted from these values for classifying the US image texture as thyroid and non-thyroid.

D.2.3 Feature Extraction

The bispectrum computed using Eq. D.3 is a complex matrix with large amounts of data for each texture patch. Since the computed bispectrum matrix contains large amounts of complex data, a feature extraction step is carried out to extract a set of 10 different features based on bispectral features already used in the literatures [12], [13]. A total of 10 linear and non-linear features were computed to classify thyroid and non-thyroid regions in a thyroid US image.

The linear features computed from the bispectrum are energy based and the purpose of using these features is to observe the dynamical bispectrum energy difference between thyroid and non-thyroid regions. For that we computed three different energy levels: the average, maximum and minimum computed from the bispectrum matrix of each

texture patch at different locations.

The non-linear features consist of a) frequency-relation-base, that shows the dynamical difference between different frequencies in the bispectrum of thyroid and non-thyroid texture patches. These features correspond to sum of logarithmic amplitude, sum of logarithmic amplitude of diagonal elements and first-order spectral moment of amplitude of diagonal bispectrum [14]; b) entropy-based features, that compute different level of entropy in order to show the difference between regularity and irregularity properties of different texture patches. These features are entropy of phase and entropy of phase domain calculated for first, second, and third power of domain [4].

D.2.4 Classification

In this work, a SVM was used as a supervised ML approach to classify the texture in thyroid US images using the 10 features extracted using bispectrum analysis.

SVM classifier is one of the supervised classifier that perform classification class by constructing hyperplanes. SVM has the capability of data classification by using linear and non-linear kernel function. In this work, linear boundaries were used to classify an image into two linearly separable classes of thyroid and non-thyroid. A Polykernel function was used to transform feature spaces into two classes. The polykernel is defined as:

$$k(x, y) = (X^T y + c)^d \tag{D.4}$$

Where x and y are feature vectors from two classes and d is the degree of the polynomial. Finally, c is a regularization parameter to control the trade-off between low error and

minimizing the norm of the weight. The classification was carried out in Matlab using a 1-fold holdout testing method where 5-fold data (i.e. from 5 patients) were used for training and 1-fold data were used for testing. This process was repeated 6 times until we included each patient data in the test set.

D.2.5 Post-Processing
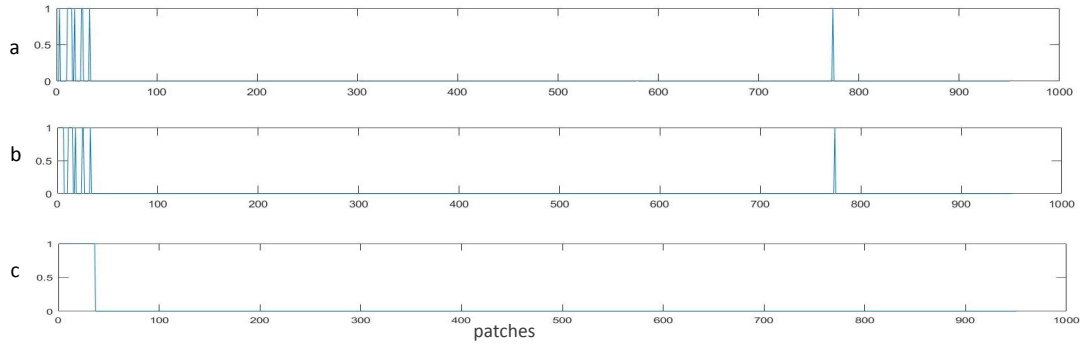
Fig. D.2: An example of post-processing method for an image in dataset D01. The x and y axis are number of patches and output labels ('1' or '0') for thyroid or non-thyroid respectively. a) shows the output labels of SVM classifier, b) is the output vector of the first post-processing step and c) is the output vector of the second post-processing step.

During classification, thyroid and non-thyroid texture patches are labeled as '1' and '0' respectively. The results of classification did not classify all the texture patches correctly. Hence, a post-processing was required after the classification step. In this regard, smoothing of the classification output vector was performed to reduce the misclassification of the SVM. This results in a better thyroid segmentation (see Fig. D.2).

Each thyroid US image consists of 950 texture patches, hence 950 output labels (i.e. either 1 or 0) are produced after classification by SVM (Fig. D.2.a) . These output labels are divided into blocks of size 6 (the block size was chosen based on the number of thyroid texture patches present in the smallest thyroid US images in the whole image dataset) resulting in a total of 158 blocks and 2 patches are left out. These two patches

150

are considered as non-thyroid and the rest of the 158 blocks are then analyzed in two steps, firstly by counting the number of '1' in each block. The whole block is labeled as 'thyroid' if there is at least one '1' label in it and 'non-thyroid' if there is no '1' label in it (Fig. D.2.b). In the second step, if any block before and after the current block is labeled as thyroid, then this current block is also labeled as 'thyroid', otherwise it is labeled as non-thyroid (Fig. D.2.c). Finally, the outer boundary of the classified thyroid texture patches were extracted to produce a segmented thyroid as shown in the fourth row of Fig. D.3.
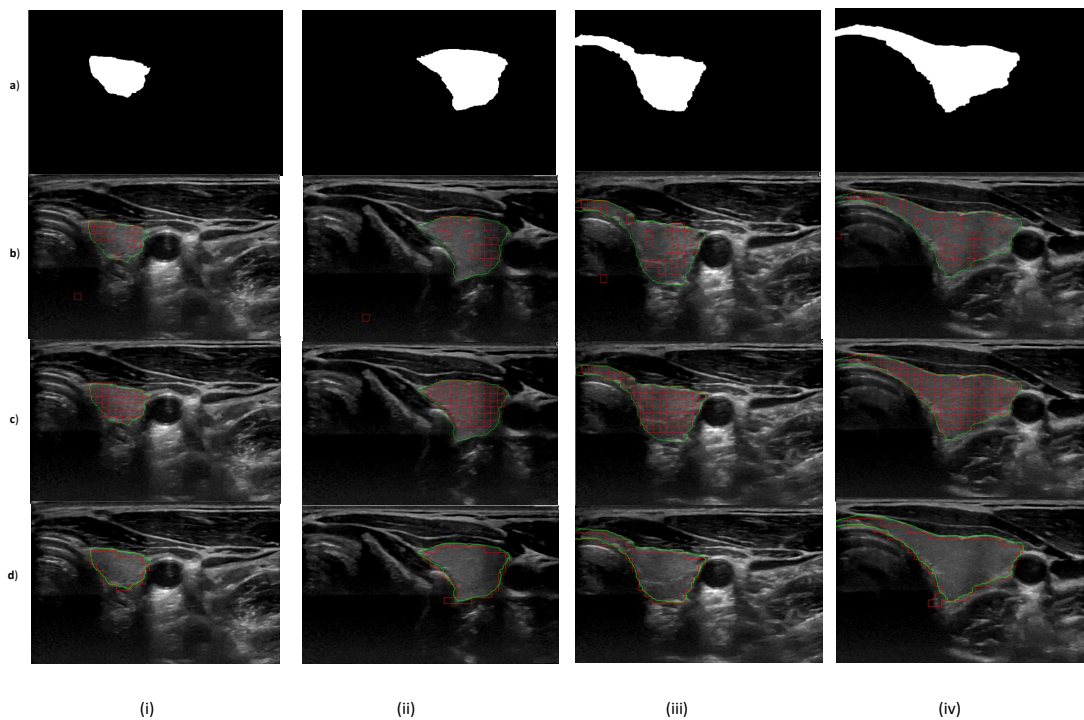
Fig. D.3: Thyroid Texture Classification and Segmentation Results in 4 images of different sizes from different patients (columnwise) where a) is the ground truth, b) is the SVM classified thyroid texture patches (in red and ground truth is green), c) is thyroid texture patches after post-processing (in red) and d) is the final segmented thyroid.

151

D.3 Results and Discussions

The output of the bispectrum analysis is shown in Fig. D.4. This figure shows four different bispectrums computed from different regions in the US image. The first two bispectrum images of a1 and a2 are from non-thyroid region while b1 and b2 are from thyroid region. The observation of the computed bispectrum shows that the dynamics of texture in these two different regions are completely different from each other. We can even see this difference in the two different non-thyroid regions (i.e. in a completely black and non-black non-thyroid region), whereas the bispectra look similar in the two different thyroid regions. This is the characteristic that we want to explore to classify the texture patches in an US image as thyroid and non-thyroid.
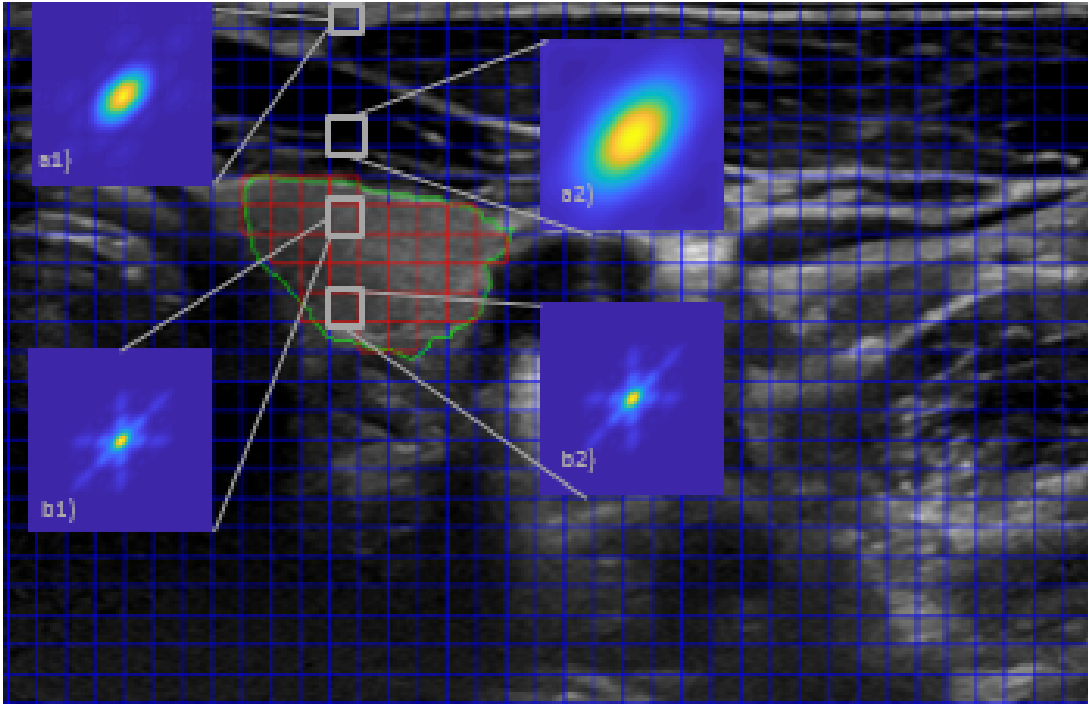
Fig. D.4: Bispectrum results of two different texture patches. a1) and a2) are the bispectrum of non-thyroid area and b1) and b2) are the bispectrum of the thyroid area.

All the 703 thyroid US images were analyzed using the proposed method. The performance of the proposed method was evaluated in the terms of accuracy, which is the

number of corrected identified patches over total number of patches. The accuracy of the texture classification using SVM was 91.6% and after post-processing was 93.27%. Similarly, the Sensitivity and Specificity were 0.83 and 0.53 before the post-processing and 0.92 and 0.62 after post-processing respectively. Fig. D.3 shows the classification results on four randomly selected thyroid images using SVM. These images were selected based on area the thyroid occupies in the US images (i.e. 2 left columns contain the smaller thyroid and 2 right columns contain the larger thyroid).

The results show that the proposed method can distinguish thyroid and non-thyroid texture in both the smaller and larger thyroid region (i.e. the second row in Fig. D.3). We also applied post-processing technique for improving the mis-classification of SVM and the results can be seen in the third row in Fig. D.3 and the final segmented thyroid can be seen in the fourth row in the same figure. The first row shows the ground truth binary images, which were manually annotated by the medical experts. The results show that the SVM classifier in conjunction with the post-processing step could classify the thyroid textures accurately. In the figure, the red patches represent the texture patches classified as thyroid while the green line represents the ground truth region.

The dataset was already used for thyroid segmentation by Poudel *et al.* [8]. They used classical segmentation methods such as active contour without edges (ACWE), graph cut (GC) and pixel based classifier (PBC) to segment the thyroid from the non-thyroid region. They obtained an accuracy of 73.1%, 74.8% and 67.2% using ACWE, GC and PBC respectively. Similar, none of the approaches used in this work could segment the thyroid region in the isthmus region and on top of that ACWE even used a pre-processing step to reduce the speckle noise in the US images. These results prove that the classical segmentation approaches are not well suited for segmentation of thyroid and non-thyroid region in the US images. The approach that we have proposed in this work is the first experimental study in thyroid image texture classification and segmentation

that estimates the features using HOSA. We have showed that this method can classify and segment the thyroid ultrasound images with high accuracy and do not require any pre-processing steps to reduce the speckle noise.

