FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO



Ovarian Structures Segmentation using a Neural Network Approach

Sónia Alexandra Cardoso Marques

Integrated Master in Bioengineering

Supervisor: Aurélio Joaquim de Castro Campilho, PhD

Second Supervisor: Catarina Brito Carvalho, PhD

July 23, 2019

© Sónia Alexandra Cardoso Marques, 2019

Ovarian Structures Segmentation using a Neural Network Approach

Sónia Alexandra Cardoso Marques

Integrated Master in Bioengineering

July 23, 2019

Resumo

A segmentação e deteção de estruturas anatómicas tem várias aplicações nas diferentes áreas da medicina. Em particular, na área de ginecologia e obstetrícia, a segmentação e deteção de estruturas do ovário são tarefas que, durante a prática clínica, se tornam demoradas, repetitivas e propensas a erro quando feita manualmente.

Uma das áreas de aplicação desta tarefa é a área de reprodução medicamente assistida, cujos tratamentos têm uma procura cada vez maior. Estes tratamentos requerem a análise frequente das imagens de ovários de exames de ultrassonografia, em relação à contagem e ao tamanho dos folículos. Também a área de investigação para a classificação de tumores no ovário beneficiaria de um método automático que efectuasse segmentação das estruturas do ovário de forma rápida e exata.

Tendo em conta a necessidade de um método de segmentação e deteção automática para estas estruturas, a investigação tecnológica encontra-se em busca de soluções eficazes para este fim.

Esta dissertação tem como objetivo o desenvolvimento de um algoritmo, baseado em redes neuronais convolucionais, que desempenhe essa tarefa.

Este documento apresenta conceitos chave para o entendimento do método desenvolvido, nomeadamente, os conceitos de ultrassonografia e *deep learning*, bem como o estado da arte na área da segmentação de folículos em imagens de ultrassom.

O modelo proposto é desenvolvido a partir de um modelo de base, para o qual são propostas alterações e passos adicionais na tentativa de melhorar o seu desempenho. O modelo de base utiliza imagens B-mode como dados de entrada de uma U-Net, otimizada utilizando uma função de perda baseada no coeficiente de Dice, para obter a segmentação de folículos e ovários, separadamente. As adaptações propostas para o modelo base incluem alterações em relação a 1) dados de entrada, nomeadamente o modo de ultrassom (B-mode e BRF), métodos de pré-processamento; 2) tamanho de imagem de entrada e parâmetros de arquitetura da U-Net, 3) implementação de uma nova função de perda e 4) métodos de pós-processamento.

Os resultados do modelo final demonstraram que este é apropriado para a execução da tarefa proposta mas demonstra limitações que impedem a sua implementação em prática clínica. Para além disso, as conclusões retiradas deste trabalho são relevantes para futura investigação na área.

No futuro, a função de perda proposta poderá ser aperfeiçoada e a informação de ultrassom em modo BRF pode ser explorada utilizando algoritmos the processamento de sinal em vez de algoritmos the processamento de imagem, para que o método possa utilizar a total informação, sem necessidade de processos de conversão para B-mode que levam à perda de informação. O aumento do conjunto de imagens e uma maior variedade the tamanhos de folículos presentes nas imagens melhorará o desempenho do modelo.

Abstract

The segmentation and detection of anatomical structures have many clinical applications in different medical fields. In particular, in the fields of gynecology and obstetrics, the segmentation and detection of ovarian structures is a task, which in clinical practice, becomes time-consuming, repetitive and error-prone when performed manually by the physicians.

One of the areas of application of this task is Assisted Reproductive Technology, whose treatments have been reporting an increase in the demand. These treatments require a frequent analysis of ultrasounds images of the ovaries, regarding the count and size of ovarian follicles. Also, the area of investigation of ovarian cancer would benefit from an automatic method that delivered the segmentation and detection of ovarian structures.

Taking into account the need for a method of automatic segmentation and detection of ovarian structures, scientific and technological investigation is working towards finding effective solutions for that end.

This dissertation's aim is the development of an algorithm based on Convolutional Neural Networks that performs this task.

This document presents key concepts for the understanding of the proposed model, namely the concepts of ultrasound imaging and deep learning, as well as the state-of-the-art in the area of follicle segmentation in ovarian ultrasound images.

The proposed model is developed based on a baseline model, to which modifications and additional steps are proposed in order to improve its performance. The baseline model uses B-mode images as input of a U-Net optimized using a loss function based on DSC, to obtain the prediction of the follicle or ovary segmentation. The proposed adaptations include alterations regarding 1) input data, namely the type of data (B-mode and BRF), and pre-processing methods, 2) input size and U-Net architecture parameters, 3) the implementation of a different loss function and 4) post-processing methods.

The results of the final model showed that it is suitable to perform the proposed task but it presents some limitations that do not allow its implementation in clinical practice. Furthermore, the conclusions taken from this work are relevant for future investigation in the area.

In the future, the proposed loss function can be refined and the US radio-frequency data can be explored using signal-processing algorithms instead of image-processing algorithms so that the model can utilize its complete information, without the data loss that the conversion from radio-frequency to B-mode entails. An increase in the size of the dataset and a more balanced variety of follicle size in the dataset would also improve the model's performance.

Acknowledgements

I would like to thank to Professor Aurélio Campilho, my supervisor, who guided me through the development of this dissertation and also inspired me to learn more about image analysis and computer vision over the years of my integrated master's degree.

To Catarina Carvalho, my co-supervisor, I thank for all the help and advice. Her continuous support was fundamental for the making and writing of this dissertation.

I thank my mother and sister for the love, patience and encouragement that shaped my path and led to the achievement of this milestone in my life.

Finally, I would like to thank to my boyfriend Sebastião and my best friends Daniela, Cátia and Filipa for their support during these last months and for all of the happy moments I shared with them throughout the years.

Sónia Marques

"People think of education as something that they can finish"

Isaac Asimov

Contents

1 Introduction **1**
- 1.1 Anatomy of the Ovaries . . . 1
- 1.2 Clinical Application of Follicle Segmentation . . . 3
- 1.3 Aim and Objectives of the Dissertation . . . 4
- 1.4 Contributions of the Dissertation . . . 5
- 1.5 Structure of the dissertation . . . 5

2 Ultrasound Imaging **7**
- 2.1 Ultrasound Acquisition . . . 8
- 2.2 Ultrasound Modes . . . 9
- 2.3 B-mode image formation . . . 10
- 2.4 Evolution of Ultrasounds for Clinical Purposes . . . 12
- 2.5 Concluding Remarks . . . 13

3 Neural Networks and Deep Learning **15**
- 3.1 Basic Neural Networks . . . 15
- 3.2 Artificial Neural Networks . . . 17
  - 3.2.1 Convolutional Neural Networks . . . 19
  - 3.2.2 Learning and Neural Network Optimization . . . 22
- 3.3 Performance Evaluation Measures for Segmentation . . . 24
- 3.4 Convolutional Neural Networks for the Segmentation of Biomedical Images . . . 27
- 3.5 Concluding Remarks . . . 29

4 Segmentation in ultrasound images. Its application for the segmentation of follicles in ultrasound images **31**
- 4.1 State-of-the-art . . . 31
  - 4.1.1 Segmentation of general hypoechoic anatomical structures . . . 32
  - 4.1.2 Segmentation of follicles . . . 39
- 4.2 Concluding Remarks . . . 45

5 Methodology **47**
- 5.1 Objectives and Challenges . . . 47
- 5.2 Dataset . . . 49
- 5.3 Equipment and tools . . . 50
- 5.4 U-Net Architecture . . . 51
- 5.5 Learning Framework . . . 52
- 5.6 Baseline Method . . . 54
- 5.7 Proposed Experiments . . . 54

5.7.1 Input Data and Pre-processing Methods . . . 55
5.7.2 Data Dimensioning and U-Net Architecture Hyperparameters . . . 58
5.7.3 Loss Function . . . 59
5.7.4 Post-processing Methods . . . 62
5.8 Performance Measures . . . 64
5.8.1 Segmentation Measure . . . 65
5.8.2 Detection Measure . . . 65
5.8.3 Partition Distance Measure . . . 66
5.8.4 Model Validation and Comparison . . . 66

6 Results and Discussion **69**
6.1 Input Data and Pre-processing Methods . . . 69
6.2 Data Dimensioning and U-Net Architecture Hyperparameters . . . 71
6.3 Loss Function . . . 73
6.4 Preliminary Model and Post-processing Methods . . . 74
6.5 Final Model . . . 77
6.6 Comparison with State-of-the-art . . . 79

7 Conclusion **83**

A Dataset Division **85**

B U-Net Convergence - Learning Rate Hyperparameters **87**

References **89**

List of Figures

1.1 Ovary, uterus and fallopian tube, posterior view. 1
1.2 Cross section of an ovary and of an antral follicle. 2

2.1 Example of the correspondence between the scan line and intensity signal in image. (a) Ultrasound scan lines. (b) A scan line in the B-mode US image. (c) Intensity signal along the scan line after conversion to B-mode. 9
2.2 Formation of a B-mode image. The image is built up line by line as the beam is stepped along the transducer array. 9
2.3 Linear B-mode transvaginal ultrasound shows a well defined hypoechoic mass. . 10
2.4 Example of the conversion of BRF to B-mode data. 11

3.1 Comparison of sigmoid functions with diferent parameter's values to illustrate the influence of the weights in the model of the neuron. 17
3.2 Comparison of sigmoid functions with diferent parameter's values to illustrate the influence of the bias in the model of the neuron. 17
3.3 Basic model of the neuron. 18
3.4 Basic architecture of a NN. 19
3.5 Neural network with a sequence of convolutional layers. 19
3.6 Illustration of the input (on the left) and output (on the right) size of a 2D convolution with a kernel of size 3×3 and stride 1 and stride 2. 20
3.7 Examples of the application of zero-padding and reflection. 21
3.8 Convolutional Neural Network with 8 layers alternating between convolution and downsampling using max-pooling. 22
3.9 Illustration of the sequence of training steps for neural network optimization. . . 23
3.10 Confusion Matrix for a two–class problem. 24
3.11 Example of four AUC-ROC Curves and the respective AUC-scores. 26
3.12 U-Net architecture. 27
3.13 Schematic representation of the V-net architecture. 28

4.1 Example of the results of the segmentation processing. 32
4.2 Example of the initial contour estimation results. 33
4.3 MAD (represented with circles) and standard deviation (represented with stars) between the segmentation results and the ground-truth. 34
4.4 Example of the segmentation results 34
4.5 Example of the segmentation results. 35
4.6 Example of the segmentation results 37
4.7 Example of the segmentation results 37
4.8 Example of the segmentation results. The method was tested in typical prostate images and prostate images with shadow artifacts and the presence of needles. . . 38

4.9 Examples of the segmentation results. 40
4.10 Illustration of the definition of major axis length and minor axis length. 41
4.11 Examples of segmentation results of different volumes (shown in 2D). 42
4.12 Example of segmentation results. 43
4.13 Example of image superimposed with the segmentation boundaries. 43
4.14 Example of image superimposed with the 3D segmentation. 44
4.15 Examples of segmentation results. Background area in red, follicle area in blue and ovarian stroma area in green. 45

5.1 Examples of linear B-mode images of the dataset to demonstrate the size and shape variability of the follicles. GT delineations of the follicles are in green. 48
5.2 Examples of linear B-mode images of the dataset. 48
5.3 Example of corresponding intensity profiles of one pixel column in a image in a) original BRF data; and b) B-mode data. 49
5.4 Example of (a) a linear B-mode image; (b) GT of a) with follicles represented in white and stroma in grey. 50
5.5 Illustration of the implemented U-Net architecture. 51
5.6 Schematic overview of the learning framework. 52
5.7 Schematic overview of the baseline method. 54
5.8 Schematic overview of the proposed methodology. The steps that comprise proposed experiments are presented in grey boxes. 55
5.9 Example of effect of the despeckling filter and of CLAHE. (a) Original B-mode image; (b) Results of the despeckling filter; (c) Result of CLAHE; (d) Result of despeckling filter followed by CLAHE. 57
5.10 Illustrations of the impact of size of the feature map in the lowest level of resolution and kernel size of U-Net convolutions, on the amount of context that the U-Net is able to apprehend. The white grid represents the number of pixels of the feature map in the lowest level of resolution and the colored squares represent the kernels of the convolution. 58
5.11 Illustrations of the influence on DSC of different segmentation errors (compared with the same segmentation reference). (a) Correct object count, with considerable distance between borders; (b) Incorrect object count due to FP objects, but minor distance between borders; (c) Incorrect object count due to incorrect union of objects and FN object, but minor distance between borders. The TP area is in yellow, FP area is in red and FN area is in green. 60
5.12 Examples of the produced masks with the areas defined as in between or near follicles in white, and corresponding B-mode images. (a) and (e) B-mode images; (b) and (f) Loss1; (c) and (g) Loss2; (d) and (h) Loss3. The masks have three distinct areas: follicle area (gray), high penalization background area (white) and low penalization background area (black). 60
5.13 Example of appropriate cases for the application of (a) Method Selection; (b) Method Removal, with a small FP object with 52 pixels (white arrow). The TP area in yellow, FN area in green and FP in red) 63
5.14 Example of the result of the Watershed algorithm. (a) B-mode image (b) Example of follicle segmentation, with TP area in yellow, FN area green and FP area in red; (c) Distance transform of the segmentation; (d) In white is the combination of the areas detected as being undoubtedly foreground and background and in black is the area called 'unknown'. (e) Delineation result of Watershed algorithm. 64

5.15 Illustrations of the computation of the TP, FN and FP objects for the object-wise analysis.The TP area is in yellow, FP area is in red and FN area is in green.. . . . 66
5.16 Illustration of boxplot interpretation. 67

6.1 Boxplot of the results of the experiment regarding input data and pre-processing methods for follicle segmentation. 70
6.2 Boxplot of the results of the experiment regarding input data and pre-processing methods for ovary segmentation. 70
6.3 Boxplot of the results of the experiment regarding input size and U-Net hyperparameters, for follicle segmentation. 71
6.4 Boxplot of the results of the experiment regarding input size and U-Net hyperparameters, for ovary segmentation. 71
6.5 Boxplot of the results of the experiment regarding the proposed definition for the loss function, for follicle segmentation. DSC in the legend of the figure corresponds the results of the baseline model. 73
6.6 Example of the results of the experiment regarding the proposed loss function. (a) B-mode image (b) Result of the baseline model; (c) Result of the model using the proposed loss function with Mask Loss2 and Parameters A. 74
6.7 Boxplot of the results of the experiment regarding the post-processing methods, for follicle segmentation. 75
6.8 Example of the results of the implementation of the Method Separation. (a) B-mode image; (b) Result from the baseline model; (c) Results after implementation of the Method Separation. 75
6.9 Example of the results of the implementation of the Method Out of Ovary. (a) B-mode image; (b) Result from the baseline model; (c) Results after implementation of the Method Out of Ovary. 76
6.10 Boxplot of the results of the experiment regarding the post-processing method, for ovary segmentation. 76
6.11 Example of the results of the implementation of the Method Selection. (a) B-mode image; (b) Result from the baseline model; (c) Results after implementation of the Method Selection. 77
6.12 Examples of (a) the best DSC results and (b) the worst DSC results, for ovary segmentation. 78
6.13 Examples of the worst DSC results (a) Class 1 images (b) Class 4 images. 79
6.14 Examples of the best DSC results (a) Class 2 images (b) Class 3 images. 79

B.1 Loss value progression for the validation set with different initial learning rates, for a factor of reduction of learning rate of 5. 87
B.2 Loss value progression for the validation set with different initial learning rates, for a factor of reduction of learning rate of 10. 88
B.3 Loss value progression for the validation set with different initial learning rates, for a factor of reduction of learning rate of 20. 88

List of Tables

3.1 Description of some of the rates derived from TP/FP/TN/FN values. 25

4.1 Overview of the described methods for ovarian follicle segmentation and respective performance measures. 46

5.1 Combinations of hyperparameters to study the influence of the amount of image context captured by the U-Net. 59

6.1 Results of the final model for the total dataset and for each class division, for follicle segmentation. 78

6.2 Overview of the performance of the state-of-the-art methods for follicle segmentation and detection. 80

6.3 Results for follicle segmentation of the proposed method for the total dataset and for each class division (regarding follicle size). 80

A.1 Dataset division for the validation of the final model. 85

Abbreviations and Symbols

1D	One Dimensional
2D	Two Dimensional
3D	Three Dimensional
ACC	Accuracy
AFC	Antral Follicle Count
ANN	Artificial Neural Network
ART	Assisted Reproductive Technology
BCE	Binary Cross-Entropy
BRF	Beam-formed Radio-Frequency
CLAHE	Contrast Limited Adaptive Histogram Equalization
CNN	Convolutional Neural Network
DSC	Dice Similarity Coefficient
DL	Deep Learning
FN	False Negative
FNR	False Negative Rate
FP	False Positive
FPR	False Positive Rate
GA	Genetic Algorithm
GLCM	Gray-Level Co-Occurrence Matrix
GT	Ground Truth
LSTM	Long Short Term Memory
ML	Machine Learning
MLP	Multilayer Perceptron
MNN	Modular Neural Networks
MSE	Mean-squared Error
NN	Neural Network
PD	Partition Distance
PSO	Particle Swarm Optimization
RBF	Radial Basis Function
ReLU	Rectified Linear Unit
RNN	Recurrent Neural Networks
ROI	Region Of Interest
SGD	Stochastic Gradient Descent
SVM	Support Vector Machines
TN	True Negative
TNR	True Negative Rate
TP	True Positive
TPR	True Positive Rate
US	Ultrasound

Chapter 1

Introduction

The segmentation of anatomical structures is a task recurrently demanded in numerous applications, in clinical practice. The segmentation task refers to the partition of a digital image into sets of pixels (or voxels, in the case of 3D imaging). The image pixels can be categorized into foreground and background. The former consists in the so called "Region of Interest" (ROI), which is a common term in medical imaging used to indicate the pixels of the object that needs to be identified, and the latter is the surrounding area, in other words, everything in the image that is not foreground. The foreground, in this case, is the ovary and a set of ovarian structures, called follicles.

The aim of this dissertation is the development of an automatic method that performs the segmentation and detection of ovarian follicles and of the ovary in ultrasound (US) images, by means of an algorithm based on a Convolutional Neural Network (CNN). CNNs are a type of Artificial Neural Network (ANN), which is an artificial intelligence framework that tries to replicate the network structure of the human brain and the behaviour of neurons transmitting electrical signals, to solve complex signal processing or pattern recognition problems [1].

1.1 Anatomy of the Ovaries

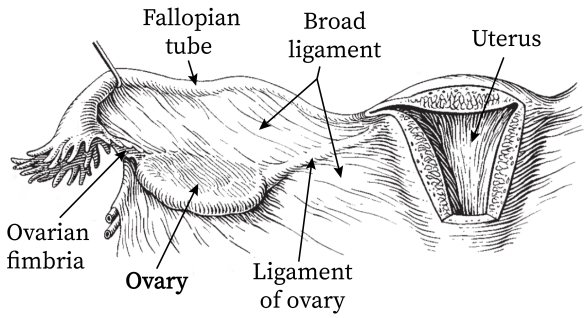


Figure 1.1: Ovary, uterus and fallopian tube, posterior view (adapted from [2]).

Ovaries are the female gonads and are paired pelvic organs with an almond shape, situated on each side of the uterus, near the pelvic wall (Figure 1.1). Their size and weight depends greatly on their follicular content, but they are roughly 3.0-5.0 cm $\times$ 1.5-3.0 cm $\times$ 0.6-1.5 cm, and 5-8 g [3].

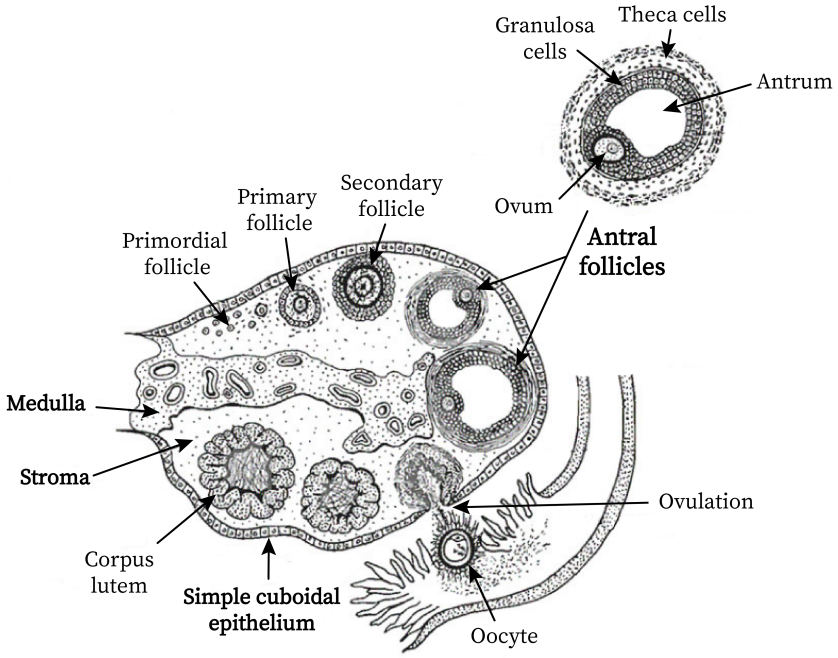


Figure 1.2: Cross section of an ovary and of an antral follicle (adapted from [4]).

The main parts of the ovary (Figure 1.2) are: 1) the surface, composed by simple cuboidal epithelium; 2) the stroma, composed by connective tissue, where the ovarian follicles, blood and lymph vessels and endocrine cells are found; 3) the medulla, also composed by connective tissue, containing neurovascular structures.

Besides, ovaries also contain follicles which are small structures that contain one immature egg, and other cell types (theca and granulosa cells) responsible for the production of hormones required for the egg's maturation[1]. It is impossible to accurately count the total number of follicles in an ovary, but before puberty there can be hundreds of thousands[2].

There are four distinct stages of follicle development: primordial, primary, secondary and tertiary (or antral) [5]. It takes almost one year for a primordial follicle to grow and develop to the ovulatory stage, which results in the release of the mature ovum[1.1]. After ovulation, the remaining granulosa and theca cells form the corpus luteum. After that, the type of hormone production carried out by the corpus luteum depends on whether or not the fertilization of the released oocyte occurs[3].

The primordial follicles are microscopic, measuring only around 0.025 mm, undetectable on an US or any other clinical test[4]. Alternatively, and more relevant to the clinical practice is the

[1] www.glowm.com/section_view/heading/Follicle%20Growth%20and%20Development/item/288
[2] https://extendfertility.com/your-fertility/egg-count
[3] http://medcell.med.yale.edu/histology/ovary_follicle.php
[4] https://ivi-fertility.com/blog/follicle-how-many-follicles-do-you-need/

study of antral follicles, which are visible and more mature. Antral follicles are characterized by the presence of an antrum, which is a cavity containing follicular fluid [5], and present a diameter between 2 and 30 mm [6].

1.2 Clinical Application of Follicle Segmentation

One of the main goals of follicle segmentation and detection in US images is to obtain the antral follicle count (AFC).

AFC is helpful for the prediction of the risk of menopause and for the investigation of ovulatory dysfunction, among others, but the main purpose for which it is requested is for infertility and Assisted Reproductive Technology (ART) treatments, as a way to evaluate ovarian stimulation response. It has proven to be one of the best indicators of women's fertility and predictors of the outcome of ART procedures, along with other ovarian parameters obtained through transvaginal US images, such as the mean follicular diameter, growth rate of dominant follicles and follicular wall thickness [7]. With this information, physicians are able to individualise and optimise the protocols of the treatment, in order to achieve better results [8, 9].

A worldwide web-based survey [10] reported that 71.2% of the physicians considered AFC as a mandatory part of their practice and 89.7% would modify the ovarian stimulation protocol based on the AFC.

More and more couples with infertility problems resort to ART procedures to overcome that issue. In 2017, the number of in vitro fertilization (IVF) procedures performed worldwide surpassed two million, and in some countries the percentage of newborns that result from this type of treatment reaches 5-6%[5].

The success rates of ART procedures depend on the patient and treatment characteristics, such as age, infertility diagnosis, type of ART procedure and history of previous births, among others[6]. According to the Portuguese Association of Fertility[7], the success rates of both intrauterine insemination and IVF (the most commonly used ART techniques) are lower than 20%, even for women under 39 years of age.

Infertility results in a prolonged and stressful period of time for the couples, putting at risk their mental health and the relationship between the two partners [11]. Having a way to get more accurate predictions for the results of this type of treatment, and to potentially improve them, would help to manage patients' expectations and make the process less frustrating and cumbersome for them.

Monitoring the growth rate of a particular follicle during the menstrual cycle and counting the number of smaller follicles requires a level of detail for which there are still no ideal solutions available. For a doctor to examine the evolution of a woman's ovarian condition and determine ART relevance, US scans must be performed and analysed every 1 to 3 days. If the assessment

[5] www.cije.up.pt/publications/debatendo-procria%C3%A7%C3%A3o-medicamente-assistida-0
[6] https://www.cdc.gov/art/artdata/index.html
[7] http://www.apfertilidade.org/Infertilidade_Tecnicas.html#text6

of the sonographic parameters is done manually, which is the norm in clinical practice, the workload becomes overwhelming and may discourage physicians to perform this investigation in some cases [7]. Thus, there is still room for research in this area, not only in the improvement of quality of the US acquisition but also of the analysis of the US images.

The ovarian follicles segmentation is also relevant for the research on ovarian cancer, which is one of the pathologies with worse prognostic in adult women [12]. The majority of ovarian cancers grows as a cystic mass, resulting from an abnormal development of ovarian follicles [13]. Ovarian cysts, unlike healthy follicles, do not contain a oocyte, and contain a different tissue composition [12].

The majority of the cystic masses detected by B-mode US screening are benign, but there are specific types of ovarian masses that cannot be diagnosed with certainty until surgery [14]. This limitation in cyst classification regarding its malignancy leads to unnecessary oophorectomys which have long term impact on both women's fertility and hormonal balance. Thus, research on computer-aided methods that can provide a second opinion to gynaecologists is valuable. The classification of these masses should be preceded by an accurate segmentation task of such ovarian structures, and hence the relevance of this dissertation.

For the assessment of ovarian follicles, transvaginal US is acquired using an US central frequency higher than 7 MHz [5]. Antral follicles are seen as round or oval hypoechoic structures, that is, they appear as a darker area in the US, as they are fluid filled and produce no internal echoes. Other structures inside the ovary may also appear as hypoechoic structures, such as small vessels, and so it is a challenge to achieve accurate follicle quantification. Other difficulties encountered for this task are: strong noise of the US signal, acoustic shadowing, follicles being too close to the probe or too close to each other, follicular borders being irregular and the large range of size of the follicles of clinical interest [15].