D.4 Conclusion

In this work, we proposed a novel feature extraction technique using bispectrum analysis to classify and finally segment the thyroid gland in an US image using a supervised learning approach. The extracted features along with a trained classifier and a post-processing showed a good performance on thyroid texture classification. The bispectral method can even work in the presence of speckle noise as no pre-processing step was used. Similarly, since the formation of different anatomical structures in an US image is a non-linear process, bispectral analysis can easily distinguish these different structures (for example thyroid and non-thyroid structures in our case) by computing robust features for texture classification. The results show that the 10 features extracted from bispectrum computation are able to accurately classify textures in thyroid US images. In future, more research can be carried out to classify different anatomical structure in US images as well as to apply this texture classification methods on images from different imaging modality.

References

1. J. Zhao, W. Zheng, L. Zhang, and H. Tian, "Segmentation of ultrasound images of thyroid nodule for assisting fine needle aspiration cytology," Health information science and systems, vol. 1, no. 1, pp. 5, 2013.

2. J. Kaur and A. Jindal, "Comparison of thyroidsegmentation algorithms in ultrasound

and scintigraphyimages," International Journal of Computer Applications, vol. 50, no. 23, 2012.

3. A. S. Agustin, S. S. Babu and K. T. Nadu, "Thyroid segmentation on us medicalimages: An overview," 2012.

4. Acharya, U. R., Sree, S. V., Swapna, G., Gupta, S., Molinari, F., Garberoglio, R., and Suri, J. S. (2013). Effect of complex wavelet transform filter on thyroid tumor classification in three-dimensional ultrasound. Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine, 227(3), 284-292.

5. Raghavendra, U., Gudigar, A., Maithri, M., Gertych, A., Meiburger, K. M., Yeong, C. H., ... and Acharya, U. R. (2018). Optimized multi-level elongated quinary patterns for the assessment of thyroid nodules in ultrasound images. Computers in biology and medicine, 95, 55-62.

6. T. E Hall and G. B. Giannakis, "Bispectral analysis and model validation of texture images," IEEE Transactions on Image Processing, vol. 4, no. 7,pp. 996-1009, 1995.

7. M. Nasrolahzadeh, Z. Mohammadpoory, and J. Haddadnia, "Higher-order spectral analysis of spontaneous speech signals in alzheimers disease," Cognitive Neurodynamics, pp. 1-14.

8. P. Poudel, A. Illanes, D. Sheet, and M. Friebe, "Evaluation of commonly used algorithms for thyroid ultrasound images segmentation and improvement using machine learning approaches," Journal of healthcare engineering, 2018.

9. A. Swami, G.B. Giannakis, and J.M. Mendel, "Linear modeling of multidimensional non-gaussian processes using cumulants," Multidimensional Systems and Signal Processing, vol. 1, no. 1, pp. 11-37, 1990.

10. M. K. Tsatsanis and G. B. Giannakis, "Object and texture classification using higher

155

order statistics," IEEE Transactions on Pattern Analysis and Machine Intelligence, , no. 7, pp. 733-750, 1992.

11. M. E. Hossain, W. A Jassim, and M.S.A. Zilany, "Reference free assessment of speechintelligibility using bispectrum of an auditory neurogram," PloS one, vol. 11, no. 3, pp. e0150415, 2016.

12. T. Ghosh, T. Biswas, and R. Khatun, "A feature extraction scheme to classify motor imagery movements based on bi-spectrum analysis of eeg," .

13. D. Karimi, "Spectral and bispectral analysis of awake breathing sounds for obstructive sleep apnea diagnosis," 2013.

14. K.C. Chua, V. Chandran, U.R. Acharya, and C.M. Lim, "Higher order spectra based support vector machine for arrhythmia classification," in 13th International Conference on Biomedical Engineering. Springer, 2009, pp.231-234.

15. R. Sharma, R.B. Pachori, and U.R. Acharya, "Application of entropy measures on intrinsic mode functions for the automated identification of focal electroencephalogram signals," Entropy, vol. 17,no. 2, pp. 669-691, 2015.

CHAPTER E

Thyroid Ultrasound Texture Classification using Autoregressive Features in Conjunction with Machine Learning Approaches

Thyroid Ultrasound Texture Classification Using Autoregressive Features in Conjunction With Machine Learning Approaches

PRABAL POUDEL[iD], **(Member, IEEE), ALFREDO ILLANES**[iD], **ELMER J. G. ATAIDE, (Member, IEEE), NAZILA ESMAEILI, (Member, IEEE), SATHISH BALAKRISHNAN, AND MICHAEL FRIEBE**[iD], **(Senior Member, IEEE)**
Faculty of Medical Engineering, Otto-von-Guericke University, 39106 Magdeburg, Germany

Corresponding author: Prabal Poudel (prabal.poudel@ovgu.de)

ABSTRACT The thyroid is one of the largest endocrine glands in the human body, which is involved in several body mechanisms like controlling protein synthesis, use of energy sources, and controlling the body's sensitivity to other hormones. Thyroid segmentation and volume reconstruction are hence essential to diagnose thyroid related diseases as most of these diseases involve a change in the shape and size of the thyroid over time. Classification of thyroid texture is the first step toward the segmentation of the thyroid. The classification of texture in thyroid Ultrasound (US) images is not an easy task as it suffers from low image contrast, presence of speckle noise, and non-homogeneous texture distribution inside the thyroid region. Hence, a robust algorithmic approach is required to accurately classify thyroid texture. In this paper, we propose three machine learning based approaches: Support Vector Machine; Artificial Neural Network; and Random Forest Classifier to classify thyroid texture. The computation of features for training these classifiers is based on a novel approach recently proposed by our team, where autoregressive modeling was applied on a signal version of the 2D thyroid US images to compute 30 spectral energy-based features for classifying the thyroid and non-thyroid textures. Our approach differs from the methods proposed in the literature as they use image-based features to characterize thyroid tissues. We obtained an accuracy of around 90% with all the three methods.

INDEX TERMS Medical imaging, support vector machine, artificial neural network, random forest classifier, texture classification, thyroid ultrasound.

157

ABSTRACT

The thyroid is one of the largest endocrine glands in the human body, which is involved in several body mechanisms like controlling protein synthesis, use of energy sources, and controlling the body's sensitivity to other hormones. Thyroid segmentation and volume reconstruction are hence essential to diagnose thyroid related diseases as most of these diseases involve a change in the shape and size of the thyroid over time. Classification of thyroid texture is the first step towards the segmentation of the thyroid. The classification of texture in thyroid Ultrasound (US) images is not an easy task as it suffers from low image contrast, presence of speckle noise and non-homogeneous texture distribution inside the thyroid region. Hence, a robust algorithmic approach is required to accurately classify thyroid texture. In this work, we propose three machine learning based approaches: Support Vector Machine, Artificial Neural Network and Random Forest Classifier to classify thyroid texture. The computation of features for training these classifiers is based on a novel approach recently proposed by our team, where autoregressive modelling was applied on a signal version of the 2D thyroid US images to compute 30 spectral energy based features for classifying the thyroid and non-thyroid textures. Our approach differs from the methods proposed in the literature as they use image-based features to characterize thyroid tissues. We obtained an accuracy of around 90% with all the three methods.

E.1   Introduction

The thyroid is a butterfly shaped gland, one of the largest endocrine glands in the body, located below Adam's apple on the front of the neck. It is involved in several body mechanisms such as controlling protein synthesis, use of energy sources and controlling the body's sensitivity to other hormones. Due to these important functionalities, the

158

thyroid is one of the important organs in the human body. However, it is susceptible to many diseases like Graves' (excessive production of thyroid hormones), subacute thyroiditis (inflammation of thyroid), thyroid cancer, goiter (thyroid swelling), etc [1]. In all of these cases, the size of the thyroid changes over time. So, it is essential to keep track of the thyroid size over time.

Ultrasound (US) imaging has been widely used for thyroid staging, as it is much safer and painless to use for the patients compared to other imaging modalities such as MRI which uses radio and magnetic waves, Computed Tomography (CT) which uses X-rays and Positron Emission Tomography (PET) which uses nuclear imaging technique [2]. Segmentation and volume computation of the thyroid have high clinical importance when it comes to the diagnosis and treatment of thyroid diseases. In this work, we will mainly focus on characterization of thyroid texture in an US image using three machine learning (ML) techniques. These approaches are Support Vector Machine (SVM), Artificial Neural Network (ANN) and Random Forest Classifier (RFC).

The features computed in this work for training the classifier are based on a novel texture characterization algorithm published previously by our team [3]. A signal based parametrical approach using Autoregressive (AR) modelling has been proposed to characterize the thyroid texture using 30 AR spectral energy ratios based features that can distinguish between thyroid and non-thyroid regions. A simple clustering algorithm has been used to show the significance of the proposed AR-based features. In this new proposed work, we go further and use our robust textural features to train three different machine learning based approaches (SVM, ANN and RFC) that have already been used to segment US images in the literature. We show in this work that using the AR features together with the proposed classifiers the obtained results outperform other thyroid segmentation algorithm already presented in the literature.

The rest of the paper is organized as following: Section II presents the reviews on the

159

related works on thyroid segmentation. Section III discusses about the novel feature extraction that we have used to extract signal based features from thyroid US images and the different texture classification methods. Section IV presents the results and compares our results with the ones from literature. Finally, Section V presents the discussion on the future works that we have planned as well as the conclusions that can be drawn from our work.

E.2   Related Works

Several approaches have been proposed on how to segment the thyroid in 2D US images. Zhao *et al.* [4] proposed several thyroid US segmentation approaches using edge detection, thresholding, region splitting and merging, watershed segmentation, active contour, graph theory, US image segmentation based on Ncut and segmentation based on improved normalized cut. Thyroid segmentation in 2D US and scintigraphy images using active contour without edges (ACWE), localized region based active contour and distance regularized level set were proposed by J. Kaur and A.jindal [5]. China *et al.* [6] explored the possibilities of using the apriori information based on the US imaging physics and segmented the thyroid using Iterative Random Walks and Random Forest (IRWRF). Similarly, segmentation using a polynomial SVM [7], local region-based active contour [8], a boundary method and local binary patterns [9] for texture analysis and level-set active contours models [10] and [11] have been proposed. H. Garg and A. Jindal worked on feed-forward neural network (FNN) to segment the thyroid in US images [12]. Similarly, Echogenicity based Quantization (EBQ) and Joint Classification-Regression (JCR) which uses speckle related pixels and imaging artefacts as a source of information to perform multi-organ (i.e. thyroid, carotid artery, muscles and trachea) segmentation in thyroid US images were proposed by Narayan *et al.* [13].

160

Apart from segmentation in 2D images, several research works have been carried out to segment a fully 3D thyroid. A semi-automated approach to classify thyroid for volumetric quantification using geodesic active contour was proposed by Kollorz *et al.* [14]. Chang *et al.* used a radial basis function (RBF) neural network to segment the blocks of thyroid gland [15]. Similarly, a complete segmentation and analysis of 3D thyroid images was performed by A. Osman [16] by thresholding the voxel intensities and then connecting similar voxels to predict the thyroid regions. Poudel *et al.* [17] have used Active Contours without Edges (ACWE), Graph Cut (GC) and Pixel Based Classifier (PBC) to segment 2D thyroid images and later reconstructed them to compute a 3D thyroid.

Most of the above mentioned approaches involved thyroid segmentation using data-driven approaches which means that, the segmentation of thyroid was carrying out by directly operating over the pixel values in the US images. Similarly, several works have been proposed for thyroid nodule classification by characterizing the thyroid tissues. These works are based on computation of Statistical features [18-20], Spectral-based features [21], [22] and Higher Order statistics based features [23], [24]. The problem with using these data-driven approaches for feature computation is that, they are generally affected by the presence of speckle noise, low signal to noise ratio (SNR) and resolution in US images and even the pre-processing steps cannot get rid of these problems completely.

Similarly, most of the methods in the literature do not explore texture based features for thyroid segmentation. We believe this is due to the heterogeneous textural patterns within the thyroid US images [25] and thus a novel texture based feature extraction method should be devised to extract robust features which could be used to train the machine learning classifiers for thyroid segmentation.

As explained earlier, different machine learning based classifiers have been trained only using statistical, spectral and higher order statistical based features in the literature

for thyroid texture classification. However, we have used a set of novel parametrical based features computed using AR modelling to classify the thyroid textures. To our knowledge, these features have not been used for training the machine learning based classifiers for thyroid texture classification. This is the main contribution of our work. We have used three widely used methods of texture classification from the literature and outperformed several other state-of-the art approaches which use different features compared to ours.

E.3   Methods

This section is divided into four sections: database generation, features computation, texture classification and post-processing. In the first section, we will mainly discuss how the 2D US image datasets were acquired and how the texture patch database for training of the classifiers was prepared. The second section will present how the features were computed from the texture patches which were used for training of the classifiers and the third section presents the thyroid texture classification approach using SVM, ANN and RFC. Finally, the fourth section will explain a simple post-processing step that we have used to get rid of the over classified thyroid texture patches from the three trained classifiers.