1.3 Aim and Objectives of the Dissertation

The aim of this dissertation is the development of an automatic method based on a U-Net which performs the segmentation and detection of the ovary and the ovarian follicles.

The objectives set to achieve this aim are:

- Study the acquisition of US images and the information those images display.
- Investigate the mechanisms of CNNs and their application on segmentation problems.
- Research the state-of-the-art of image analysis and computer vision techniques for the automatic or semi-automatic segmentation of ovarian follicles in US images.
- Develop a baseline model for follicle and ovary detection and segmentation.
- Examine the influence of the type of input data and pre-processing methods on the model's performance.
- Examine the importance of the image context analysis on the model's performance.

- Develop a loss function to prevent the incorrect union of neighboring follicles in the segmentation predicted by the model.
- Examine of the effect of post-processing methods on the model's performance.
- Establish and validate the final model for follicle and ovary detection and segmentation.

1.4 Contributions of the Dissertation

This work is a detailed analysis of the potential of the U-Net architecture to perform the segmentation and detection of the ovarian follicles and of the ovary.

The contributions of this dissertation can be summarized as follows:

- Report of the impact on model's performance of radio-frequency data as an alternative and as complement to B-mode data, as input data.
- Report of the impact on model's performance of the application of pre-processing methods, namely a despeckling filter and contrast-limited adaptive histogram equalization.
- Report of the impact on model's performance of the image context analysis performed by the U-Net, through the adjustment of the size of the input data and architecture hyperparameters such as the number of levels of resolution and kernel size.
- Development a novel loss function to prevent the incorrect union of neighboring follicles in the segmentation predicted by the model.
- Development an innovative post-processing method to combine the advantages of the implementation of the loss function of the baseline model and of the aforementioned novel loss function.
- Implementation of post-processing methods based on *a priori* information about the characteristics of the follicles and the ovary.
- Implementation and validation of the final model for follicle and ovary segmentation and detection.
- Writing of paper as co-author, titled "Deep learning approaches for gynaecological ultrasound image segmentation: a radio-frequency vs B-mode comparison" accepted for the 2019 International Conference on Image Analysis and Recognition (ICIAR).
- Writing of paper abstract as first author, titled "Post-processing and architecture influence on ovary and follicle ultrasound image segmentation", accepted for the 2019 IEEE International Ultrasonics Symposium (IUS).

1.5 Structure of the dissertation

This document is composed by seven chapters. The first chapter introduces the main concepts of this dissertation, giving a brief description of the anatomy of the ovaries and the clinical application

of follicle segmentation, which leads to the motivation for the dissertation. The aim, objectives and contributions of this dissertation are presented as well.

In Chapter 2, US Imaging is presented, including its acquisition process and the basic physics for its formation. Also, the different modes of US and the evolution on US technology for clinical purposes are introduced.

In Chapter 3, the NN and Deep Learning concepts are explained, namely the concept of CNN, including the description of the U-Net and V-Net architectures. A description of NN optimization is given as well. The most common performance evaluation measurements for segmentation are presented.

In Chapter 4, the State-of-the-art of segmentation in US images is presented. It includes segmentation methods of other anatomical structures whose ultrasonic representation is similar to that of ovarian follicles, and then segmentation methods of ovarian follicles specifically.

Chapter 5 describes the implemented methodology for the baseline model and the conducted experiments regarding the proposed alterations to improve performance and select the final model. Chapter 6 presents the corresponding results and discussion.

In Chapter 7, the conclusions and future work are presented.

Finally, Appendixes A and B present, respectively, the division of the implemented dataset and the sensitivity study regarding learning rate hyperparameters.

Chapter 2

Ultrasound Imaging

Ultrasound (US) Imaging or Sonography is a medical imaging technique that uses high-frequency sound waves, that range from 1 to 20 MHz, to acquire images of internal body structures, which are called sonograms [16].

This technology's early developments can be traced back to the 1940s, when the Austrian neurologist, Dr. Karl Theo Dussik, applied it for the first time as a medical diagnostic tool to obtain transcranial images [17]. Nowadays there are several applications of US in medical diagnosis. Sonograms and sonometry are valuable tools that can be used to evaluate bone fragility and blood flow, listen to the fetal heartbeat, guide minimally invasive biopsies and needle placement for anesthesia and, in general, assess the anatomical and physiological conditions of many different organs and structures of the human body[1].

US imaging offers many advantages, which make it one of the leading imaging modalities worldwide [17]. First off, it is rather inexpensive and its equipment is easily transported. Also, for the patient, it is a very practical exam as it requires little to no preparation and it is non-invasive [16].

Probably the biggest advantage of this modality is that it is based on non-ionizing radiation, so it does not have any cumulative biological side effects and there is no strict limit for the number of sonograms that a patient can take per year. This is very important for ART treatments as the monitorization requires frequent assessment of sonographic parameters through transvaginal US [16].

Transvaginal US is indicated for the visual evaluation of pelvic structures, namely in the cases of follicle monitoring, assessment of pelvic masses, endometrial studies, early pregnancy observations, among others [18].

[1] https://www.fda.gov/radiation-emittingproducts/radiationemittingproductsandprocedures/medicalimaging/ucm115357.htm

2.1 Ultrasound Acquisition

US acquisition relies on the combination of physical principles and signal processing methods, but it can be simply said that it is the result of the reflection of the sound waves in body structures [16].

It begins with an US transducer that contains piezoelectric elements, which vibrate in response to an applied electric current (piezoelectric effect) and produce sound waves that propagate through the body tissue. These waves propagate in the body, where they can be partly transmitted to deeper structures, reflected back to the transducer as echoes, scattered, or transformed into heat [17].

An US transducer not only generates sound waves but also receives them. The reflected portion, which goes back into the transducer when the wave hits a interface between different materials, is the responsible for the image formation. The intensity of the reflected portion is proportional to the difference between the acoustic impedance of the two mediums in the interface. The acoustic impedance is a property defined as the density of the medium times the velocity of wave propagation in that medium, and represents the resistance of a medium to the flow of acoustic waves. So the higher the frequency of the wave is (which is proportional to its velocity), the higher the impedance that each tissue will present [16]. On the other hand, the higher the frequency, the higher the resolution of the image produced by them [19].

Air-containing organs, such as the lungs, have very low impedance, while soft tissues have higher values, and bone has the highest level of impedance in the body [17]. Given that, US image showcases the acoustic impedance gradient in the body, for instance, interfaces with air and bone can induce total reflection, causing acoustic shadows behind them [20]. Irregular surfaces conduce to a larger portion of the waves to be scattered and this ultimately results in the image's speckle noise [17].

The sound waves emitted by the transducer are focused either by the shape of the transducer, a lens in front of the transducer, or a complex set of control pulses from the ultrasound scanner machine. This produces an arc-shaped sound wave from the face of the transducer [19]. In terms of visual access of the US, it is restricted to a relatively small window and its penetration power depends on the wave frequency used and can go from 3 to 25 cm. Lower frequency allows for higher penetration into the tissues, as they can endure more attenuation caused by the propagation through the tissues. Images of superficial structures, such as muscles and breast, can be acquired with higher frequencies, which is beneficial as they will consequently provide higher axial and lateral resolution [16].

Then the reflected waves return to the transducer and the received data is processed and used to construct a 2D representation of the slice of the body. The amplitude of the received sound signal and the time it took to travel through the body provide the necessary information for image formation[2]. The scan in performed in a sequence of directions and each received wave corresponds to a vertical line in the final image, as illustrated in Figure 2.1.

[2]https://www.fda.gov/radiation-emittingproducts/radiationemittingproductsandprocedures/medicalimaging/ucm1-15357.htm

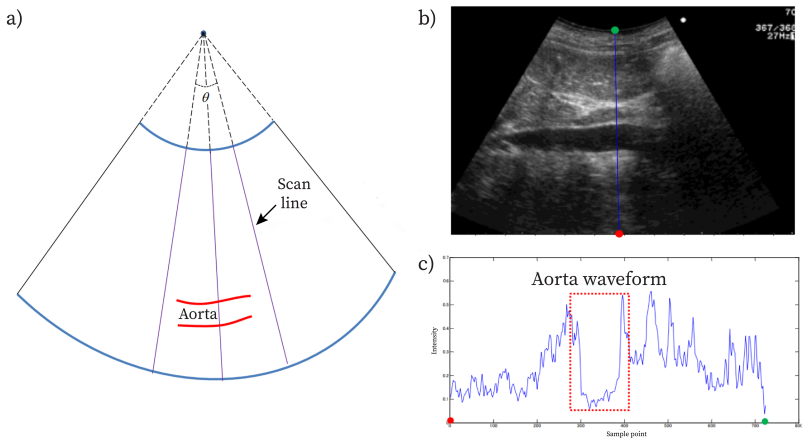


Figure 2.1: Example of the correspondence between the scan line and intensity signal in image. (a) Ultrasound scan lines. (b) A scan line in the B-mode US image. (c) Intensity signal along the scan line after conversion to B-mode. [21].

2.2 Ultrasound Modes

There are three modes in which the US acquisition can be displayed: A-mode, B-mode, and M-mode [17].

A-mode is the simplest way to present an US and it corresponds to a 1D scan along the axis of the transducer [22]. The “A” stands for “Amplitude”, as the A-mode consists on a signal which represents the variation of the amplitude of the wave received by the transducer over time. The time course can be perceived as the depth on a line of tissue where the reflection that formed that part of the signal took place [19]. It results from amplitude demodulation of the beam-formed radio-frequency (BRF) data, which is the raw data acquired by the transducer [23].

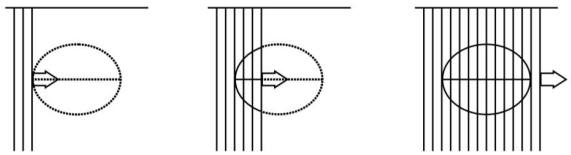


Figure 2.2: Formation of a B-mode image. The image is built up line by line as the beam is stepped along the transducer array [24].

The B-mode scan provides 2D information, presented as a grayscale image, and it is the most frequently used method in clinical practice. The “B” in B-mode stands for “Brightness”. The additional dimension, with regards to A-mode, is a consequence of the sequential coplanar pulses emitted by the transducer in this mode [16]. For this image construction, each line in the image corresponds to a pulse–echo sequence, as described in Figure 2.2. An example of a B-mode acquisition is shown in Figure 2.3.

The M-mode, or "Motion" mode, is a sequential set of A or B-scans, characterized by a very high sampling rate, which results in a high time resolution. It is used to examine phenomena

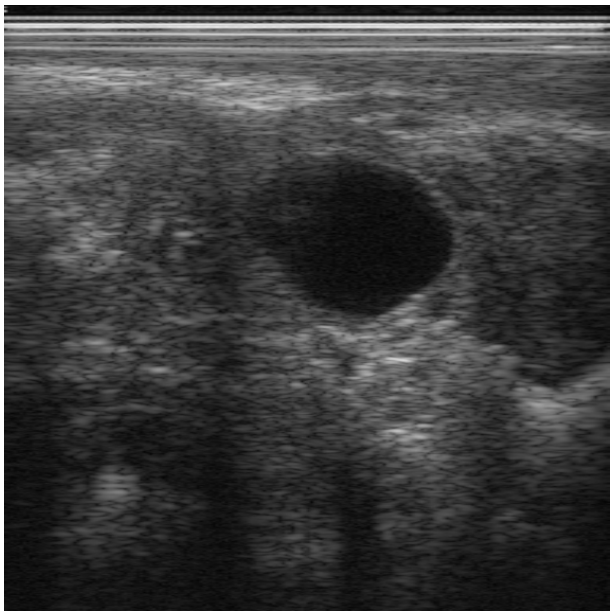


Figure 2.3: Linear B-mode transvaginal ultrasound shows a well defined hypoechoic mass.

associated with moving body parts, as it allows for rapid motion to be recorded, displayed, and measured[3]. Its main application is in cardiac and fetal cardiac imaging [4].

2.3 B-mode image formation

The sequence of signals acquired from the transducer is used to construct a 2D representation of the captured area.

Several processing steps are applied to the BRF signals in order to make the data visually comprehensible as an image. In Figure 2.4, it is illustrated an example of the results of the processing steps required for the conversion of BRF data to B-mode data.

In the first step, the original BRF signal is filtered with a low-pass filter using a Hamming window (Figures 2.4.e and 2.4.f). Then, the filtered signal is demodulated by envelope detection, and submitted to logarithmic compression to fit within the range of values that can be attributed to a image pixel (Figures 2.4.g to 2.4.j). The compression is performed logarithmically in order to enhance the contrast in the areas with lower ranges of intensity.

After the aforementioned steps, the result is a linear B-mode image (Figure 2.4.a), which is already visually perceptible. The last step is optional and consists in the rearrangement of the data based on the angle of the transducer used for the acquisition. It results in the "scan-converted" B-mode image (Figure 2.4.b), which is a more accurate representation of the spacial layout of the captured area.

[3] https://www.123sonography.com/book/351
[4] http://sonodriftzone.blogspot.com/2010/01/ultrasound-modes-b-m.html

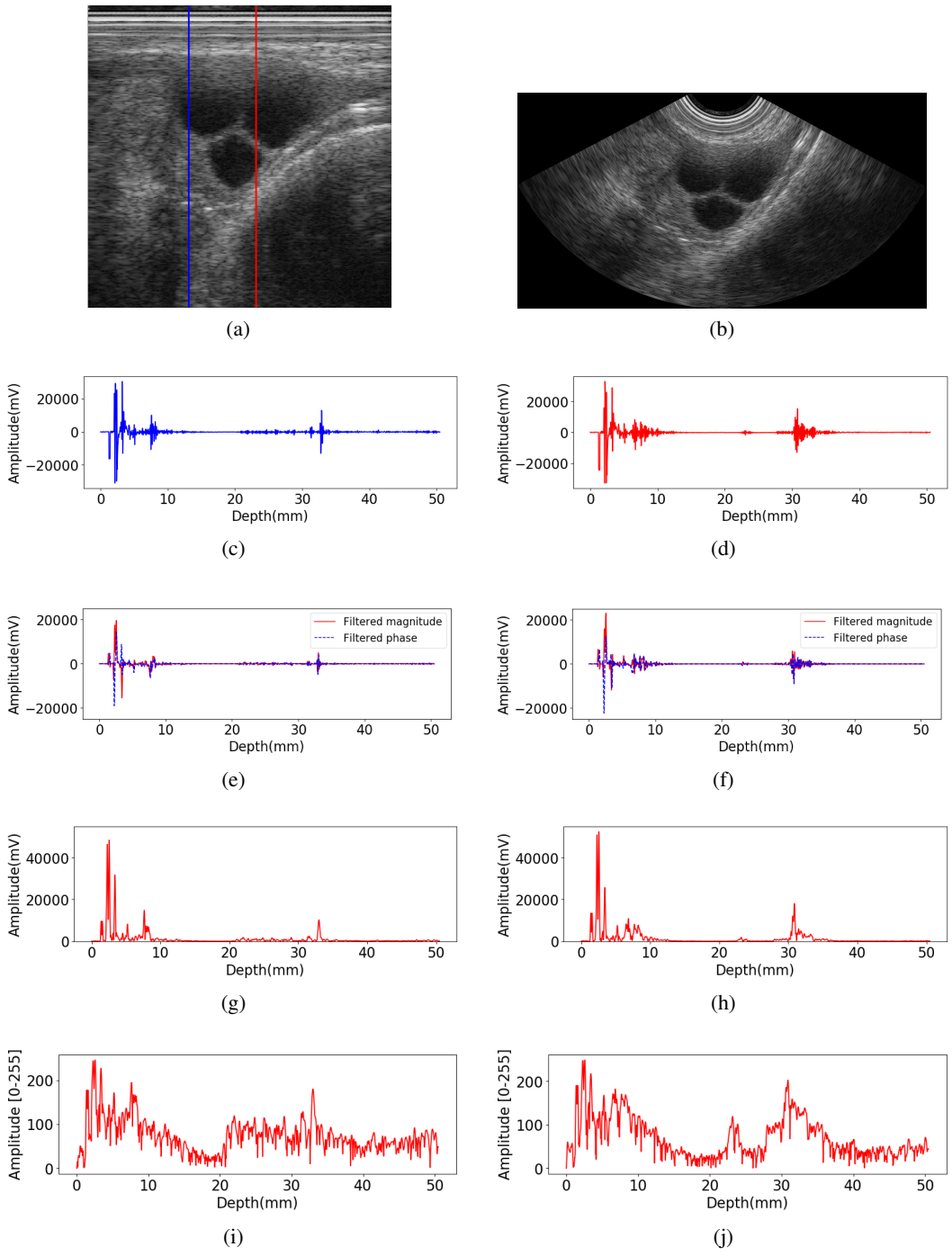


Figure 2.4: Example of the conversion of BRF to B-mode data. (a) Linear B-mode image with the two lines which correspond to the signal data shown next. The representations on the left (c, e, g and i) correspond to the blue line, on the right (d, f, h and j) correspond to the red line. (b) Scan-converted B-mode image. (c-d) Representation of the BRF signals. (e-f) Magnitude and phase of the filtered signals. (g-h) Filtered signals after envelope detection. (i-j) Final B-mode data, after logarithmic compression.

2.4 Evolution of Ultrasounds for Clinical Purposes

Research for the improvement of the quality of US imaging and production of sharper and clearer images focuses on the data acquisition level. Given that US is very cost-effective, widely available globally and does not require ionizing radiation, it has big potential as a diagnostic tool and recent advances in US technology are opening the door to real-time dynamic physiologic imaging. The fact that this imaging modality is so versatile in terms of application leads it to require research for a broad range of investigative methods [25].

Advances in the transducer design and electronic phase correction have allowed efficient focusing of US for transcranial treatment [26]. US with low intensity and high frequency (2 MHz) allows a non-invasive approach to create an optimal lesion for maximum efficacy without inducing adverse effects in otherwise inaccessible brain areas. This can potentially be used for the treatment of challenging neurological and psychiatric disorders.

Interesting discoveries in US biophysics with applications in cellular bioeffect investigations are being carried out by the Laboratory on Innovative Technology in Medical UltraSoundy (LITMUS)[5]. One particular phenomenon called Sonoporation, defined as cavitation-induced membrane perforation, has been investigated for its ability to change cell-cycle dynamics and disrupt cytoplasmic signaling pathways, which give it potential to enhance gene and drug delivery to cells [27, 28, 29].

In [30], it is proposed a different 2D array transducer design that uses a reduced number of elements on the 2D array. This reduction is compensated by the adjustment of the relative positions of the active elements in the probe and also by the application of a stochastic optimization algorithm for data processing. This lightens the acquisition process which is critical for enhancing the quality of real-time 3D US imaging.

Improvements have been made towards a more accurate estimation of low blood velocities, like the ones found near the wall in recirculation or disturbed flow regions. These estimations are critical for the evaluation of conditions such as atherosclerosis and risk of stroke. For this purpose, in [31], it is proposed a transducer design with 5 receiver beams in a asymmetric configuration.

For the acquisition of US images with higher signal-to-noise ratio, it is proposed, in [32], a frequency compounding method that uses a high-frequency dual element transducer. Frequency compounding can be used to reduce the speckle noise by averaging two or more images acquired with different US frequencies.

Also, various signal processing techniques such as adaptive beamforming [33], and eigencomponent analysis [34] have been proposed to improve contrast and resolution in imaging, namely to study dynamic events like blood flow and tissue motion[6].

[5] https://lit-mus.org/
[6] https://lit-mus.org/research

2.5 Concluding Remarks

US imaging is one of the leading medical imaging techniques worldwide, as it is inexpensive, non-invasive and based on non-ionizing radiation. Transvaginal US images are frequently used to monitor ovarian parameters in patients undergoing ART treatments.

A US transducer generates sound waves and then receives the reflected portion of those waves. The intensity of the reflected echo is proportional to the mismatch in acoustic impedance between two mediums. This information is then used for image formation.

There are three modes for US display. A-mode is a 1D scan along the axis of the transducer. B-mode consists of a grayscale image, in which each line corresponds to a beam-formed radio-frequency signal demodulated and compressed. M-mode is used to obtain information on phenomena associated with moving body parts.

Chapter 3

Neural Networks and Deep Learning

Deep Learning (DL) is a class of Machine Learning (ML) techniques that comes even closer to the reasoning of the human brain. DL can mimic human intellect and solve problems that for us require intensive training and intuition. ML approaches allow a computational model to learn from data or knowledge, but are limited to process natural data in its raw form and it is dependent on the choice of data representation/features. On the other hand, DL approaches enable learning from experience and adjustments to the results it needs to achieve [35].

DL adds a key component: feature learning, which can be supervised or unsupervised [36]. This allows the system to automatically extract and organize the discriminative information that is needed for effective detection or classification from the raw data.

ML approaches are on the rise and are applied to numerous fields, like drug-design, speech recognition and web search/social network algorithms. In the context of this work, the field of computer vision is the most important one. The investigation in this field has been focusing more and more on DL specifically.

Using conventional ML techniques for pattern recognition would require a lot of feature engineering and specific expertise to design a feature extractor to perform the transformation of raw data into an appropriate feature vector that allowed for the system to do a proper detection and classification [37]. DL's ability for pattern recognition tasks goes beyond the human capacity and that's probably why it has become so popular to solve complex problems. The DL architecture is a hierarchical multilayer stack of simple modules that are subject to learn from data and compute input–output mappings [37].

3.1 Basic Neural Networks

The basis of the most sophisticated NNs are two simple algorithms: Logistic Regression and Perceptrons [38]. Such models were the first to be built that were able to learn from "experience", and were both developed around the 1950's.

The perceptron is fundamentally a linear binary classifier that combines a set of weights ($\mathbf{w}$) with a feature vector (inputs, defined by $\mathbf{x}$). Given a set of features, it computes their weighted

sum to which is added a bias value. The result of this computation is thresholded and finally it returns the corresponding binary output.

$$\begin{aligned} y = 1 \quad if \quad \sum_{i=0}^{\mathrm{N}} w_i x_i \geq t \\ y = 0 \quad if \quad \sum_{i=0}^{\mathrm{N}} w_i x_i < t \end{aligned} \tag{3.1}$$

The mathematical formulation of the perceptron is described in Equation 3.1, where y is the output of the perceptron, N is the number of input values, x and w are the values of the inputs and their respective weights and t is the selected threshold[1]. The drawback of this algorithm is that it requires the two classes to be linearly separable for it to be able to converge. This makes perceptrons appropriate to solve only a small set of problems.

The single-perceptron learning algorithm is quite simple and requires only the pre-definition of one parameter, the learning rate (η). The update of the weights is given by:

$$w_i(k+1) = w_i(k) + \eta \max(0, -y(k)\, d) x_i \tag{3.2}$$

where k is the number of the iteration and d is the desired output. Equation 3.2 is computed until the entire training set is classified correctly. Note that the same formula is used to determine the update of the threshold by simply replacing w with t[2].

Logistic Regression works in a very similar manner in terms of data flow and structure. It corresponds to a smoothed version of the perceptron due to the nonlinearity introduced by the sigmoid function (instead of the step function of the perceptron) [39]. Its ouput it computed by the following expression:

$$y = \sigma\left(\Theta^{\mathrm{T}}\mathbf{x}\right) = \frac{1}{1 + e^{-\Theta^{\mathrm{T}}\mathbf{x}}} \tag{3.3}$$

where σ is the sigmoid function, Θ^{T} is the transposed vector of the parameters (weights and bias) and $\mathbf{x}$ is the vector of the inputs. In the Equation 3.3, the bias corresponds to the first element of the vector Θ^{T}, and it is multiplied by the first element of the vector $\mathbf{x}$, whose value is 1.

The learning algorithm is based on gradient descent and backpropagation, which will be discussed later in this chapter. The cost function that is applied to the gradient descent in the case of Logistic Regression is the following:

[1] https://towardsdatascience.com/perceptron-the-artificial-neuron-4d8c70d5cc8d
[2] http://130.243.105.49/ lilien/ml/seminars/2007_02_01b-Janecek-Perceptron.pdf

$$c(\Theta) = -\frac{1}{M}\sum_{j=0}^{M}[d^{(j)}log(y^{(j)}) + (1-d^{(j)})log(1-y^{(j)})] \tag{3.4}$$

where M is the number of training examples[3].

From these basic units, multilayered structures can be formed. With the liberty of choosing the implemented loss function, activation function and the structure (number of neurons, layers, connections between neurons, etc.), a variety of networks can be established [38].

3.2 Artificial Neural Networks

Artificial Neural Networks (ANN) or just Neural Networks (NN) are DL frameworks with a directed graph topology, inspired in the network's structure of the human brains and the behaviour of neurons transmitting electrical signals [1].

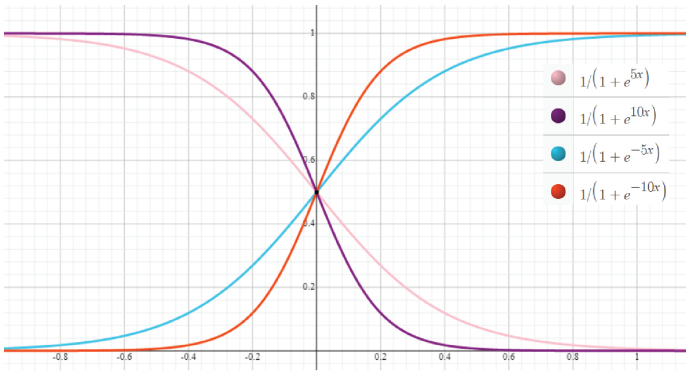


Figure 3.1: Comparison of sigmoid functions with diferent parameter's values to illustrate the influence of the weights in the model of the neuron.



Figure 3.2: Comparison of sigmoid functions with diferent parameter's values to illustrate the influence of the bias in the model of the neuron.

The basic unit of a NN is the artificial neuron that has inputs just like a real neurons have dendrites and an output analogous to the axon. These inputs and outputs can be seen as links between

[3] https://hackernoon.com/introduction-to-machine-learning-algorithms-logistic-regression-cbdd82d81a36

neurons, since, in the context of multilayered structures, one neuron's outputs will become another neuron's input, resulting in a complex causal chain or net. The neuron is like a processor that has its own function to transform input values into an output value. This function is composed by a set of connection weight values, that are multiplied by their respective input. To this weighted sum is added a bias value[4]. This result is passed on to an activation function which usually corresponds to a sigmoid function in binary classification problems[5]. In the sigmoid function, the connection weights define the "steepness" (Figure 3.1) and the bias (positive or negative) define the shift to the left or to the right (Figure 3.2). The result of these operations corresponds to the output of the neuron. Without the activation function, the neuron would simply perform linear regression and the NN would have a lower ability to learn complex functional mappings from data[3.2].