E.3.1   Thyroid Datasets and Texture Database Generation

A total of two 2D thyroid US datasets were used in this work. The first dataset (Dataset 1) consisted of six subjects with each subject containing between 53 and 189 2D thyroid US images. A total of 675 thyroid images with an image size of 760 x 500 pixels were used. This dataset was acquired by a medical expert in SurgicEye GmbH [36] and has

been published and available in [26] The second dataset (Dataset 2) consisted of sixteen subjects with each subject containing between 156 and 289 2D thyroid US images. The second dataset was obtained by a thyroid specialist medical doctor at University Clinic of Magdeburg, Germany and contains a total of $3,370$ thyroid US images with an image size of 760 x 1020 pixels. It has been presented in [27] and can be downloaded from http://opencas.webarchiv.kit.edu/?q=node/29 Along with the US images, we also acquired manually annotated ground truth images from the respective clinical experts who acquired the thyroid images. All the images were acquired using a General Electric (GE) Logiq E9 US machine equipped with Electromagnetic Tracking system. The acquired tracking data could be used for 3D reconstruction of segmented thyroid images and volume assessment over time.

The two datasets were further processed to compute the features for training of the ML classifiers. Each image from the two datasets were first divided into non-overlapping texture patches of size 20 x 20 and following the ground truth, each patch was labelled either thyroid (=0) or non-thyroid (=1). The size of the texture patch was set in such a way that it captured important dynamical changes that allowed to involve a number of main frequency components that can help to spectrally differentiate thyroid and non-thyroid regions (see [3]). On top of that, it should also cover all the smaller regions inside the thyroid (for example the isthmus region as marked by yellow solid lines in Fig. E.1).

For the labelling, a thresholding technique was used. Each pixel inside the patch was compared against the ground truth pixels. Hence, if a patch consisted of more than 70% (i.e. 280 pixels) of the total pixels, the patch was labelled as thyroid. Similarly, any patch that consisted of only black pixels (i.e. sum of all the pixel intensities inside a patch = 0) were not used as these patches could not be used to compute the features. A total of $90,816$ and $1,791,397$ texture patches were computed from Dataset 1 and

2 respectively to prepare a final texture database. An example of separation of a 2D thyroid US image into texture patches is shown in Fig. E.1. In the figure, the green patches represent the thyroid and the blue patches represent the non-thyroid patches. The thyroid patches are always present inside the thyroid region which is marked as red using the ground truth images.
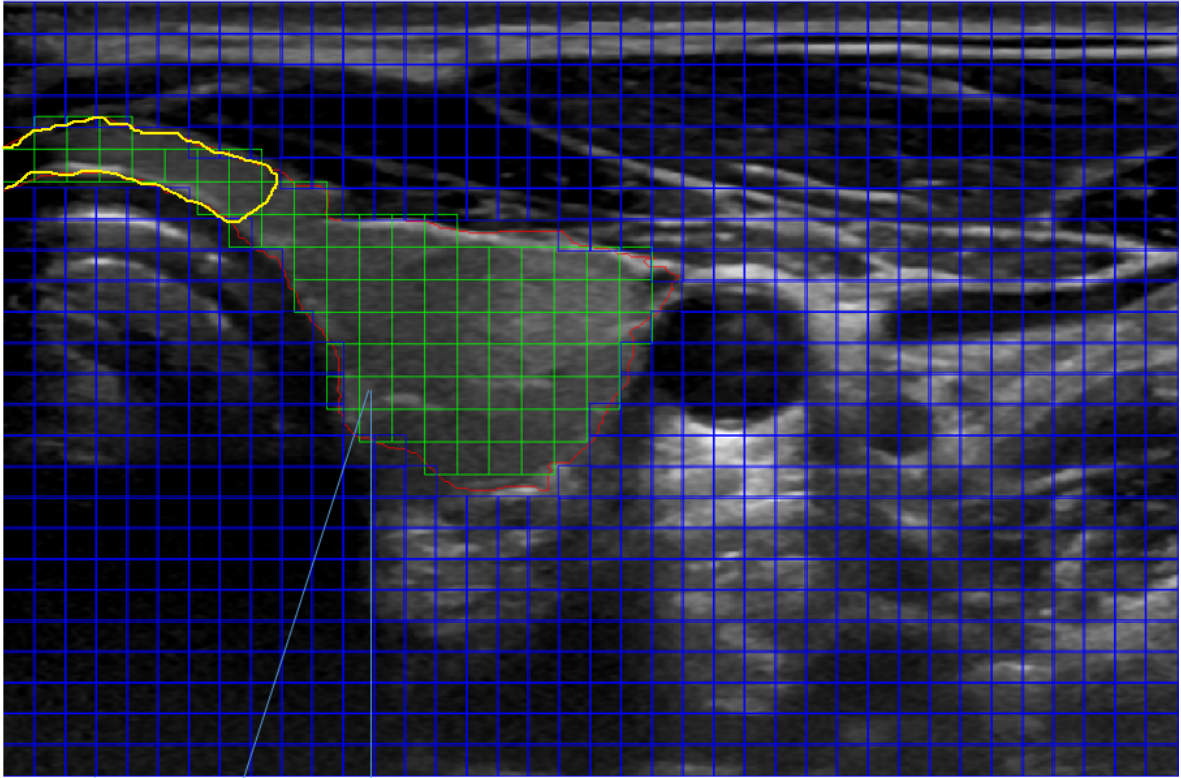
Fig. E.1: The figure represents the division of a 2D thyroid US images into smaller texture patches. In the figure, Green: Thyroid Patches, Blue: Non-Thyroid Patches, Red:Ground Truth and Yellow: Isthmus Region.

E.3.2 Features Computation

In this section, we will mainly discuss on how the features were computed from the thyroid images which were used for the training of the classifiers for thyroid texture classification. A detailed explanation on AR modelling, feature computation and prominent

features selection have been explained in our recent work [3] but we will only introduce the main steps here. We used AR modelling to compute the features from the texture patches. The advantage of AR modelling is that the features are computed not directly from the image data (which in general contain speckle noise and have low SNR and contrast) like in Fast Fourier Transform based techniques, but using a parametrical version of the image data. This allows computing robust features in noisy images and less data compared to the standard data-driven methods.

First of all, the texture patches are converted into four different types of signals which capture the texture dynamics within the patch. The transformation from matrix to signal has been performed using ZigZag (obtained by following the rows direction) and Spiral transformation and also using their 90 degree rotated patch version (see Fig. E.2).
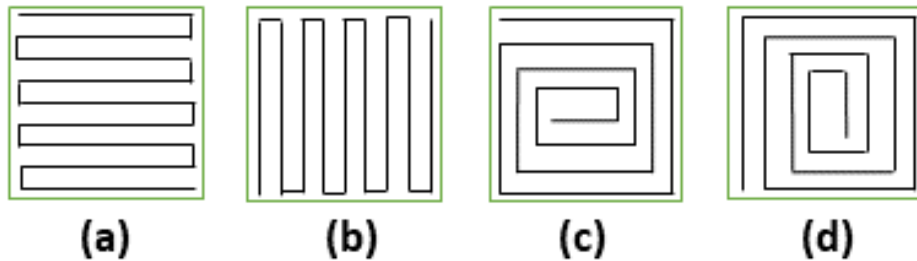
Fig. E.2: Conversion of texture patch to four different signals. ((a)ZigZag, (b)ZigZag 90 degree rotated, (c) Spiral and (d) Spiral 90 degree rotated respectively. Adopted from [3]

These signals were then decomposed into four narrowband signals (i.e. low, middle, high and total band frequency components - LF, MF, GF and TB respectively) by applying Continuous Wavelet Transformation (CWT). These signals represent the dynamic textural characteristics such as smoothness or roughness in the texture patches. A total of 16 narrowband signals were obtained which were modeled using a parametrical AR model [28]. A set of 30 different features were computed from the AR parameters using the energy ratios between different frequency bands for each texture patch in the texture database. A detailed explanation on AR modelling, features computation and

prominent features selection have been explained in [3]. A flowchart representing the entire feature computation process is shown in Fig. E.3.

Fig. E.3: Flowchart representing the entire feature extraction process.

E.3.3   Texture Classification

This section will present all the three algorithms that were used to classify the thyroid texture in US images.

Support Vector Machine (SVM)

In this work, SVM with radial basis function (rbf) (aka gaussian) kernel was used to classify the thyroid texture patches in US images. The features that were obtained from the feature extraction procedure were used to train the SVM. A total of 30 features were used to train the SVM classifier. The trained classifier was later used to test the input images by classifying the texture patches as thyroid or non-thyroid.

Let $x \in R^n$ be a vector of all the features extracted from the texture patches to be classified and let a scalar y denote its class label (i.e. whether the texture patch belongs to thyroid or not, $y \in \{0, 1\}$). Also, let $\{(x_i, y_i), i = 1, 2, 3, ...l\}$ be a set of l training data. For the simplest case, when the training patterns are linearly separable, there exist a linear function:

$$f(x) = W^T x + C \tag{E.1}$$

which separates the two different classes by a hyperplane:

$$f(x) = W^T x + C = 0 \tag{E.2}$$

where, C is the regularization parameter which controls the cost of misclassification on the training data.

However, there might exist many hyperplanes that maximize the separating margin between the two classes. The hyperplane that causes the largest separation between the different classes is computed by the SVM using minimizing the cost function [29]:

$$f(W) = \frac{1}{2} W^T W = \frac{1}{2} ||W^T||^2 \tag{E.3}$$

However, when the data are not linearly separable, a hyperplane cannot separate the data correctly. Thus, kernel functions are analysed to achieve this separation. In this work, a radial basis function (rbf) kernel is used which is given by:

$$k(x, z) = exp(-\frac{||x - z||^2}{2\sigma^2}) \tag{E.4}$$

where,$(\gamma = \frac{1}{2\sigma^2})$ is the kernel parameter that defines how far the influence of a single training example reaches. In other words, if the value of γ is low, then the far away points from the hyperplane carry more weights and if it is low, the nearer points carry more weights.

Using this kernel, all the features are sent as the input to the SVM classifier to train it. The features are represented in the vector form as:

$$x_i = [f_{i,1}, f_{i,2}, ..., f_{i,n}] \tag{E.5}$$

where $f_{i,n}$ is the nth feature of the ith texture patch.

These features are used as the training vector to train the SVM which is later used for testing the input images. The three parameters that could be optimized while using SVM are the kernel, C and the gamma γ. A grid search method with a 10-fold cross validation technique on the training data was employed to find the optimum parameters. We found that the SVM performed the best with 'rbf' kernel, $C = 0.7$ and $\gamma = 1.0$. A total of 75% training and 25% testing data were used to train and test the SVM classifier. The training and testing of SVM were carried out in Matlab 2017a using the Image Processing Toolbox.

Artificial Neural Network (ANN)

The classification of thyroid texture patches in US images was also done using ANN that is primarily an interconnected web of input nodes, hidden nodes and output nodes called artificial neurons.

The first step was to pre-process the data. The dataset was first split into dependent and independent variables. The independent variables consisted of the 30 features that were computed in the feature extraction section. The dependent variables consisted of data (represented as 1 or 0) that indicated whether the features belonged to thyroid or non-thyroid patches. Following this, the dataset was split into the training and testing sets by employing the train test split (75% and 25%) from scikit-learn model selection. Feature scaling was employed on the training and test sets to ensure that all the values were in the same range.

The ANN (Multi-layer Perceptron) was built with the Keras library using TensorFlow on the backend based on the Stochastic Gradient Descent Algorithm (SGD). SGD was used as an iterative method to adjust the weights and obtain a minimum cost function

and hence an optimal neural network. The SGD is represented by the following equation:

$$Q(w) = \frac{1}{n} \sum_{i=1}^{n} Q_i(w)_i \qquad \text{(E.6)}$$

where, $Q_i(w)$ is a loss function based on the training data indexed by i [30].

The optimization of the parameters of ANN involved some empirical analysis. After few experiments, we found that the ANN outperformed SVM and RFC. Hence, the following parameters were chosen as the optimum ones: 100 epochs, learning rate of 0.1, momentum of 0.9 and 4 layers. The 4 layers consisted of an input layer, two hidden layers (each of them containing 15 nodes) and an output layer. The Sequential module was used for the initialization of the network as a sequence of layers and the Dense module was used to build the layers.A Rectified Linear Unit (ReLU) activation function was used for the activation of the hidden layers. A sigmoid activation function was used for the output layers and is represented by:

$$f(x) = \frac{1}{1 + e^{-x}} \qquad \text{(E.7)}$$

where, x = value of the weighted sums and e = Euler's number (= 2.71828) [31].

The classifier was saved after training the network with a batch size of 32 and 100 epochs. The trained ANN was used for testing the input images. The classification using ANN was carried out using Python 3.6 with the help of libraries such as Scikit-learn, Keras and TensorFlow.

Random Forest Classifier (RFC)

In our approach, we trained a random forest classifier for a binary classification problem, which classifies each of the patches extracted from the US images as thyroid and non-

thyroid. RFC is basically a type of ensemble learning method which usually constructs a final classifier by using a set of M individual weak classifiers. In this case, these weak classifiers are the binary decision trees. A train-test split of 75% to 25% was used.

The input from the training data for each of the trees, $x \in \{1, ..., M\}$ in the ensemble was created using bootstrapping of the samples (bagging) from the training dataset and randomly sampling the subset of the features supplied to each tree. Introducing this level of randomness helped this classifier in reducing to an extent, the dependency between training and testing data. Each tree is a collection of nodes N and features F, which aid to the final classification result. A decision tree is made up of a single parent node $N_{p,x}$ and multiple splitting nodes $N_{s,x,i} \forall i \in \{1, ..., k\}$ and leaf nodes $N_{l,x,j} \forall j \in 1, ..., p$. During the splitting of the nodes, the best split was not chosen based on all the features but a random subset of features from the training dataset.

All the leaf nodes inside a decision tree have a final probabilistic model $\phi_{x,j} \in [0, 1]$ associated with it. The final decision of a forest for each of the patches extracted from the US images were made by averaging the individual decisions $(\phi_{x,j}(p))$ from all the individual trees in the forest.

$$P^{RF}(y(p) = 1) = \frac{1}{M} \sum_{x=1}^{M} \phi_x(p) \tag{E.8}$$

We have used the most common and recognized method to train the classifiers [30], [32]. Just like ANN, the classification using RFC was carried out in Python 3.6 using Scikit-learn, Keras and TensorFlow libraries.

There are many parameters that can be optimized in RFC. However, we optimized only the 5 important parameters which were the depth of the trees, minimum number of samples required to split a node, minimum number of samples required at each leaf node, number of trees in the random forest and whether to use bootstrap or not. The

optimum parameters that were obtained after using Randomized Search method were depth of 10, minimum samples at each leaf node of 2, minimum samples to split a node 4, 200 trees and using the bootstrap method for sampling the training data points.

E.3.4   Post-Processing

The texture classification step produced some over-classified thyroid texture patches. Hence, to get rid of these over-classifications, a post-processing step was employed. A largest connected component analysis was performed on the classified texture patches. For that, the total number of texture patches were obtained by counting the patches that were classified as thyroid (i.e. the output label = 1). Then a threshold value was chosen empirically to identify the thyroid patches from the over-classified thyroid patches. The blocks of texture patches that contained more than the threshold amount of thyroid patches were considered to be thyroid and the rest were disregarded. Section IV C presents the results from before and after post-processing steps in details.

E.4   Results

E.4.1   Experimental Setup

For the evaluation and quantitative and qualitative analysis of the proposed feature extraction and texture classification technique, we performed two-steps experiments. The two datasets were trained and tested separately. A total of $90,816$ and $1,791,397$ texture patches corresponding to Dataset 1 and 2 respectively were used for this evaluation. Out of these patches, only $68,112$ patches were used for training and $22,704$ patches were used for testing in Dataset 1. Similarly, $1,343,548$ patches were used for training and $447,849$ patches were used for testing in Dataset 2. In both datasets, to

ensure there was no over-fitting while training of the classifiers, it was made sure that the training and testing processes did not involve images or texture patches from the same subjects. The training and testing processes involved the 75% and 25% of all the texture patches respectively.

The feature extraction part was performed using MATLAB 2017a and the training and testing of the classifiers was performed in Python 3.6. All the experiments were carried out using a Lenovo T430 ThinkPad Notebook with Intel Core i5-3320M CPU, NVIDIA NVS 5400 graphics card, 2.60 GHz processor and 8.00 GB RAM.

E.4.2 Quantitative Analysis

For the quantitative analysis, we have compared our results with the approaches in state of art that used the same datasets. Similarly, we have also compared our approaches with other approaches but which do not use the same datasets. For the performance metric, we have used Dice's Coefficient (DC), Sensitivity (SE) and Specificity (SP).

DSC is a measure of how similar two objects are, which in our case is the computation of the overlap area between the ground truth images and classified thyroid texture patches. Similarly, **SE** is the measure of the proportion of actual positives that are correctly identified as such. **SP** is the measure of the proportion of actual negatives that are correctly identified as such. They can be computed using the following equations:

$$DSC = \frac{2TP}{2TP + FP + FN} \tag{E.9}$$

$$SE = \frac{TP}{TP + FN} \tag{E.10}$$

$$SP = \frac{TN}{TN + FP} \qquad \text{(E.11)}$$

where, TP = True Positive (Thyroid Patches identified as Thyroid), FP = False Positive (Non-Thyroid Patches identified as Thyroid), TN = True Negative (Non-Thyroid Patches identified as Non-Thyroid) and FN = False Negative (Thyroid Patches identified as Non-Thyroid).

Table E.1: Performance Comparison of SVM, ANN and RFC with State of Art Methods on Dataset 1.

Methods	DSC	SE	SP
ACWE [17]	0.805	-	-
GC [17]	0.745	-	-
PBC [17]	0.666	-	-
RFC - Volume Based [17]	0.855	-	-
CNN - Volume Based [17]	0.872	-	-
KMEANS [3]	0.897	0.950	0.700
SVM	**0.895**	**0.896**	**0.818**
ANN	**0.930**	**0.928**	**0.970**
RFC	**0.925**	**0.925**	**0.866**

Table E.2: Performance Comparison of SVM, ANN and RFC with State of Art Methods on Dataset 2.

Methods	DSC	SE	SP
IRWRF [6]	0.854	0.989	0.923
KMEANS [3]	0.869	0.890	0.620
SVM	**0.887**	**0.887**	**0.556**
ANN	**0.894**	**0.935**	**0.535**
RFC	**0.891**	**0.935**	**0.517**

Using these performance metrics, we have presented the results of SVM, ANN and RFC and compared them with state of arts in the tables below. These comparisons are

Table E.3: Performance Analysis of Different State of Arts for Thyroid Segmentation using Different Datasets

Methods	DSC	SE	SP
EBQ [13]	0.839	0.955	0.889
JCR [13]	0.479	0.564	0.926
RBF [15]	0.512	0.874	0.560
FNN [12]	0.400	0.473	0.864

carried out in a 2-step procedure. The first step involved the comparison between all the approaches that use either Dataset 1 or 2 and in the second step, all the approaches were used for thyroid segmentation but using different datasets. The comparison of performance between SVM, ANN and RFC and state of arts are presented in Table E.1 and E.2 and Table E.3 shows the comparison between different approaches that use different datasets. Table E.4 summarizes all the parameters we used after the optimization process in SVM, RFC and ANN classifiers for texture classification.

Table E.1 represents the comparison between the approaches we have used in our work with the works in [17] and [3] using Dataset 1. Active Contours without Edges (ACWE), Graph Cut (GC), Pixel based classifier (PBC), Random Forest Classifier (RFC) and Convolutional Neural Network (CNN) were used in [17] for thyroid segmentation. Out of these 5 approaches, the first three were non-machine learning (NML) based methods and the last two methods used machine learning (ML). However, these last two approaches were operated directly on 3D thyroid images. Similarly, kmeans (a simple clustering algorithm) was used in [3] to cluster and segment thyroid region in 2D thyroid US images.

Similarly, Table E.2 presents the comparison between our three approaches and Iterative Random Walks and Random Forest (IRWRF) from [6], a ML based and kmeans from [3], a NML based approaches using Dataset 2. We also present the results of

Table E.4: Summary of all the optimized parameters used in SVM, RFC and ANN.

SVM	RFC	ANN
• kernel = 'rbf' • $C = 0.7$ • $\gamma = 1.0$	• tree depth = 10 • minimum samples to split leaf node = 2 • minimum samples to split a node = 4 • number of trees = 200 • bootstrap = 'true'	• epochs = 100 • learning rate = 0.1 • momentum = 0.9 • number of layers = 4

thyroid segmentation using four other algorithms in Table E.3. It presents the results using Echogenicity-based Quantization [13], Joint Classification-Regression (JCR) [13], RBF Neural Network (RBF) [15] and Feedforward Neural Network (FNN) [12] in terms of DSC, SE and SP using different thyroid US datasets. Despite the fact that these approaches use different datasets than we use, we present these results just to see how these algorithms perform in the domain of texture classification in thyroid US images.

All these metrics were computed using confusion matrix (CM) for each of the approaches used in our work. We present the CM for all the three algorithms when used on both the datasets below. In terms of TP, FN, FP and TN, the CM can be represented as below. The CM were computed during the tests we carried out in the test sets which consisted of $22,704$ and $447,849$ texture patches in Dataset 1 and 2 respectively.

$$CM = \begin{bmatrix} TP & FN \\ FP & TN \end{bmatrix} \qquad (E.12)$$

Dataset 1:

$$SVM = \begin{bmatrix} 20317 & 2368 \\ 16 & 3 \end{bmatrix}$$

$$ANN = \begin{bmatrix} 20311 & 1576 \\ 22 & 795 \end{bmatrix}$$

$$RFC = \begin{bmatrix} 20290 & 1654 \\ 43 & 717 \end{bmatrix}$$

Dataset 2:

$$SVM = \begin{bmatrix} 397213 & 50600 \\ 34 & 2 \end{bmatrix}$$

$$ANN = \begin{bmatrix} 375914 & 26060 \\ 21333 & 24542 \end{bmatrix}$$

$$RFC = \begin{bmatrix} 374365 & 26133 \\ 22882 & 24469 \end{bmatrix}$$

From Table E.1 and E.2, we can see that all the three classifiers can classify the thyroid texture patches with better if not comparable accuracies. SVM has the lowest accuracy out of the three classifiers with a DC of 0.895 and 0.887 in Datasets 1 and 2

respectively. Similarly, ANN has the highest accuracy out of the three classifiers with a DC of 0.930 and 0.894 in Datasets 1 and 2 respectively. RFC produces almost the same accuracy as ANN with a DC of 0.925 and 0.891 in Datasets 1 and 2 respectively. These results can be visually accessed in the section below (see Section IV C). Similarly, all the three approaches outperformed ACWE, GC, PBC, RFC - Volume Based, CNN - Volume Based and KMEANS on Dataset 1 and KMEANS and IRWRF on Dataset 2 (except for KMEANS outperforming SVM on dataset 1). Apart from other methods, the tests with RFC and CNN - volume based were tested on the 3D thyroid volumes corresponding to Dataset 1 instead of individual 2D images.

We have also presented the performance matrices in terms of DSC, SE and SP from four different approaches in the literature such as EBQ, JCR, RBF and FNN despite the fact that they were tested on different datasets compared to what we are using in this work. These results are displayed in Table E.3. Compared to these approaches too, SVM, ANN and RFC achieve better DSC and similar SE and SP in both the datasets. These results prove the robustness of the feature extraction process for thyroid texture classification.

Apart from the accuracy of classification, the feature extraction and training and testing of the approaches are fully automatic compared to ACWE, GC and PBC which use some level of human interaction. ACWE requires the user to draw an initial contour, GC requires the user to scribble the thyroid and non-thyroid region as a initialization process and PBC requires the users to click inside and outside of the thyroid regions to extract features from these regions. Also, the initializations are very important in these approaches as a wrong initialization could result in a misclassification of the different regions.