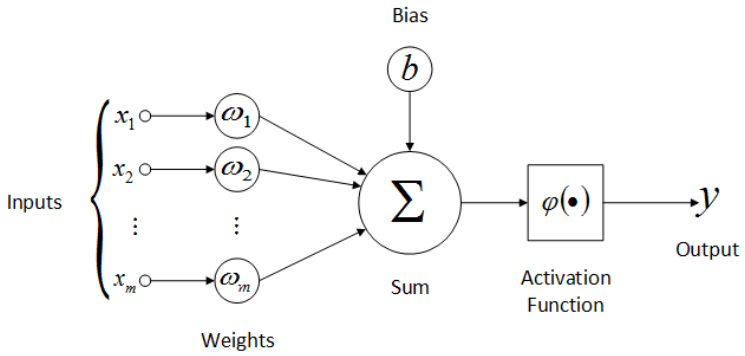


Figure 3.3: Basic model of the neuron [40].

The mathematical model of a neuron from a generic NN can be found in Equation 3.5, which is illustrated in Figure 3.3.

$$y = \varphi\left(\sum_{i=1}^{m}(x_i w_i) + b\right) \tag{3.5}$$

A NN is a sequence of layers in which the neurons are grouped in. The first layer is the input layer, whose input is the raw data. Then there are the hidden layers, where the neurons of one layer receive the output of all the neurons of the previous layer, forming a fully connected NN. Finally, the output layer receives the outputs that resulted from the last hidden layer and returns the final result, concluding the flow of data. This process describes a simple feedforward NN but there are many variations to this.

A illustrative representation of a generic NN is shown in Figure 3.4.

[4]https://developer.nvidia.com/discover/artificial-neural-network

[5]http://dataaspirant.com/2017/03/07/difference-between-softmax-function-and-sigmoid-function/

[6]Adapted from https://stackoverflow.com/questions/28288489/neural-networks-does-the-input-layer-consist-of-neurons

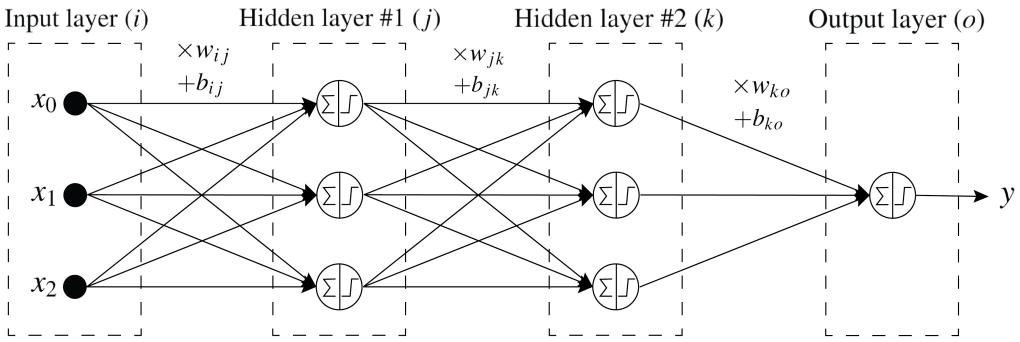


Figure 3.4: Basic architecture of a NN[6].

3.2.1 Convolutional Neural Networks

Among the vast variety of unsupervised and supervised feature learning algorithms designed computer vision problems [41], Convolutional NNs (or CNNs) have gained the most attention due to their good performance [42]. The main idea behind CNNs is to create a deep hierarchical feature detection system that is able to represent and detect patterns in the input data [37]. The deeper the network goes, that is, the more hidden layers for feature learning it is composed by, the more sophisticated is its pattern detection ability. However, a configuration with too many hidden layers can be detrimental, as it will cause the NN to be more computationally expensive, to require a lot of training data, and may lead the model to overfit to the training set [43].

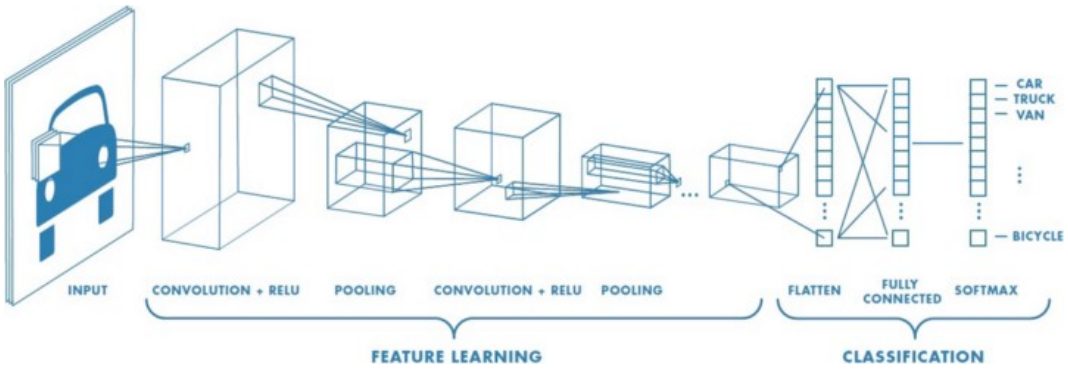


Figure 3.5: Neural network with a sequence of convolutional layers[7].

What sets CNNs apart from regular NNs is that their hidden layers are convolutional layers and max-pooling layers, as illustrated in Figure 3.5. Convolutional layers do not consist of a set of neurons in which regular weighted sum operations are performed, but instead they are set convolution operations and Rectified Linear Units (ReLU) [37].

To understand the formulation of a CNN it is key to understand the basic concepts of the convolution operation (including its kernel, padding and stride), ReLU and max-pooling.

[7] https://medium.com/@RaghavPrabhu/understanding-of-convolutional-neural-network-cnn-deep-learning-99760835f148

Convolution is a mathematical operation that preserves the spatial relation between data points. The most important in this context is the 2D convolution, in which data points correspond to pixels. The mathematical formulation of 2D convolution operation is the following:

$$H(x,y) = F(x,y) * G(x,y) = \sum_{i=0}^{K-1} \sum_{i=0}^{K-1} F(i,j)G(x-i,y-j) \tag{3.6}$$

where $H(x,y)$ is the obtained image, $F(x,y)$ is the original image, $G(x,y)$ is the applied filter or kernel, and K corresponds to the width and height of the filter (whose size is $K \times K$).

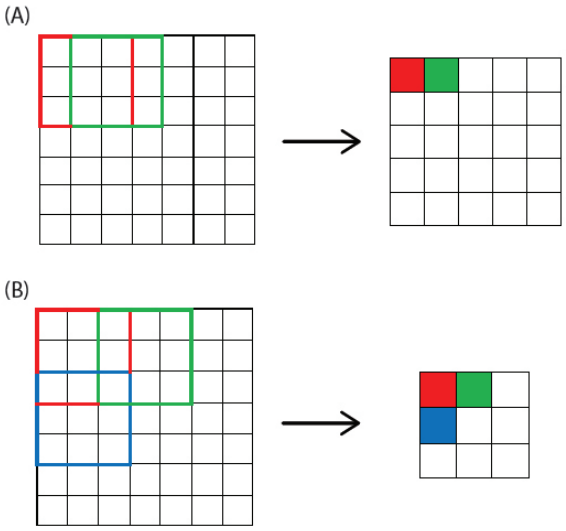


Figure 3.6: Illustration of the input (on the left) and output (on the right) size of a 2D convolution with a kernel of size 3×3 and stride 1 (A) and stride 2 (B)[8].

As consequence of the convolution formulation, if K is higher than 1, the output of the operation will be smaller than the input, as the borders of the input will have no correspondence in the output. The amount of pixels lost not only depends on the size of the kernel, but also on the stride of the operation. Stride corresponds to the number of pixels that the kernel shifts by in the input between the computation of consecutive output pixels. The demonstration of stride variation is found in Figure 3.6. The output size will be equal to $((N-K)/S)+1$, where N is the width and length of the input image and S is the stride.

To avoid size reduction after the convolution, the borders of the input can be padded before the convolution. Padding, in the context of image processing, means generating new border pixels, whose values can be extrapolated in several ways. The input size is increased as much as necessary so that the output of the convolution is returned with the original size of the input. The expression to determine the output size is $((N+2P-K)/S)+1$, where P is the padding applied (number of pixels added around the border). Some of the simplest and most common padding strategies are

[8]https://adeshpande3.github.io/A-Beginner%27s-Guide-To-Understanding-Convolutional-Neural-Networks-Part-2/
[9]https://medium.com/machine-learning-bites/deeplearning-series-convolutional-neural-networks-a9c2f2ee1524

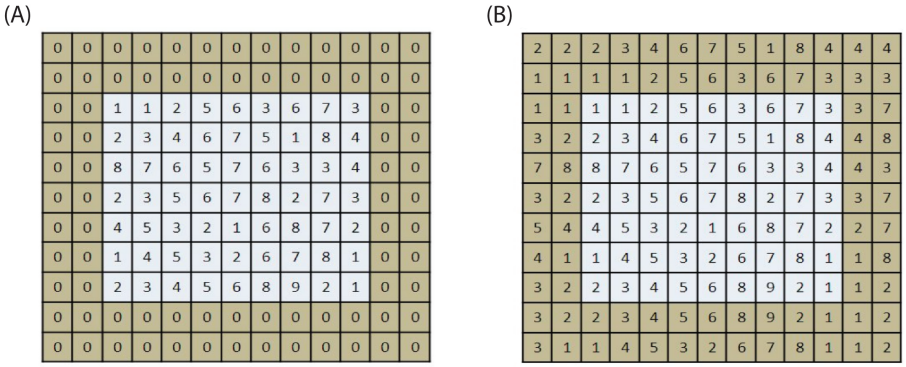


Figure 3.7: Examples of the application of zero-padding (A) and reflection (B). Padding size is 2 in both examples[9].

zero-padding (new pixels are set to zero) and reflection (border pixels of the original image are repeated in the added border pixels)[10], which are demonstrated in Figure 3.7.

The Rectified Linear Unit, or ReLU, is the activation function typically used in convolutional layers, written as $f(x) = max(0,x)$. The use of ReLU instead of activation functions with saturating nonlinearities like sigmoid, helps accelerate the convergence of stochastic gradient descent [43]. The drawback of its implementation is that it can get "stuck" in the negative part of the function, where the slope is 0. When this happens it is unlikely that the optimization algorithm adjusts the weights involved in the computation that led to a negative value, in order to bring it back to the positive range of outputs. This way that specific computation no longer plays a meaningful role in the NN, since the result it produces won't change according to its input and be propagated so that is influences the final result of the CNN[11].

Pooling layers perform downsampling through one of a few possible operations, namely max-pooling, subsampling, average-pooling and attentive-pooling. The purpose of pooling layers is to reduce the spatial size of the data that goes through the hidden layers and, consequently, reduce the amount of parameters in the NN, which helps to avoid overfitting[12]. The max-pooling operation is the most widely used and the one with better performance in convolutional architectures [44, 45]. This is a simple operation, described as a pooling window of arbitrary size that slides over the input image with a certain stride, and selects the maximum value from each patch to form the corresponding output image.

In convolutional layers, instead of weights and biases, the parameters that are optimized during training are the weights in the filters used for the convolution operation[13]. These filters can be perceived as the "pattern detectors" of the CNN.

A filter bank is assigned to each convolutional layer. It is part of the configuration of the architecture of the CNN to set the number of filters and their spatial dimensions (e.g. $5\times5\times5$) for

[10] http://cse19-iiith.vlabs.ac.in/theory.php?exp=neigh
[11] https://medium.com/tinymind/a-practical-guide-to-relu-b83ca804f1f7
[12] http://cs231n.github.io/convolutional-networks/
[13] https://www.intechopen.com/books/green-electronics/optimizing-of-convolutional-neural-network-accelerator

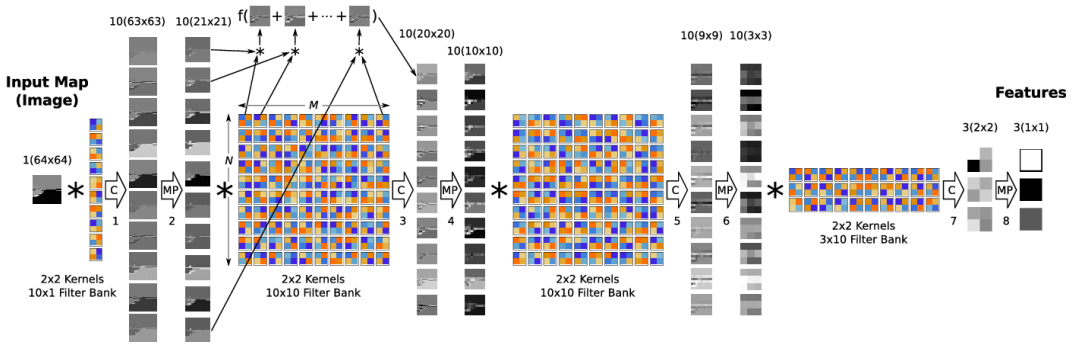


Figure 3.8: Convolutional Neural Network with 8 layers alternating between convolution (C) and downsampling using max-pooling (MP) [46].

each filter bank. Each filter of the bank convolves with all of the input channels [47]. An example of a CNN applied to a image (2D), highlighting the filter banks used, is illustrated in Figure 3.8.

3.2.2 Learning and Neural Network Optimization

The training of a NN depends on its structure and its algorithm, but one clear distinction is that the learning can be supervised or unsupervised, that is, trained using labeled or unlabeled data.