The computation time for feature extraction in our work is higher compared to the state of art techniques. This is mainly because we compute the wavelet spectrum for

all the scales (or frequencies) in the LF, MF, HF and TB bands. An optimization step can be carried out to compute the spectrum at a scale that best represents these bands. Similarly, during AR modelling, instead of computing the power spectral densities (PSD) at all the frequency components in the complex plane, a set of non-repetitive frequency components could be chosen. On top of that, we have computed all the features using MATLAB which makes the process a lot slower. The optimization processes and the computation of these features in C++ could increase the frequency computation speed by a factor of 100. However, it is worth to mention that these features need to be computed only once and can be stored in a .csv file for training the networks in future. The time taken for classifying a new thyroid US image is however faster compared to the state of art methods. This makes it applicable for clinical use as the doctors and radiologist can just take a set of individual US images and segment the thyroid regions using the trained classifiers.
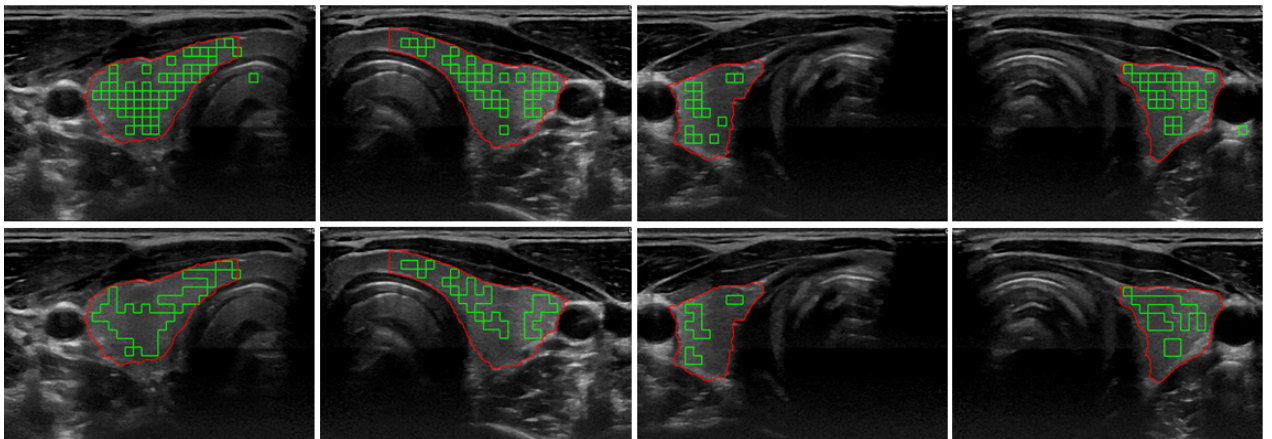

Fig. E.4: Examples of thyroid texture classification and segmentation using SVM and comparison with ground truth on Dataset 1.
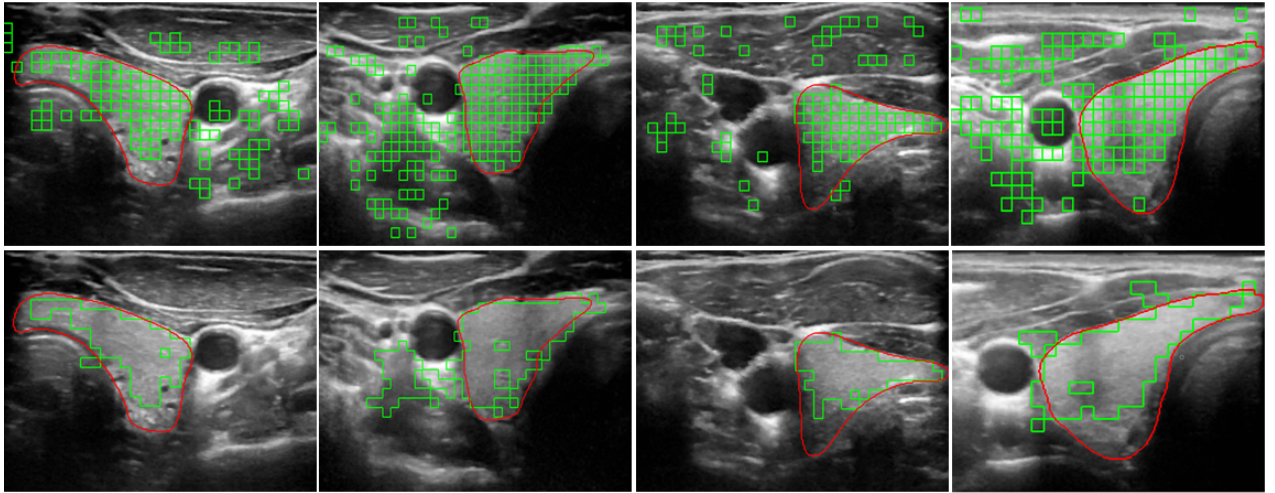
Fig. E.5: Examples of thyroid texture classification and segmentation using SVM and comparison with ground truth on Dataset 2.
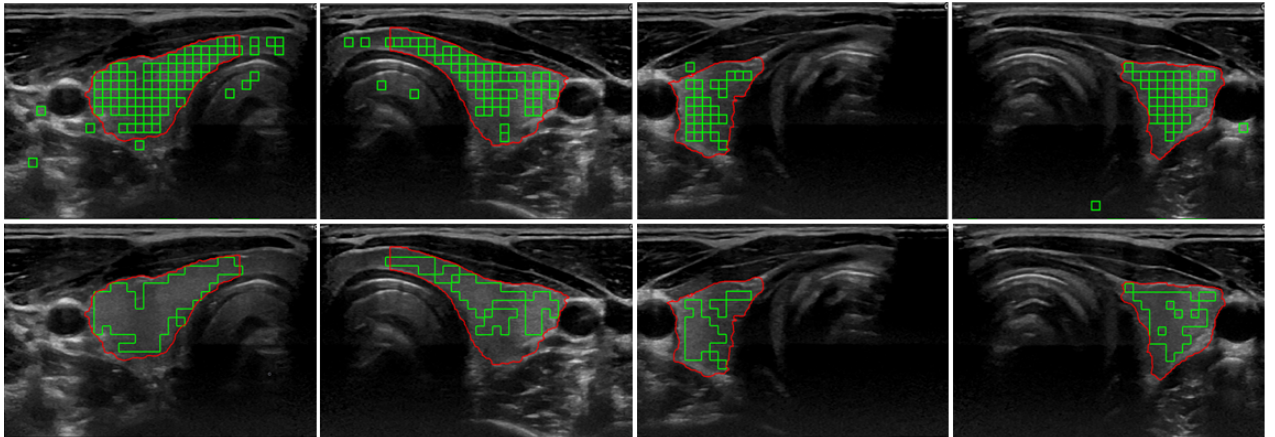
Fig. E.6: Examples of thyroid texture classification and segmentation using ANN and comparison with ground truth on Dataset 1.

E.4.3 Visual Analysis

The training of the three classifiers were followed by testing of individual images which were not part of the training set. An example of texture classification (first row) and segmentation (second row) on a total of 8 (4 from each dataset) different thyroid US images using SVM on Dataset 1 and 2 are shown in Fig. E.4 and E.5 respectively.

179

Fig. E.7: Examples of thyroid texture classification and segmentation using ANN and comparison with ground truth on Dataset 2.

Fig. E.8: Examples of thyroid texture classification and segmentation using RFC and comparison with ground truth on Dataset 1.
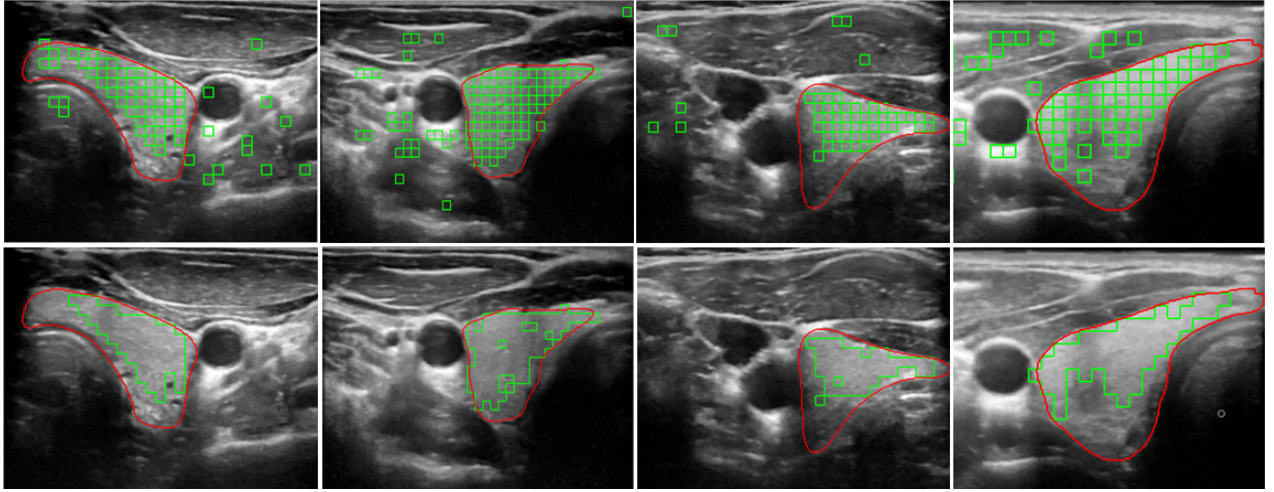
Similarly, the results using ANN on Dataset 1 and 2 on the same images as in SVM are shown in Fig. E.6 and E.7 respectively. Fig. E.8 and E.9 show the results using RFC on Datasets 1 and 2 respectively. The images in the first row in all the figures from E.4-E.9 show the results of texture classification using the trained classifiers and the images in the second row present the segmented thyroid regions after the post-processing step. In the figures, the green squares represent the 20 x 20 pixel texture patches classified as
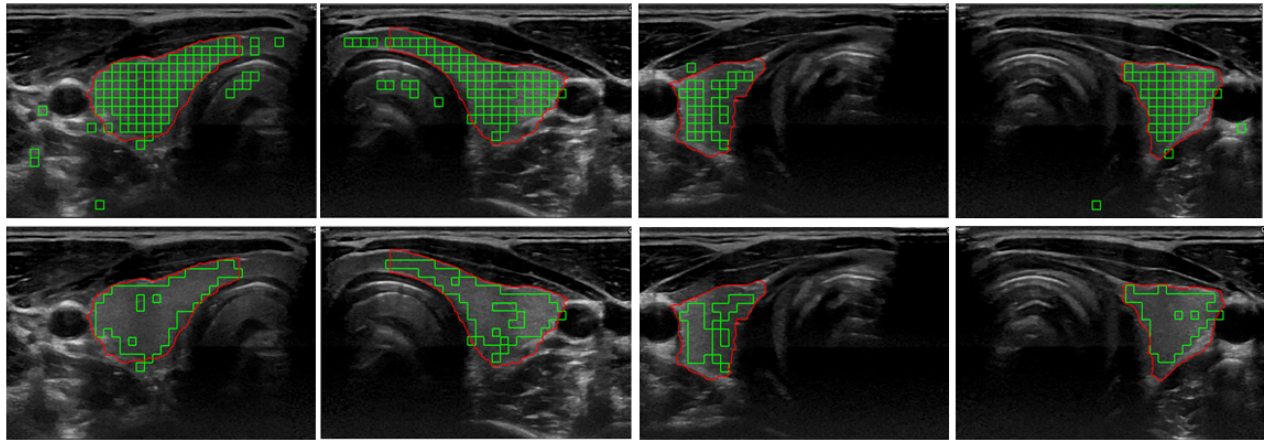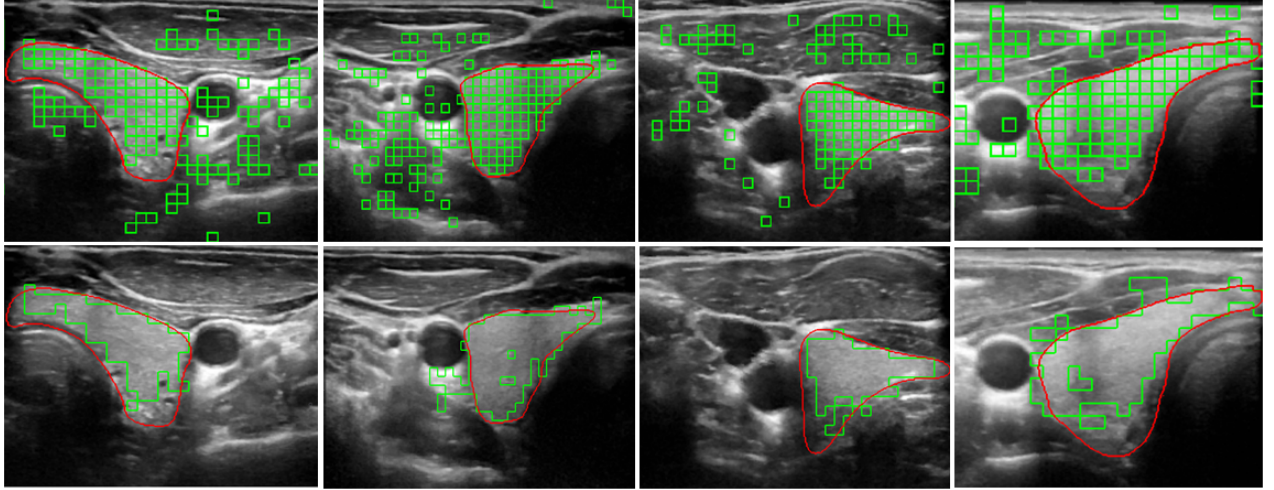
Fig. E.9: Examples of thyroid texture classification and segmentation using RFC and comparison with ground truth on Dataset 2.

thyroid and the solid red line represents the ground truth region manually annotated by the expert clinicians. For testing purposes, we took the thyroid images from different locations with respect to the thyroid volume and from different patients.

The images from the first row in all the figures (i.e. Fig. E.4, E.5, E.6, E.7, E.8 and E.9) show the texture classification results from the trained classifiers and the images from the second row show the post-processed segmented thyroid region (marked with solid green lines). As evident in the figures, there are some over-classifications of texture patches as thyroid. Hence, a post-processing step was carried out to get rid of these over-classified texture patches.

The sample test images along with their ground truth have been shown above. From these test images, we can see that this way of texture classification obtains the larger thyroid region compared to the thyroid segmentation using ACWE, GC and PBC as they fail to segment the isthmus region inside the thyroid [17]. Despite classifying the regions in the isthmus, our approach achieves few under-classified results inside the thyroid. This problem could be solved by calculating more features (energy based, entropy based,

statistical features, etc.) and using some extensive pre-processing techniques to choose the most prominent features like Principal Component analysis [33] and Information Gain Techniques [34].