The general learning rules, in other words, the algorithms for updating the parameters of a network, are designed in order to get better predictions, minimize an error or energy function, both in supervised and unsupervised learning. The NN is initialised with a random set of parameters which will return very inaccurate outputs, and the training allows it to get progressively better results.

An error or cost function is associated with every ML algorithm, and it is a way for the NN to quantify how different its results are from what they should be. There are several options for the error function, from mean squared error to entropy, which will be detailed in the Subsection 3.3.

NN frameworks such as Autoencoders and Self-organizing Maps apply unsupervised learning[14], and so do many clustering and association algorithms, as they return the output only by searching for hidden patterns and structure in the dataset. With these techniques it is possible to group data or to discover relationships between data points based on its similarity. Although this is less complex and quicker to run, and it is easier to find data without the requirement of labels, these methods are usually less accurate than the ones that use prior knowledge of the desired result.

In the context of this work, the focus should be supervised learning[14]. This type of learning has the goal of minimizing a loss/cost function that uses labeled data to be computed. There are a few different ways to perform supervised learning, but the concept of backpropagation is particularly important as it is the standard algorithm for NN optimization. It is a procedure that computes the gradient of the loss function according to the weights and biases of a multilayer stack of modules (neurons), which is fundamentally a chain rule for derivatives. In other words: it quantifies the rate

[14]http://intellspot.com/unsupervised-vs-supervised-learning

of change that the loss function experiences due to the variation of the parameters. It is important to state that it requires the node's transfer functions to be continuous.

Along with backpropagation, the learning algorithm must have an optimization algorithm to update the parameters based on the gradient of the objective function[15].

In the context of NNs, one of the most commonly used optimization algorithms is the Stochastic Gradient Descent (SGD)[16]. The backpropagation determines the actual function (Loss vs. Parameters) and Stochastic Gradient Descent allows the determination the parameters that minimize it, which are then used to update of the NN. The gradient descent consists on the repetition, until convergence, of the following parameter update rules:

$$\mathbf{w}^l := \mathbf{w}^l - \frac{\eta}{\mathrm{N}} \sum \delta^{\mathbf{x},l} (\mathbf{a}^{\mathbf{x},l-1})^T \tag{3.7}$$

$$\mathbf{b}^l := \mathbf{b}^l - \frac{\eta}{\mathrm{N}} \sum \delta^{\mathbf{x},l} \tag{3.8}$$

where η is the learning rate, N is the number of training examples, l denotes the number of the layer in the network that the backpropagation is reaching, $\mathbf{a}^{\mathbf{x},l-1}$ is the vector of activations of the previous layer when $\mathbf{x}$ is the input, and $\delta^{\mathbf{x},l}$ is the derivative of the error (backpropagated to that layer). Equation 3.7 computes the update of the weights values ($\mathbf{w}$) and Equation 3.8 of the biases ($\mathbf{b}$)[3.2.2].

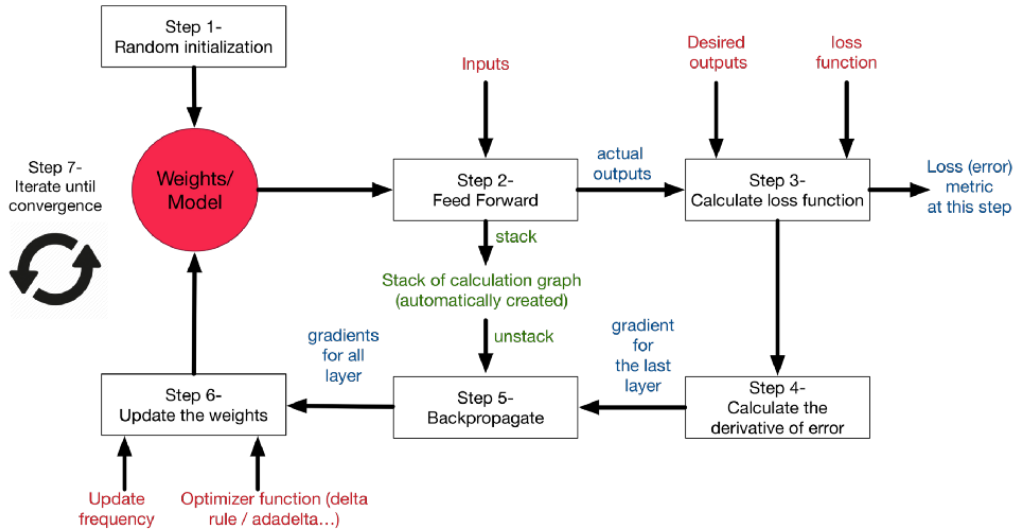


Figure 3.9: Illustration of the sequence of training steps for neural network optimization[17].

The learning rate determines how fast the algorithm converges. Basically, its purpose is to control how far the change in the parameters value can go in each iteration.

[15] https://towardsdatascience.com/types-of-optimization-algorithms-used-in-neural-networks-and-ways-to-optimize-gradient-95ae5d39529f

[16] http://neuralnetworksanddeeplearning.com/chap2.html

[17] https://medium.com/datathings/neural-networks-and-backpropagation-explained-in-a-simple-way-f540a3611f5e

During the training process, the loss function will fluctuate, which may helps the model discover new and possibly better local minima. On the other hand, it makes it difficult for the updates to convergence to the exact minimum due to overshooting [48].

To sum-up, the general learning algorithm of a NN (Figure 3.9), for each training pair, repeats the following steps until convergence [48]:

- Feed-forward - Feed input and return output;
- Computation of the error;
- Backpropagation of the error to hidden layers;
- Parameter update.

Adaptive learning rate algorithms or adaptive gradient descent optimization algorithms are an optimization of gradient descent methods that update the learning rate in each step[18]. One example of these algorithms is Adam Optimizer [49] which adapts the step size individually for each weight of the NN and employs the concept of momentum, that is, it uses the magnitude of the gradients of previous steps (not only the gradient of each individual step) to update the parameters. Adaptative learning rate algorithms have a better performance for very noisy and/or sparse gradients.

3.3 Performance Evaluation Measures for Segmentation

Quantitative measures to evaluate the performance are crucial for segmentation problems. These criteria are not only used to assess the overall result of an algorithm implementation, in order to validate a certain algorithm or to make well founded comparisons between algorithms, but also, particularly in the field of DL, to be used for model optimization.

For the evaluation of segmentation results as it will be performed in this work (through supervised learning), the most relevant criteria are based on discrepancy evaluation methods, which compute the error by comparing the predicted segmentation against a manually-segmented reference image (also know as ground-truth) [50].

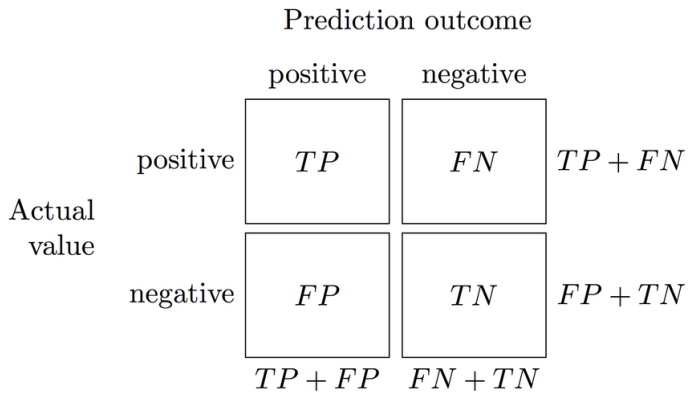


Figure 3.10: Confusion Matrix for a two–class problem [51].

[18] https://wiki.tum.de/display/lfdv/Adaptive+Learning+Rate+Method

The most common evaluation measures for classification problems are obtained through the manipulation of the four values found in the Confusion Matrix. The Confusion Matrix is in itself a performance measure, which can be determined for problems with two or more classes. In the case of two classes - foreground and background - where foreground is considered positive and background is negative, the Confusion Matrix is as shown in Figure 3.10. In this matrix, true positives (TP) and false positives (FP) correspond to the number of correct and incorrect predictions of foreground, respectively, and true negatives (TN) and false negatives (FN) correspond to the number of correct and incorrect predictions of foreground, respectively[19]. The description of the measures obtained from the number of TP/FP/TN/FN is presented in Table 3.1.

Table 3.1: Description of some of the rates derived from TP/FP/TN/FN values[20].

Rate	**Definition**	**Technical significance**
Accuracy	(TN+TP)/(TN+FP+FN+TP)	Probability of a correct prediction
Misclassification Rate	(FN+FP)/(TN+FP+FN+TP)	Probability of a incorrect prediction
Precision	TP/(FP+TP)	Probability of a positive prediction to be correct
False Positive Rate	FP/(FP+TN)	Probability of a positive prediction to be incorrect
Sensitivity/ Recall/ True Positive Rate	TP/(TP+FN)	Probability of the positive class to be correctly predicted
Specificity/ True Negative Rate	TN/(TN+FP)	Probability of the negative class to be correctly predicted
False Negative Rate	FN/(FN+TP)	Probability of a negative prediction to be incorrect

A way of representing the classifier's performance is through the AUC (Area Under The Curve) ROC (Receiver Operating Characteristics) Curve, determined by matching pairs of sensitivity and specificity values in a plot. As shown in Figure 3.11, the ROC Curve illustrates the trade-off between sensitivity and specificity, and the AUC-score is a measure of how good this trade-off is.

Other relevant measures that may be used for segmentation performance evaluation are the Mean-Squared Error, the Dice Similarity Coefficient and the Binary Cross-Entropy.

The Mean-Squared Error (MSE) is one of the most commonly used performance measures, useful for both classification and regression problems. It corresponds to the average squared difference between the predicted values and its ground-truth, and so the lower the MSE is, the better the classifier is. The MSE formulation is in Equation 3.9, where N is the number of predictions, f_i is the prediction of the classifier and y_i is its actual value (ground-truth)[22].

[19] https://towardsdatascience.com/understanding-confusion-matrix-a9ad42dcfd62
[20] https://www.dataschool.io/simple-guide-to-confusion-matrix-terminology/
[21] https://classeval.wordpress.com/introduction/introduction-to-the-roc-receiver-operating-characteristics-plot/
[22] https://medium.freecodecamp.org/machine-learning-mean-squared-error-regression-line-c7dde9a26b93

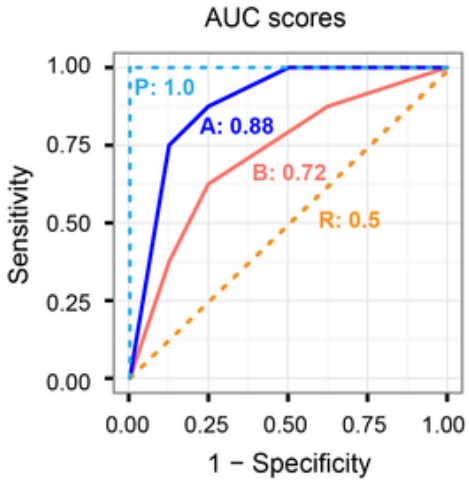


Figure 3.11: Example of four AUC-ROC Curves and the respective AUC-scores. The perfect classifier (P) corresponds to 1.0 AUC-score and a random classifier corresponds to 0.5 AUC-score. The classifier A has better performance that classifier B, since the score is higher and the line of the plot of A is always above the plot of B[21].

$$\text{MSE} = \frac{1}{\text{N}} \sum_{i=1}^{\text{N}} (f_i - y_i)^2 \tag{3.9}$$

The Dice Similarity Coefficient (DSC) is a metric of spatial overlap between two segmentations, defined as $\text{DSC}_{(\text{A,B})} = 2(\text{A} \cap \text{B})/(\text{A} + \text{B})$, where A and B are the two compared segmentations. DSC values range between 0 and 1, where 1 means total overlap and 0 means no overlap [52]. Another way to put its expression, in terms of TP/FP/TN/FN values is:

$$\text{DSC} = \frac{2\text{TP}}{2\text{TP} + \text{FP} + \text{FN}} \tag{3.10}$$

Binary Cross-Entropy (BCE) quantifies the difference between the predicted probability distribution and the actual probability distribution (where each point of the distributions corresponds to one prediction)[23].

$$\text{BCE} = \text{mean}(\{l_1, \ldots, l_\text{N}\}^\top) \tag{3.11}$$

$$l_n = -\left[y_n \cdot \log f_n + (1 - y_n) \cdot \log(1 - f_n)\right] \tag{3.12}$$

where N is the number of predictions, f_n is the prediction of the classifier and y_n is its actual value (ground-truth)[24]. Both prediction and ground-truth values must be numbers between 0 and

[23] https://medium.com/@vijendra1125/understanding-entropy-cross-entropy-and-softmax-3b79d9b23c8a
[24] https://pytorch.org/docs/stable/_modules/torch/nn/modules/loss.html

1 for BCE to be applied this way.

3.4 Convolutional Neural Networks for the Segmentation of Biomedical Images

U-Nets [53] and V-Nets [54] are CNNs specifically designed for the segmentation of biomedical images in 2D and 3D, respectively. Their formulations take into account two big limitations faced in the biomedical field for image segmentation: the training sets available have a limited number of annotated images and there are overlapping objects of the same class in the images. These NN are also adapted to the requirement of pixel-by-pixel classification, which is the case for segmentation tasks.

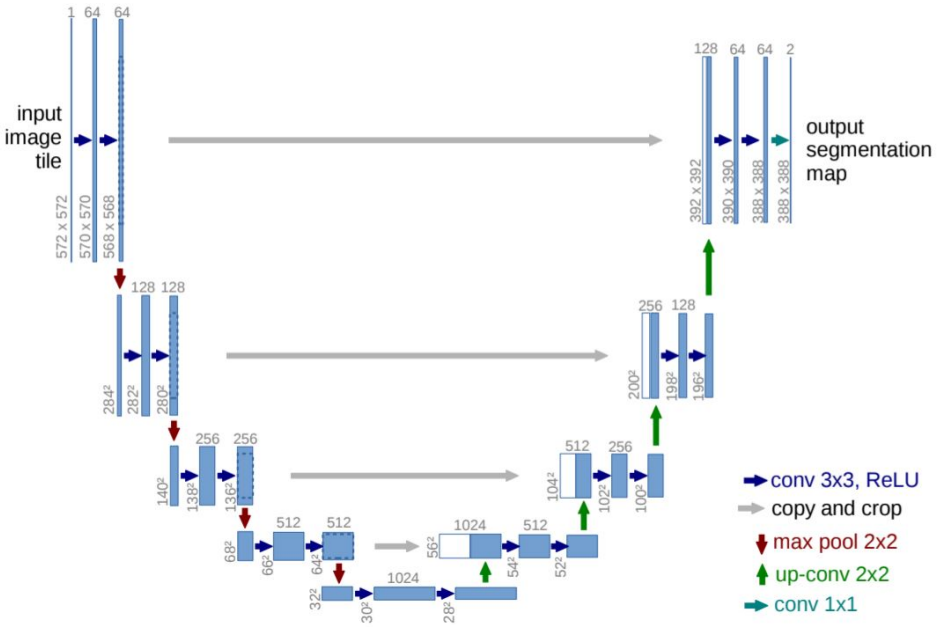


Figure 3.12: U-Net architecture. Each blue box corresponds to a multi-channel feature map. The number of channels is denoted on top of the box. The x-y-size is provided at the lower left edge of the box. White boxes represent copied feature maps [53].

The structure of the U-Net, as illustrated in Figure 3.12, has a U shape.

It is composed by an contracting path, on the left side, where the input goes in and its goal is to capture the context of the image. The spatial information is reduced and feature information is increased due to a sequence of convolutions, rectified linear units and max-pooling layers for downsampling. With each downsampling step the number of features is doubled.

Then an expanding layer, on the right, picks up the data from its lowest resolution and results in its output. In this path, it is applied a series of convolutions and rectified linear unit combined with “up-convolutions”, each of them reducing the number of feature channels by half.

The result of each up-convolution is concatenated with the spatial information of the correspondingly cropped high-resolution feature map from the contracting path.

The final result has two channels, one for the foreground and one for the background class. This output will be smaller than the input images, which can be avoided by extrapolating the raw data of the borders by mirroring.

To perform the training of these type of NN, the authors suggest a sliding-window setup to provide a local region (patch) around each pixel for which the class label will be predicted. It uses SGD with backpropagation to optimized the parameters of the NN.

The recommended energy function is the cross-entropy loss function transformed by the softmax function. An additional step that is proposed for this task is to pre-compute a weight map for each ground-truth segmentation, to compensate the different frequency of pixels from a certain class in the training. This combined with data augmentation using shifts, rotations and smooth deformations with random displacements sampled from a Gaussian distribution (with 10 pixels standard deviation), will lead to more robust and accurate results. Also during the training process, the pixels in the weight map that correspond to border separation of touching objects will be assigned a very high weight. This way the network is very penalized for errors in these type of pixels, and so it will learn to ensure the correct segmentation between touching objects of the same class.

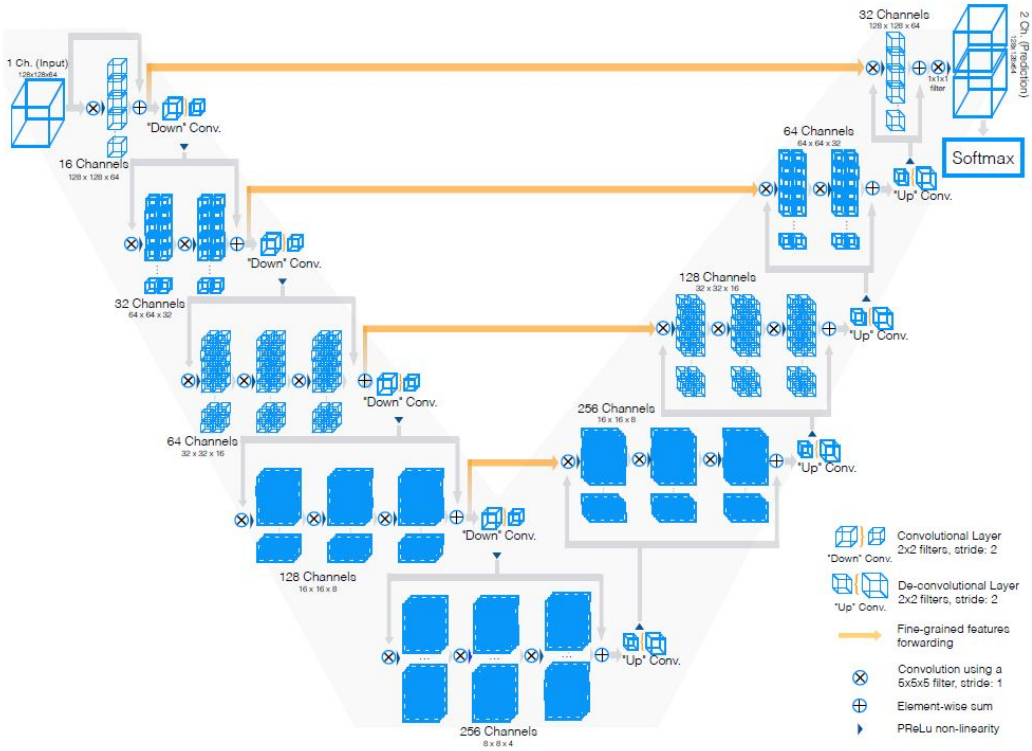


Figure 3.13: Schematic representation of the V-net architecture [54].

V-Nets have a very similar structure (Figure 3.13), but are designed to handle 3D input data.

In this case, the convolutions are performed with 3D kernels and padding is done so that the two channels of the output of the NN have the same dimensions as the input. This network's predictions (channels of foreground and background) are then processed through a softmax layer which outputs the probability values of each class.

In this case, to overcome the fact that most of the voxels belong to the background, which usually causes the learning process to get trapped in local minima of the loss function, and consequently makes the predictions biased towards background, instead of using the weighted map like the U-Net, it is proposed a modified objective function based on the dice coefficient, as presented:

$$D = \frac{2\sum_i^N p_i g_i}{\sum_i^N p_i^2 + \sum_i^N g_i^2} \tag{3.13}$$

where p and g are the real and obtained values, whose quantified comparison is obtain by the coefficient and N is the number of pairs to be compared. This index ranges from 0 to 1 and should be maximised.

A particular characteristic of V-Net's formulation is the use of a so called "residual function": as described in Figure 3.13, the input of each block is added to the output of the last convolutional layer of that block. It was concluded that, when the "residual function" is applied, the network reaches convergence in a fraction of the time.

Another improvement done for the V-Net in comparison to the U-Net is related to data-augmentation. Besides shifts, rotations and smooth deformations with random displacements, it also employs variations in the intensity distribution of the data by performing histogram matching of the intensity distributions of the image that will be used for training in each iteration with another random image of the dataset.

3.5 Concluding Remarks

This chapter addressed the functionality of NNs, from its basic elements to the overall structure, flow of data and optimization process. CNNs are the most relevant type of NN for computer vision problems, for they ability to identify spatial patterns. The models of two NN's specifically designed for the segmentation of biomedical images are also presented.

NNs are a very powerful tool and, although their structure is based on the connectivity system within the human brain, they are able to solve complex problems that go beyond the human brain capacity.

Chapter 4

Segmentation in ultrasound images. Its application for the segmentation of follicles in ultrasound images

The automatic or semi-automatic segmentation task for US images is particularly challenging due to the significant amount of noise, weak edges, tissue-related textures and artifacts found in this imaging modality [55].

Several approaches can be employed for this purpose, which go from more traditional image processing based approaches, which rely on rule-based approaches and specific assumptions, to deep learning approaches. Traditional methods include clustering, region-growing and graph-based algorithms, which are specially sensitive to the noise and contrast in an image, may rely on its initialization and require the user to select a criterion. On the other hand, deep learning approaches are the most promising as they do not require such strong assumptions, and have shown a superior performance in object segmentation and detection tasks [56]. This chapter presents an overview of the existing methods for US image segmentation of ovarian follicles, and other structures whose ultrasonic representation is similar to that of ovarian follicles (during this document referred to as general hypoechoic anatomical structures).

4.1 State-of-the-art

Regarding the methodology, the available solutions for image segmentation for this image modality are grouped in methods applied to general hypoechoic anatomical structures, in order to cover a broader range of relevant methods, and then methods specifically for the segmentation of follicles. Both subsections include traditional image processing methods and deep learning methods.

4.1.1 Segmentation of general hypoechoic anatomical structures

In this section are presented automatic and semi-automatic methods based on energy minimization, region growing, level-set and deep learning algorithms, applied to the segmentation anatomical structures such as breast lesions, left ventricle and prostate.

The method in [57] consists on a four step algorithm for a semi-automatic detection of breast lesions. The first one is a preprocessing step, which entails the cropping of the image by a certain number of pixels on the borders (empirically determined) to remove the subcutaneous fat (specifically necessary for breast images, not applicable to ovary US) and median filtering to minimize speckle noise. Next, the image gray-scale values are inverted (lesion changes from dark to light) and the result is multiplied by a Gaussian constraint function whose level surfaces are ellipses, centered on the lesion center. The width and depth directions of this function are varied adaptively and automatically for each image. The constraint function has higher values in the center and gradually lower values as the distance from the center increases. To correctly place the Gaussian, the center is manually defined. The third step consists on a sequential thresholding to obtain partitions that correspond to potential lesion margins. Finally, the chosen partition is the one that maximizes the utility function, which is the Average Radial Derivative (returns the average directional derivative in the radial direction along the margin).

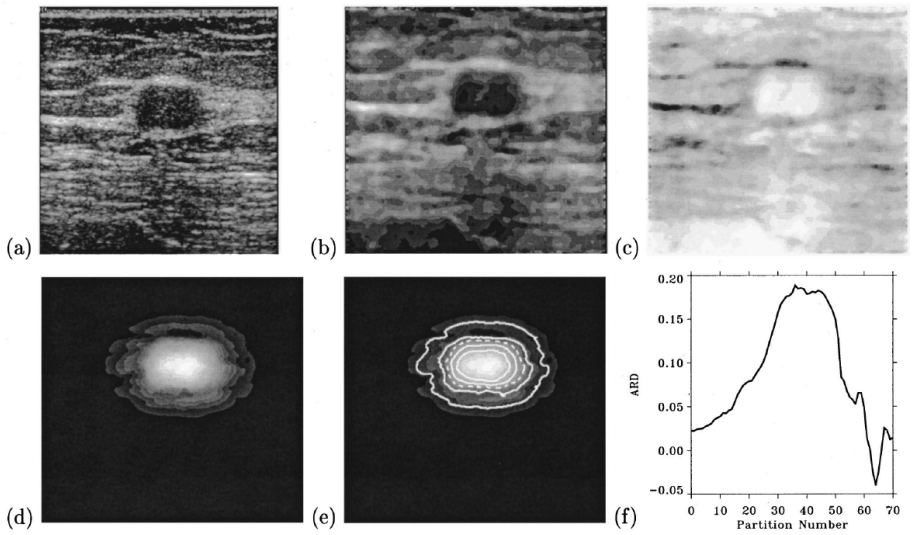


Figure 4.1: Example of the results of the segmentation processing steps in [57]: (a) the original image, (b) the preprocessed image, (c) the inverted preprocessed image, (d) the inverted preprocessed image multiplied by a Gaussian, (e) the partitions resulting from gray-value thresholding, and (f) the Average Radial Derivative as a function of partition number. In (f), the smaller the partition number, the smaller the area enclosed by the partition. For this example, the computer-chosen partition is number 35 (shown as a dashed line in (e)).

An illustration of the results of this steps is shown in Figure 4.1. One of the performance metrics used to validate this method was the overlap ratio, which ranges between zero and one, being zero in the case of no overlap and one in the case of exact overlap. The mean overlap of the

results with the manual segmentations was 0.73.

An algorithm that takes advantage of the Genetic Algorithm (GA), applied to the active contour model for boundary detection of left ventricle is described in [58]. The active contour model is defined by a parameterized deformable curve which iteratively evolves toward the target locations according to energy minimization. This model is quite sensitive to initialization and is not able to cope with significant topological changes.

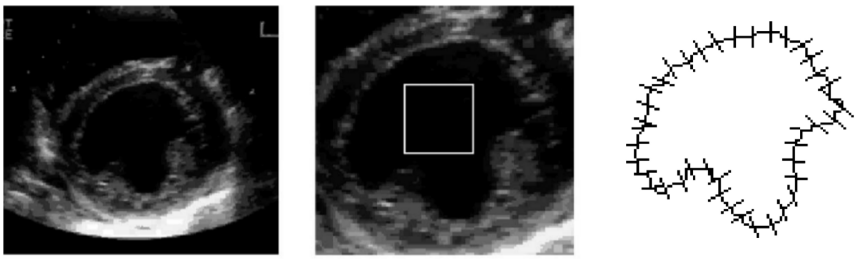


Figure 4.2: Example of the initial contour estimation results in [58]: (a) the original image, (b) the preprocessed image, (c) Search area defined by approximate perpendicular lines.