E.5   Discussion and Conclusion

In this paper, we have compared the three different machine learning techniques (SVM, ANN and RFC) for thyroid texture classification and segmentation. We computed the features for training of these classifiers using a very novel feature extraction technique. A signal based version of the US image was used and parametrically modelled to compute AR features. This transformation of the image to signal possesses many advantages compared to the image based analysis. With this way of treating the images, the extracted features are not affected by the presence of speckle noise, low contrast issues and low SNR in thyroid US images. This allows the classifiers to classify the thyroid region even in the smaller areas (for example in the isthmus of the thyroid) which was difficult using some of the state of art methods (ACWE, GC and PBC).
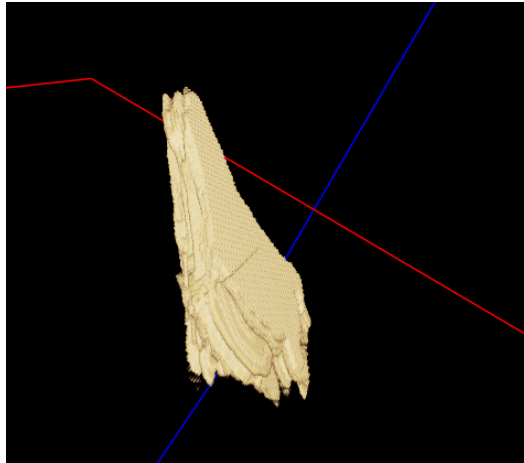

Fig. E.10: A 3D reconstructed thyroid after texture classification and segmentation using Imfusion.

We also performed a comparison analysis between our approaches and various approaches in the literature. Two different comparison analysis were performed, first by comparing the performance on the same datasets and second on different datasets. In all of the tests, our approach outperformed the approaches in the literature in terms of DSC and had similar SE and SP. The results that we have obtained show a close correlation to the ground truth data. While the accuracy of training of the classifiers are similar, ANN slightly outperformed SVM and RFC. Our approaches were fully automated, so the user did not have to invest time in tracking the progress of the segmentation like in ACWE where the user had to stop the process and run it again, if the initialization of the contour was outside of the thyroid region. Similarly, in graph cut the user had to remove the over-classified regions after the segmentation and in PBC, the user had to make more clicks inside and outside the thyroid regions to get a better estimate of the features in order to train their decision trees.

One of the main drawbacks of the proposed approach is that it has only been evaluated with thyroid images from healthy subjects. In the future, we will explore how nodules can change the spectral behaviour in the US image. Similarly, we have used the images from a high-end machine (i.e. Logiq E9) for our task and the images from low-end machine might not have the same segmentation accuracy as we have shown in this work. Additionally, the classification of the texture patches always produces a non-smooth boarder in the segmented images. However, the clinical relevance could be established by training the classifiers on pathological images and the problem of the rough boarder in the segmented images could be solved by taking overlapping patches or by using a multi-resolution patch size and using the best size that produces a smooth appearance in the boarder region.

As future works, more features can be computed by not modelling the US images using AR model but by other methods as well like Bispectral model [35] and these features

can be pre-processed by other pre-processing techniques such as Principal Component Analysis and Information Gain to select the prominent features. We could also combine the features from different modelling techniques and use them for the classification task. Similarly, the feature computation time can be reduced by optimizing the wavelet computation and AR modelling steps.

As mentioned above, the classified thyroid images can be reconstructed to a 3D volume as we also acquired the tracking data during the image acquisition phase. An example of the 3D reconstructed thyroid using Imfusion [37] after texture classification and segmentation is shown in Fig. E.10.The Imfusion software allows the user to input all the binary images obtained from the segmentation as a video file along with the tracking matrices associated with each image frames. The reconstruction is then carried by using a technique called volumetric compounding where an interpolation is carried out between the corresponding image frames to fill the empty spaces. The 3D volume information can be used clinically by the medical experts to monitor the state of thyroid over time. Since most of the thyroid diseases involve change in the shape and volume of thyroid over time, the 3D reconstruction and volume computation has a clinical relevance.

References

1. Understanding Thyroid Problems - the Basics. https://webmd.com/women/picture-of-the-thyroid.

2. T.L. Szabo. "Diagnostic Ultrasound Imaging: Inside Out, New York: Elsevier - Academic Series", 2004, Ch. 1.

3. A. Illanes, N. Esmaeili, P. Poudel, S. Balakrishnan and M. Friebe. "Parametrical modelling for texture characterization - a novel approach to Ultrasound thyroid segmentation", *PLOS ONE*, 2018.

4. J. Zhao, W. Zheng, L. Zhang, H. Tian, "Segmentation of ultrasound images of thyroid nodule for assisting fine needle aspiration cytology". *Health information science and systems*. Vol. 1, no. 1, 5, 2013.

5. J . Kaur, A. Jindal, "Comparison of thyroid segmentation algorithms in ultrasound and scintigraphy images". *International Journal of Computer Applications*. Vol. 50, no. 23, 2012.

6. D. China, A. Illanes, P. Poudel, M. Friebe, P. Mitra, D. Sheet "Anatomical Structure Segmentation in Ultrasound Volumes using Cross Frame Belief Propagating Iterative Random Walks". *IEEE journal of biomedical and health informatics.* 2018.

7. D. Selvathi, V. S. Sharnitha, "Thyroid classification and segmentation in ultrasound images using machine learning algorithms". *In Signal Processing, Communication, Computing and Networking Technologies (ICSCCN), International Conference on IEEE.* pp. 836-841, July, 2011.

8. N. H. Mahmood, A. H. Rusli, "Segmentation and area measurement for thyroid ultrasound image". *International Journal of Scientific and Engineering Research.* Vol. 2, no. 12, 2011.

9. E. G. Keramidas, D. K. Iakovidis, D. Maroulis, S. Karkanis, "Efficient and effective ultrasound image analysis scheme for thyroid nodule detection". *In International Conference Image Analysis and Recognition.* pp. 1052-1060. Springer, Berlin, Heidelberg, August, 2007.

10. D. E. Maroulis. M. A. Savelonas, D. K. Iakovidis, S. A. Karkanis, N. Dimitropoulos, "Variable background active contour model for computer-aided delineation of nod-

ules in thyroid ultrasound images". *IEEE Transactions on Information Technology in Biomedicine.* Vol. 11, no.5, pp. 537-543, 2007.

11. M.A. Savelonas, D.K. Iakovidis. I. Legakis, D. Maroulis, "Active contours guided by echogenicity and texture for delineation of thyroid nodules in ultrasound images". *IEEE Transactions on Information Technology in Biomedicine*, Vol. 12, no. 4, pp. 519-527, 2009.

12. H. Garg, A. Jindal, "Segmentation of thyroid gland in ultrasound image using neural network". *In Computing, Communications and Network Technologies (ICCCNT), Fourth International Conference on IEEE.* pp. 1-5, July, 2013.

13. N. S. Narayan, P. Marziliano, J. Kanagalingam, C. G. Hobbs, "Speckle Patch Similarity for Echogenicity-Based Multiorgan Segmentation in Ultrasound Images of the Thyroid Gland". *IEEE Journal of biomedical and health informatics.* Vol. 21, no. 1, pp. 172- 183, 2017.

14. E. N. Kollorz, D. A. Hahn, R. Linke, T. W. Goecke, J. Hornegger, T. Kuwert, "Quantification of thyroid volume using 3-D ultrasound imaging". *IEEE Transactions on medical imaging.* Vol. 27, no. 4, pp. 457- 466, 2008.

15. C. Y. Chang, Y. F. Lei, C. H. Tseng and S. R. Shih, "Thyroid Segmentation and volume estimation in ultrasound images". *IEEE Transactions on biomedical engineering.* Vol. 57 no.6, pp. 1348-1357, 2010.

16. A. Osman, "Automated evaluation of three dimensional ultrasonic datasets.", Doctoral dissertation, INSA de Lyon, 2013.

17. P. Poudel, A. Illanes, D. Sheet and M. Friebe, "Journal of healthcare engineering", 2018.

18. D.K. Iakovidis, E.G. Keramidas, D. Maroulis, "Fusion of fuzzy statistical dis-

tributions for classification of thyroid ultrasound patterns". *Artificial Intelligence in Medicine*, Vol. 50, pp. 836-841, 2010

19. D. Selvathi, V. S. Sharnitha, "Thyroid classification and segmentation in ultrasound images using machine learning algorithms". *In Signal Processing, Communication, Computing and Networking Technologies (ICSCCN), International Conference on IEEE.* pp. 836-841, July, 2011.

20. R. Koprowski, A. Korzynska, Z. Wrobel, W. Zieleznik, A. Witkowska, J. Malyszek, W. Wojcik. "Influence of the measurement method of features in ultrasound images of the thyroid in the diagnosis of Hashimoto's disease". *Biomedical engineering online.* Vol. 81, no. 1, pp. 91, 2012.

21. U.R. Acharya, S.V. Sree, M.M. Krishnan, F. Molinari, R. Garberoglio J.S. Suri. "Non-invasive automated 3D thyroid lesion classification in ultrasound: a class of ThyroScan™ systems". *Ultrasonics.* Vol. 52, no. 4, pp. 508-520, 2012.

22. U.R. Acharya, C. Pradeep, H. Fujita, B. Shreya, D. Sumeet, J.E.W. Koh, L.W.J Eugene, P. Kongmebhol, K.H. Ng. "Thyroid lesion classification in patient population using Gabor transform features from high resolution ultrasound images". *Knowledge-Based Systems.* Vol. 107, pp. 235-245, 2016.

23. U.R. Acharya, S.V. Sree, G. Swapna, S. Gupta, P. Molinari, R. Garberoglio, A. Witkowska, J. Suri. "Effect of complex wavelet transform filter on thyroid tumor classification in three-dimensional ultrasound". *Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine.* Vol. 227, no. 3, pp. 284-292, 2013.

24. U. Raghavendra, A. Gudigar, M. Maithri, A. Gertych, K.M. Meiburger, C. Hong Yeong, C. Madla, P. Kongmebhol, F. Molinari, K.H. Ng, U.R. Acharya. "Optimized multi-level elongated quinary patterns for the assessment of thyroid nodules in ultra-

sound images. Computers in Biology and Medicine". Vol. 95, pp. 55-62, 2018.

25. P. Poudel, A. Illanes and M. Friebe. "Ultrasound thyroid texture classification using a simple texture pattern characterization". *In Book of abstracts 51st annual conference of the German Society for Biomedical Engineering*, Vol. 62, Dresden, Germany, September, 2017.

26. P. Poudel, E. Ataide, A. Illanes, M. Friebe. "Linear Discriminant Analysis and K-Means Clustering for Classification of Thyroid Texture in Ultrasound Images". *In Proc IEEE Eng Med Biol Soc*, 2018.

27. T. Wunderling, B. Golla, P. Poudel, C. Arens, M. Friebe, C. Hansen. "Comparison of thyroid segmentation techniques for 3D ultrasound". *Proc. SPIE Med. Imaging*, pp. 1013317-1013317, 2017.

28. D.G. Manolakis, V.K. Ingle and S.M. Kogon. "Statistical and adaptive signal processing: spectral estimation, signal modelling, adaptive filtering, and array processing". *McGraw-Hill.* Boston, 2000.

29. S. Tsantis, D. Cavouras, I. Kalatzis, N. Piliouras, N. Dimitropoulos and G. Nikiforidis. "Development of a Support Vector Machine Based Image Analysis System for Assessing the Thyroid Nodule Malignancy Risk on Ultrasound". *Ultrasound in Medicine and Biology.* Vol. 31, no. 11, pp. 1451-1459, 2005.

30. A. Criminisi, J. Shotton. "Decision forests for Computer Vision and Medical Image Analysis". *Springer Science and Business Media*, 2013.

31. L. Bottou. "Stochastic gradient descent tricks". *In Neural networks: Tricks of the trade.* pp. 421-436. Springer, Berlin, Heidelberg, 2012

32. L. Breiman. "Random forests". *Machine Learning.* 45(1). pp. 5-32, 2001.

33. H. Abdi, L. J. Williams. "Principal component analysis". *Wiley interdisciplinary*

reviews: computational statistics. Vol. 2, no. 4, pp. 433-459, 2010.

34. S. Lei. "A feature selection method based on information gain and genetic algorithm". *In Computer Science and Electronics Engineering (ICCSEE), International Conference on IEEE.* Vol. 2, pp. 355- 358, 2012.

35. T. E. Hall, G. B. Giannakis. "Bispectral analysis and model validation of texture images". *IEEE Transactions on Image Processing.* Vol. 4, no. 7, pp. 996-1009, July, 1995.

36. SurgicEye GmbH - https://www.surgiceye.com/.

37. ImFusion GmbH - https://www.imfusion.de.

CHAPTER F

Patch Based Texture Classification of Thyroid Ultrasound Images using Convolutional Neural Network

Prabal Poudel[1], Alfredo Illanes[1] and Michael Friebe[1]

Abstract— Ultrasound (US) is an affordable and important imaging modality in medical imaging without potential hazards for patients and medical practitioners as compared to computed tomography (CT), magnetic resonance imaging (MRI), nuclear imaging, etc. Texture classification of anatomical structures in US images is an essential step for disease diagnosis and monitoring. In this work, we employed a convolutional neural network to segment thyroid gland in US images. This is particularly important for thyroid diseases diagnosis as they involve changes in the shape and size of the thyroid over time. The training of the CNN was not done directly on the acquired US images but on texture database that is created by dividing the thyroid US images of size 760 x 500 pixels into smaller texture patches of size 20 x 20 pixels. We obtained a Dice coefficient (DC) of 0.876 and Hausdorff Distance (HD) of 7.3 using the trained CNN that classifies the thyroid tissues as thyroid or non-thyroid. This approach was compared to the classic image processing approaches like active contours with edges (ACWE), graph cut (GC) and pixel based classifier (PBC) which obtained a DC of 0.805, 0.745 and 0.666 respectively and Volumetric and Mass Spring Models which obtained a HD of 11.1 and 9.8 respectively.

Augustin *et al.* [3] used fuzzy c-means algorithm, histogram clustering, QUAD tree, region growing and random walker to segment thyroid US images. Similarly, thyroid segmentation in US images using Iterative Random Walks and Random Forest was performed by China *et al.* [4].

Apart from segmentation in 2D, several research works have been proposed to segment a full 3D thyroid image. Dornheim *et al.* [5] proposed a 3D mass-spring model for thyroid cartilage segmentation by creating a 3D deformable shape models. Similarly, Osman [6] performed segmentation and analysis of 3D thyroid images by thresholding the voxel intensities and then connecting similar voxels to predict the segmenting regions.

The approaches described above use classical non machine learning methods to segment thyroid. However, in the recent years, there have been significant researches in thyroid segmentation using machine learning approaches. Selvathi et al. [7] have used a polynomial support vector machine (SVM) to segment the thyroid gland in US images. A feedforward

Abstract

Ultrasound (US) is an affordable and important imaging modality in medical imaging without potential hazards for patients and medical practitioners as compared to computed tomography which uses X-rays, magnetic resonance imaging which uses magnetic field and radio waves that could heat up the patient's body during long examinations, nuclear imaging, etc. Texture classification of anatomical structures in US images is an

essential step for disease diagnosis and monitoring. In this work, we employed a convolutional neural network to segment thyroid gland in US images. This is particularly important for thyroid diseases diagnosis as they involve changes in the shape and size of the thyroid over time. The training of the Convolutional Neural Network (CNN) was not done directly on the acquired US images but on texture database that is created by dividing the thyroid US images of size 760 x 500 pixels into smaller texture patches of size 20 x 20 pixels. We obtained a Dice coefficient (DC) of 0.876 and Hausdorff Distance (HD) of 7.3 using the trained CNN that classifies the thyroid tissues as thyroid or non-thyroid. This approach was compared to the classic image processing approaches like active contours with edges (ACWE), graph cut (GC) and pixel-based classifier (PBC) which obtained a DC of 0.805, 0.745 and 0.666 respectively and Volumetric and Mass-Spring Models which obtained a HD of 11.1 and 9.8 respectively.

F.1   Introduction

Thyroid is a butterfly shaped gland located below the Adam's apple on the front of the neck. It is involved in several human body mechanisms such as synthesis of proteins, energy sources usage and controlling body's sensitivity to other hormones. Thyroid diseases like Graves, sub-acute thyroiditis, thyroid cancer, goitre and thyroid nodule often involve changes in the shape and size of thyroid. Hence, volume computation of thyroid over time after segmenting the thyroid using texture classification helps us to monitor the state of thyroid. This is useful for the medical practitioners to diagnose the thyroid diseases. US imaging is widely used for data acquisition as it is much safer in terms of ionising radiation and painless when used on the patients.

Numerous approaches for segmentation of medical images in 2D and 3D using classical image processing techniques as well as machine and non-machine learning methods

191

have been proposed. Segmentation by edge detection, thresholding, region splitting and merging, watershed segmentation, active contour, graph theory and normalized cut were proposed by Zhao *et al.* [1] in 2D thyroid US images. Similarly, segmentation of thyroid in 2D US and scintigraphy using ACWE, localized region based active contour and distance regularized level set was performed by Kaur *et al.* [2]. Augustin *et al.* [3] used fuzzy c-means algorithm, histogram clustering, QUAD tree, region growing and random walker to segment thyroid US images. Similarly, thyroid segmentation in US images using Iterative Random Walks and Random Forest was performed by China *et al.* [4].

Apart from segmentation in 2D, several research works have been proposed to segment a full 3D thyroid image. Dornheim *et al.* [5] proposed a 3D mass-spring model for thyroid cartilage segmentation by creating a 3D deformable shape models. Similarly, Osman [6] performed segmentation and analysis of 3D thyroid images by thresholding the voxel intensities and then connecting similar voxels to predict the segmenting regions.

The approaches described above use classical non machine learning methods to segment thyroid. However, in the recent years, there have been significant researches in thyroid segmentation using machine learning approaches. Selvathi et al. [7] have used a polynomial support vector machine (SVM) to segment the thyroid gland in US images. A feedforward neural network [8] was used by Garg et al. to segment thyroid gland from US images. Similarly, Chang *et al.* [9] proposed radial basis function (RBF) neural network to segment the blocks of thyroid gland. Xiang *et al.* [10] similarly have used a cascaded CNN involving a U-shaped CNN and deep fully connected CNN which has also been used by Ma *et al.* [11] to segment nodules from thyroid images.

Since segmentation of thyroid plays a critical role in diagnosis of thyroid diseases, accurate segmentation is imperative for clinical analysis. The classical non-machine learning algorithms are not automatic and affected by the presence of noise (particularly speckle)

and intensity inhomogeneity of thyroid texture in US images. For example, all the approaches that use active contours require user initialization and wrong initialization can lead to wrong segmentation results. Approaches using edge detection, thresholding, histogram clustering, region growing require the US images to be pre-processed which is very difficult to get rid of all the speckle noise. Similarly, the machine learning methods are mainly based on training of the hand-crafted features extracted from the images which is highly time consuming and complicated.

In our work, we employ a deep convolutional neural network method to classify thyroid texture in US images. The novelty of our approach is that we use this classifier to train on the texture patches from a texture database that was created by dividing the images into smaller texture patches rather than working in a whole image. This possess greater advantage as compared to training the image as a whole because it allows the network to learn more localized features in a smaller region and also in segmenting the smaller regions of the thyroid (for example: the isthmus region), which are otherwise very difficult to segment using the classic approaches [13].
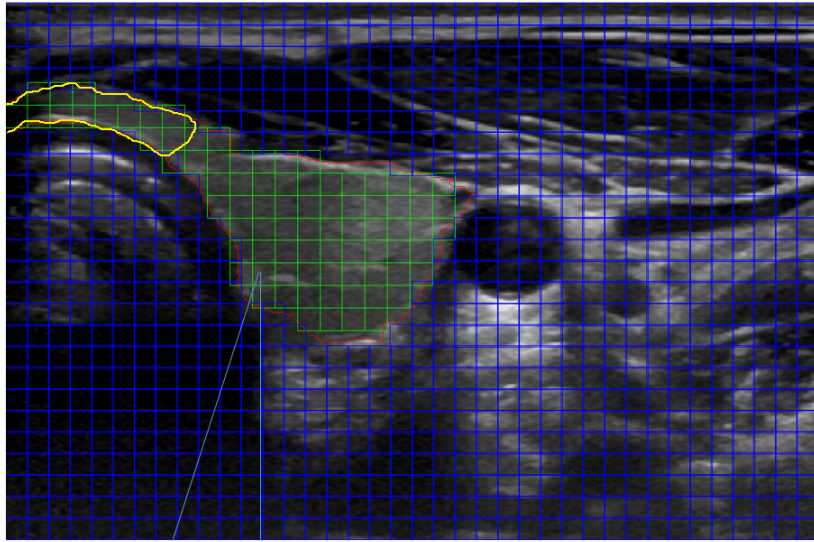

Fig. F.1: Separation of thyroid US image into texture patches. Blue: Non-Thyroid Patches, Green: Thyroid Patches, Red: Ground Truth, Yellow: Isthmus Region.

F.2   Methods and Procedures

F.2.1   Image Acquisition and Database Generation

The database that we use to evaluate our approach in this work has been introduced in
[12]. It consists of a total of *675* 2D thyroid US images from 6 healthy subjects with
an image size of *760 x 500* pixels. The images were acquired using a General Logiq
E9 system which was equipped with an Electromagnetic tracking system. A ML6-15
linear probe was used to acquire the US images. While acquiring the images, a tracking
matrix associated to each image was also acquired. These matrices could be used in the
future for volume reconstruction using the segmented thyroid images.

Table F.1: 2D US Images and Generated Texture Database

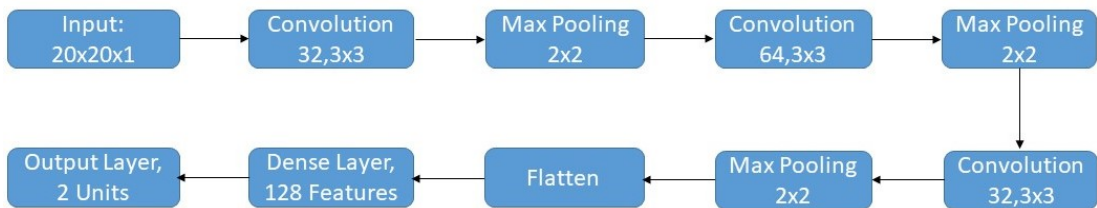| Dataset	Number of Images	Number of Texture Patches
D01	96	91,200
D02	53	50,350
D03	102	96,900
D04	63	59,850
D05	171	162,450
D06	210	199,500
Total	**695**	**660,250**

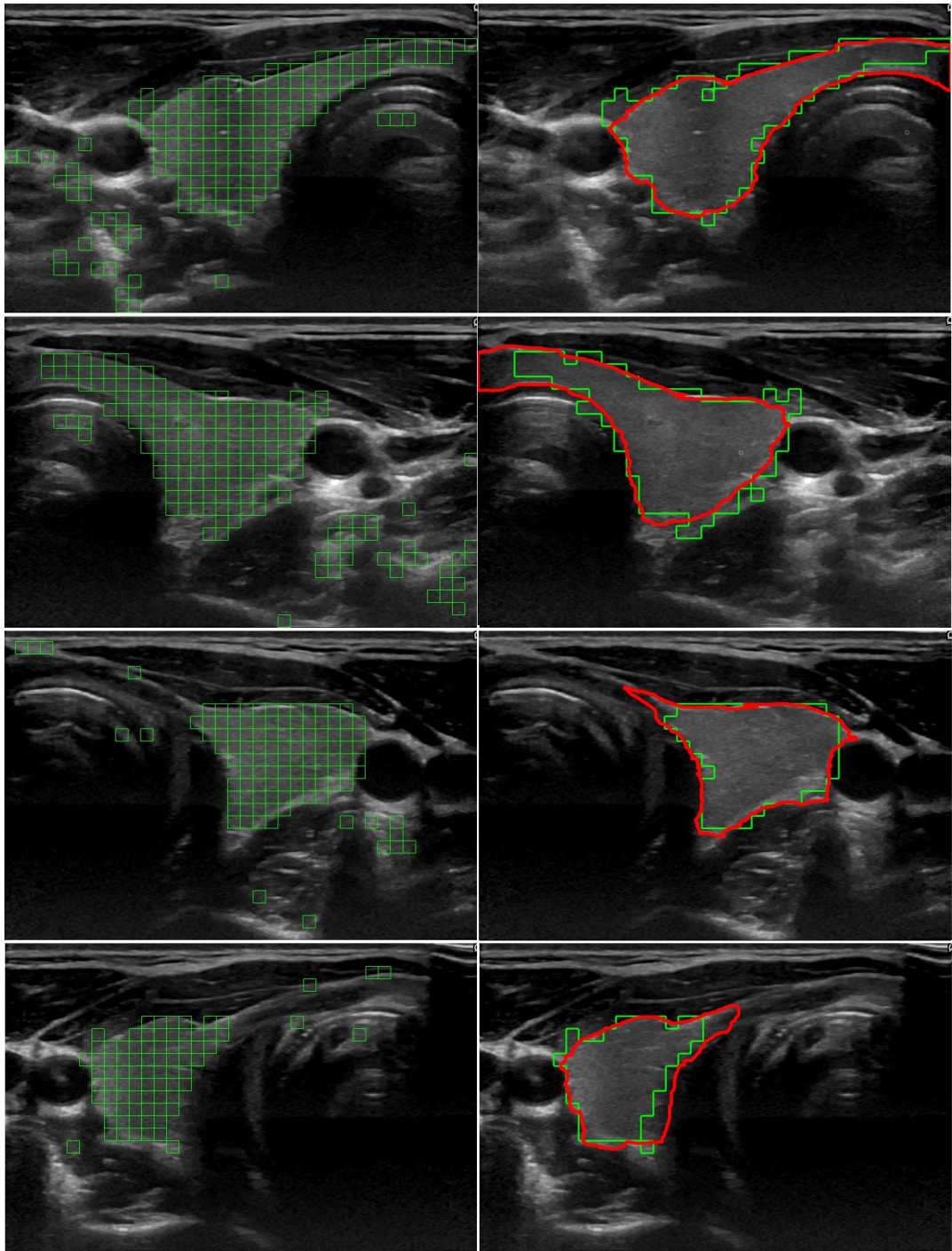Fig. F.2: Architecture of the CNN.

Fig. F.3: Left Column: Result of CNN Texture Classification, Right Column: (Green: Segmented thyroid region after post-processing step, Red: Ground Truth).
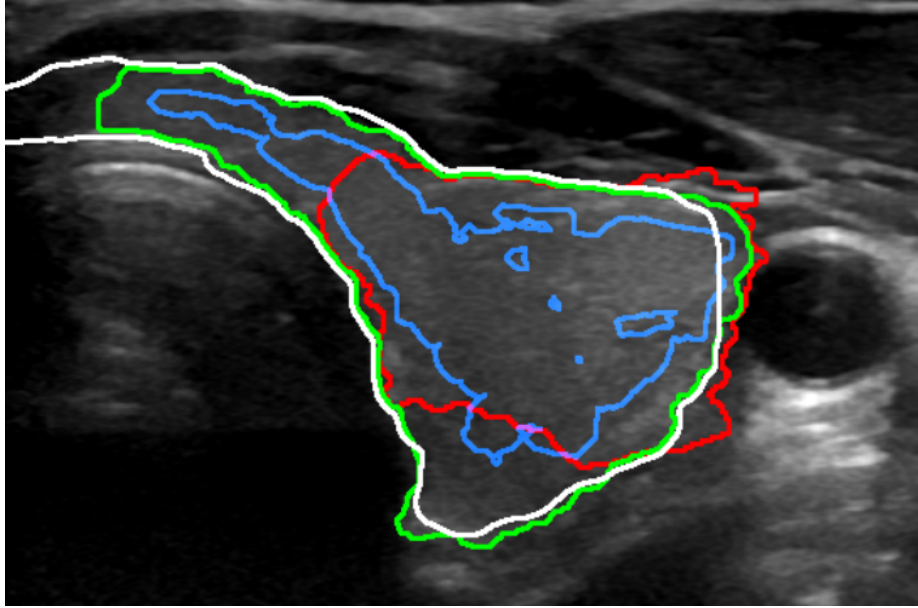
Fig. F.4: Segmentation of thyroid using ACWE (Red), GC (Green), PBC (Blue) and Ground Truth (White)

The images were then used to generate a database for the training of the CNN classifier. We divided each image in the dataset into smaller texture patches of size *20 x 20* pixels. This size was chosen empirically so that each texture patch has sufficient pixel information for feature extraction by the CNN and also perfectly cover the smaller parts of thyroid region (for example in the isthmus area as shown in Fig. F.1). Each texture patch was assigned a label (i.e. thyroid = 1 and non-thyroid = 0) by comparing to ground truth images obtained from the same clinical expert who acquired the thyroid US images. Finally, a texture database with *660,250* texture patches was generated (Table F.1). An example of separation of an image into texture patches along with the annotation of isthmus region can be seen in Fig. F.1. In the figure, blue patches represent non-thyroid patches and the green patches represent the thyroid patches and red lining represents the ground truth region obtained my manual segmentation by expert clinicians.

F.2.2 Architecture of the CNN

A deep, feed-forward CNN was implemented for the classification of thyroid regions and ultimately its segmentation. Fig. F.2 shows the network architecture which consists of a series of convolutional layers where the first convolution layer extracts very generalized features (such as edges, curvatures, pixel intensities, etc.) that can also be accessed by human eyes while the deeper layers extract more sophisticated and highly specific features from each texture patch. The input for the CNN consist of texture patches from the database and its corresponding label. The input can be represented as $X = (T_n, L_n)$, where T_n represents a texture patch and L_n represents its label.

The network consists of 3 convolution layers, starting with 32 filters of size *3 x 3*. We increase the number of filters by 2 times in each layer producing 128 in the final layer. The convolution layers are followed by a Leaky Rectified Linear Unit (Leaky ReLU) and max-pooling layer of size *2 x 2* with 0 stride. We choose Leaky ReLU as it attempts to solve the problem of ReLU as some units might die when a large gradient flows through a ReLU neuron. 'Categorical Crossentropy' was set as the loss function and 'adam' optimizer was used for the minimization of the loss function. In the last layer, a softmax activation was used with two units to predict the output probability of each pixel in each texture being either a thyroid or non-thyroid.

Data augmentation was not carried out as we had sufficient texture patches. In order to avoid the problems of over-fitting, we added a dropout of 0.25 after each max-pooling layer. Additionally, all the texture patches were normalized using Z-score method (i.e. subtracting the mean and then dividing by the standard deviation of the pixel intensities of each patch).

F.3   EXPERIMENTAL SETUP

F.3.1   Experimental Setup

In this work, we trained a CNN for classification of thyroid texture as a 2-class classification task. We also implemented three non-machine learning based approaches (ACWE, GC and PBC) to compare the performance of CNN with them. All of these approaches require some human interactions making them semi-automatic approaches. In ACWE, the user needs to initialize a mask by drawing a rectangle/square which evolves over different iterations to produce a segmented thyroid. Similarly, in GC, the user needs to scribble a thyroid and non-thyroid region which are modelled by different Gaussian Models to predict the probability of each pixel being either a thyroid or non-thyroid. In PBC, the user needs to click in the thyroid and non-thyroid region from which different features are extracted to separate the thyroid regions from non-thyroid regions.

For evaluation and quantitative analysis of the trained CNN, 10-fold cross validation experiments were performed. The overall dataset was first divided into 10 non-overlapping groups so that the same images are not used for both training as well as validation and testing. Out of the 9-fold data for training, 8-fold were actually used for training and 1-fold for validation. The validation set was used to fine-tune the parameters of the CNN and test set was used for performance analysis of the trained CNN. During the testing phase, the classified pixels in each textures were post-processed using a thresholding technique. Only the block of texture patches that contained more than 60% thyroid (=1) texture patches were considered as the thyroid patch while the rest as non-thyroid patch. Finally, the largest block of texture patches was considered as the thyroid region and the rest were disregarded. The classified texture patches (green square boxes) using CNN and the final segmented thyroid regions (green solid lines) after the post-processing steps are shown in first and second columns respectively in Fig. F.3.

All the experiments were carried out using a Lenovo T430 ThinkPad Notebook with Intel Core i5-3320M CPU, 2.60 GHz processor and 8.00 GB RAM. The operating platform were MATLAB2017 and Python 3.6. The training time was 2.45 hours where the CNN was trained for 20 epochs. DC and HD were used as the performance metrics.

F.3.2 Classification Results and Comparison with Other Methods

As mentioned earlier, 10-fold cross validation approach was used for the performance analysis of the CNN. We obtained a DC of 0.937 for training and 0.935 for validation and an average DC of 0.876 for testing on comparison with the ground truth for the 10 different tests we carried out. The result of texture classification and segmentation in four thyroid images from different subjects are shown in Fig. F.3. In the figure, green patches in the left column represent the classified thyroid patches before largest component analysis while the green region in the right column represent the segmented thyroid region after largest component analysis and extracting the borders from the thyroid patches. Similarly, the result of thyroid segmentation using ACWE, GC and PBC is shown in Fig. F.4. In the figure, white solid line represents the ground truth, ACWE is represented by red line, GC by green and PBC by blue region. From the figure, we can see that these approaches cannot segment the thyroid in narrow regions (for example isthmus) and often leads to under segmentation while CNN does not fail to segment these regions. The dividing of the image into smaller texture patches allows for these smaller to be segmented using CNN.

Table F.2: Comparison of Average Computation Time and Number of User Interactions

Method	Computation Time (sec)	Number of User Interactions
ACWE	369	8
GC	98	36
D03	10	5
CNN	13.65	0

Table F.3: 2D US Images and Generated Texture Database

Method	DC	HD (mm)
ACWE	0.805	8.2
GC	0.745	8.4
PBC	0.667	9.6
CNN	0.876	7.3

We also compared the computation time and number of user interaction required for all the 4 approaches (CNN, ACWE, GC and PBC) and present them in Table F.2. We can see that all the three non-machine learning based approaches require some level of human interaction for segmenting the thyroid. CNN however, does not require any human interaction and is fully automatic, which makes it potentially easier for the medical professionals to use this method. Also, the computation time for the three methods are higher compared to the CNN (exception: PBC) where ACWE takes 369

seconds, GC takes 98 seconds and PBC takes 10 seconds, while CNN only takes 13.65 seconds. These times are taken for segmentation of all the thyroid US images on average per test set. Similarly, the number of user interactions required are also on average per test set. The comparison of segmentation accuracy in terms of DC and HD in all the four methods are presented in Table F.3. These results prove that CNN outperforms all the other five methods in terms of accuracies, computation time and ease of use.

F.4 CONCLUSIONS

In this work, we employed a deep learning based approach (i.e. CNN) to classify the thyroid texture and finally segment the thyroid region after dividing the thyroid US images into smaller texture patches. The results obtained proved that this way of dividing the images into smaller texture patches achieves better thyroid segmentation as compared to classic non-machine learning approach which fail to segment the complex structures (for eg: isthmus) of thyroid. This is clearly visible in the obtained results in Fig. 3. However, a quantitative analysis could also be carried out by computing the accuracy of segmentation in the isthmus area. On top of that, CNN is fully automatic, fast and robust as compared to other methods which aids the medical practitioners in diagnosis and monitoring of thyroid over time with less effort. Similarly, there were no any cases that proved to be difficult to segment with this approach. However, all the US images were acquired using a single US machine (GE Logiq E9). Hence, in future, more images have to be acquired from different US devices using different acquisition settings (by varying the depth, gain, and other acquisition parameters) to prove the robustness of the applied approach. Nevertheless, we recommend using this method for classification and segmentation of textures from other organs in clinical settings.

References

1. J. Zhao, W. Zheng, L. Zhang and H. Tian, "Segmentation of ultrasound images of thyroid nodule for assisting fine needle aspiration cytology," Health information science and systems, volume 1(1), 5, 2013.

2. J. Kaur and A. Jindal, "Comparison of thyroid segmentation algorithms in ultrasound and scintigraphy images," International Journal of Computer Applications, volume 50(23), 2012.

3. A. S. Augustin, S. S. Babu and K. T. Nadu, "Thyroid segmentation on us medical images: An overview," 2012.

4. D. China, A. Illanes, P. Poudel, M. Friebe, P. Mitra, D. Sheet, Anatomical Structure Segmentation in Ultrasound Volumes using Cross Frame Belief Propagating Iterative Random Walks. IEEE journal of biomedical and health informatics, 2018

5. J. Dornheim, L. Dornheim, B.Preim, K.D. Tonnies, I.Hertel and G. Strauss, "Stable 3D mass-spring models for the segmentation of the thyroid cartilage," Otto-von-Guericke-Universitat Magdeburg, Germany, 2005.

6. A. Osman, "Automated evaluation of three dimensional ultrasonic datasets," Ph.D thesis, INSA de Lyon, 2013.

7. D. Selvathi and V. Sharnitha, "Thyroid classification and segmentation in ultrasound images using machine learning algorithms," IEEE, pages 836-841, 2011.

8. H. Garg and A. Jindal, "Segmentation of thyroid gland in ultrasound image using

neural network," IEEE, page 1-5, 2013.

9. C. Y. Chang, Y. F. Lei, C. H. Tseng and S. R. Shih, "Thyroid segmentation and volume estimation in ultrasound images," IEEE Transactions on biomedical engineering, volume 57(6), pages 1348-1357, 2010.

10. X. Ying, et al. Thyroid Nodule Segmentation in Ultrasound Images based on cascaded Convolutional Neural Network. in International Conference on Neural Information Processing (pp. 373-384). Springer, Cham, December 2018.

11. J. Ma, F. Wu, J. Zhu, and D. Kong, Cascade convolutional neural networks for automatic detection of thyroid nodules in ultrasound images. Medical physics, 44(5), pp.1678-1691, 2017.

12. P. Poudel, E. Ataide, A. Illanes, M. Friebe, Linear Discriminant Analysis and K-Means Clustering for Classification of Thyroid Texture in Ultrasound Images. Proc IEEE Eng Med Biol Soc, 2018.

13. P. Poudel, A. Illanes and M. Friebe, Evaluation of Commonly Used Algorithms for Thyroid Ultrasound Images Segmentation and Improvement Using Machine Learning Approaches. Journal of Healthcare Engineering, 2018.