In this case, the optimization problem of active contour model uses a GA. The method begins by employing low-pass filtering and morphological operations to each image to roughly estimate its contour. Spatial discretization is performed on this contour for a number of samples of equal length. This initial estimation is used to define the search area for the next step, limited by the ends of the perpendicular lines drawn across each sample in the contour, as in Figure 4.2. This limitation of the search area helps to reduce the computational cost of the GA algorithm. Firstly, only the internal energy of the snake is used as the fitness function to apply the GA and the weights in the function are adjusted to obtain a smooth contour line. The GA is a multidimensional search algorithm particularly suitable for nonlinear problems. It starts its search from a fixed population of candidate solutions (which in this case, are the equidistant samples) and each of the candidates is evaluated with a fitness function so that genetic operators, namely selection, crossover and mutation are employed among them. Process is stopped when fitness variance is under a certain value or when a predetermined maximum number of iterations is reached. After the initial contour is optimized, using only the internal energy as the fitness function, the same algorithm is applied again but using the internal and external energy combined, at the end of which the final segmentation is determined.

The authors defined a metric of mean average distance (MAD) to measure the absolute distance between the result contour and the average of two manual delineations done by expert observers, defined by the following expression:

$$\mathrm{MAD}(a,b) = \frac{1}{\mathrm{N}} \sum_{i=1}^{\mathrm{N}} \sqrt{(x_{ai} - x_{bi})^2 + (y_{ai} - y_{bi})^2} \tag{4.1}$$

where x_{ai}, y_{ai}, x_{bi} and y_{bi} are the Cartesian coordinates of the contours a (algorithm's result) and b (ground-truth) de at ith sample location (measured by unit pixels). The value of N is the total number of samples of the contour.

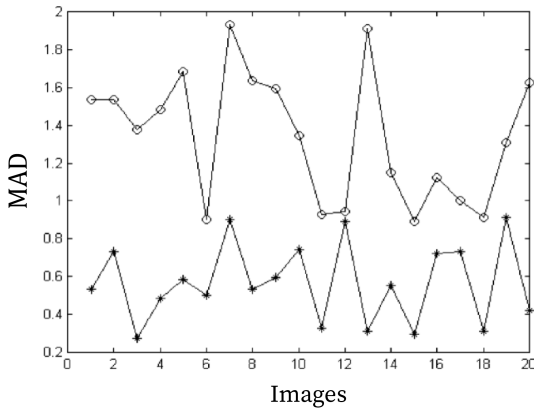


Figure 4.3: MAD (represented with circles) and standard deviation (represented with stars) between the segmentation results and the ground-truth [58].

The evaluation results computed for 20 different images are found in Figure 4.3. The variability between the algorithm's results and ground-truth is similar to inter-observer variability of the two experts.

In [59] another application of active contour model is described, in this case to perform the classification of breast masses as benign or malignant, using either 2D or 3D data.

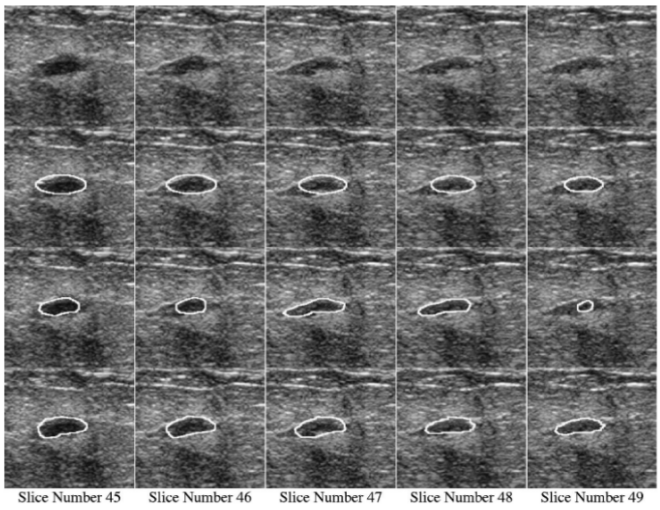


Figure 4.4: Example of the segmentation results in [59]. In row 1 there are the five slices of the original image and in row 2, the respective cross section of the initial 3D ellipsoid. The result of the 2D active contour segmentation method is presented in row 3 and then in row 4 is the result of the 3D active contour segmentation method.

The mass segmentation task consists in a energy minimization problem, which the authors propose to solve through an iterative method. The cross-section of the ellipsoid is defined by a

radiologist, and each image slice is used as the initial contour for the active contour method. The active contour's energy terms are derived from the image gray-level information. The internal energy terms related to features such as the continuity and the smoothness of the contour are defined according to the *a-priori* knowledge of the mass shape. The 3D segmentation is performed independently for each slice but then it uses the shape information (namely curvature energy term) across the slices to enhance the 2D results. The segmentation step is followed by the classification of benign and malignant masses. The segmentation results are used for feature extraction and classification of the masses.

Although the results of the segmentation in particular are not evaluated in this article, the 2D to 3D adaptation that the authors developed is interesting, as the results typically improve when the shape information is used across the slices (Figure 4.4). This adaptation can be used while applying other 2D segmentation methods.

A region-based method for the segmentation of breast lesions is presented in [60]. Region-growing is a region based method which evolves iteratively from an initial set of seed points, by examining their neighboring pixels.

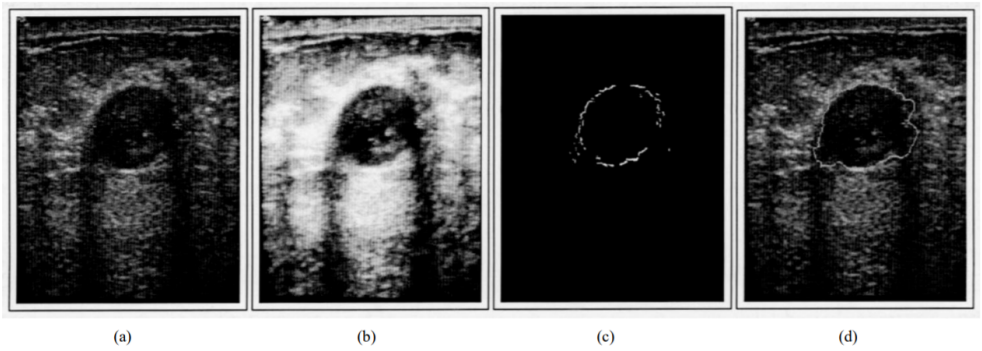


Figure 4.5: Example of the segmentation results in [60]. (a) Original Image. (b) Preprocessed image. (c) Detected boundary points. (d) Final result of the segmentation.

The first step is to perform denoising and contrast enhancement. To automatically determine the seed point of the lesion, it presents and applies a mathematical formulation of empirical domain specific knowledge. This formulation does the probabilistic analysis of a set of random points (concerning their likelihood to be real seed points), based on this knowledge, which consists on three specific discriminating sonographic features - spatial distribution of the various anatomic structures within the breast, echogenicity of the lesion, and its internal echo pattern. After that it is performed region-growing and so an initial segmentation is obtained. The directional gradient is computed to get the boundary points, which then form the initial estimate for a active contour model. As for the external forces for active contour, taking into account the complexity of the background found in sonographic images, the authors concluded gradient magnitude would not be sufficient for a stable convergence. So instead they decided to apply the so-called "balloon forces", proposed in [61], which operate on the directional gradient. An example of the application of this algorithm is shown in Figure 4.5. The results of this algorithm were computed for 42 images and

compared with manual delineations of a radiologist. The mean sensitivity with ground-truth was of 75.1% and the false positive rate was of 20.9%.

In [62] it is proposed a multiscale level-set based segmentation for echocardiographic images. The level-set method is similar to active contour, since it consists on a moving/evolving contour and it is also dependent on the user to define some parameters in the algorithm, which will determine the speed in which this contour travels in the image until it converges in the final solution. To begin a level-set algorithm, there's is the need for a initial boundary. The boundary curve is defined as the zero-level set in a space with $n+1$ dimension (where n is the number of dimensions of the original image). The added dimension is defined by a "distance function" (it is the dimensions where the curve is set to the zero-level). The update of this boundary is performed according to the derivative of the distance function and the speed.

The advantage that the level-set method presents is a better capacity of handling sharp corners and changes in topology like splits, merges and other protrusions, than other region-based methods. This way, the final results obtained with level-set segmentation don't depend so much on the arbitrary initial curve boundary like snakes (due to the existence of multiple local minima in the active contour approach). The disadvantage of the level-set method is that the computation time is higher[1].

The multiscale framework is obtained by applying a Gaussian pyramid of kernels to the images. At a coarse scale level, region homogeneity and image gradient are combined to defined the "distance function" of the level-set method and obtain an initial boundary. Then interpolation of the initial boundary to a finer scale level is done, and it is used to initialize boundary detection and serve as an external constraint to guide contour evolution. This way, at finer scale levels, the "distance function" for the level-set is defined by the image gradient and the distance to the initial boundary previously determined. The iterations continue until the finest scale level has been reached. The obtained results were analysed using MAD (Equation 4.1), for comparison with the mean manual segmentation performed by three experienced observers. The mean MAD value was of 1.643 and its standard deviation 0.503.

In [63], it is suggested a deformable model-based Bayesian segmentation for endocardial boundary detection.

Prior information about the shape of the endocardial contour is used to define the initial deformable template, along with a set of global and local transformations to take into account the natural variability. Then the data likelihood model is assessed to find the optimal template for each image. For this model, since it relies on gradient measures, the authors propose a region oriented statistical approach that involves the use of the gray level statistical distribution of each homogeneous region, and *a priori* information about the location of the endocardium with respect to other anatomical structures. Since this algorithm uses the Bayesian framework, it ends up being a cost function minimisation problem, where the fitness function is based on the inverse of the energy function (so it is maximized), where the parameters are the transformations applied to

[1] https://www2.imm.dtu.dk/ mbs/downloads/levelset040401.pdf

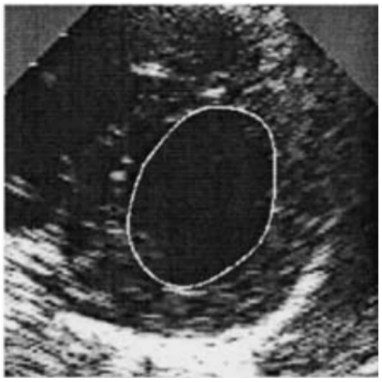


Figure 4.6: Example of the segmentation results in [63].

the template. In this case, the GA-based optimisation technique combined with a steepest ascent procedure is used.

The accuracy obtained for 50 images varied between 77.6% and 80.2%. An example of the segmentation result obtained through this algorithm is shown in Figure 4.6.

The algorithm in [64] also resorts to the Bayesian framework, in this case to perform adaptive region-growing for segmentation of the left ventricle in echocardiographic sequences. For the energy function, it is implemented a weight function accounting for both local and global statistics and an adaptive property with slow spatial variation to compensate for image noise. It also makes use of a multiresolution framework and so, starting from the highest resolution image, a multiresolution discrete wavelet transform (DWT) pyramid is built. The segmentation starts from the coarsest resolution using the K-means algorithm to obtain the initial boundary. Then it is passed on successively to finer resolution images, based on [65].

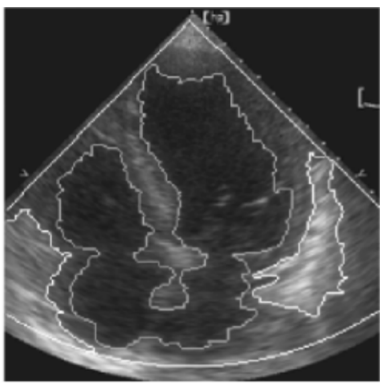


Figure 4.7: Example of the segmentation results in [64].

An result of this algorithm is found in Figure 4.7. MAD (in pixels) was on of the criteria used to evaluate the algoritm and compare of generated segmentation relative to measurement made manually by a medical expert. The MAD was of 5.96 and its standard deviation 2.20.

The article [66] proposes a solution for automatic segmentation of breast lesions that combines the advantages of Self-organizing map (SOM) classification and morphological watershed

segmentation. SOM is a type of NN architecture, which uses unsupervised learning. Watershed transformation is more sensitive to noise and contrast in the image, but it is a reliable unsupervised model. Pre-processing methods can be used to attenuate this problem and so, given the variety of tissues in breast US images, an appropriate texture-based preprocessing filter is developed for each tissue texture. The inputs of the SOM are the autocovariance-coefficients and the variance of each pixel (computed with the gray levels of its neighboring pixels). The output of SOM is the textural class map of the image. According to each textural class, a specific preprocessing filter is applied and then the watershed segmentation is finally performed.

The performance measures used for this method were precision (subsection 3.3) and match rate (MR), which is defined as the following:

$$\mathrm{MR} = \left(1 - \frac{|\mathrm{Area_M} - \mathrm{Area_W}|}{\mathrm{Area_M}}\right) \times 100 \tag{4.2}$$

where $\mathrm{Area_M}$ corresponds to the manual segmentation and $\mathrm{Area_W}$ to the automatic segmentation of this method. The average precision of the 20 images that were tested was 81.70% with minimum of 67.56% and the mean MR was 94.66%, with a minimum of 83.11%.

The NNs introduced in section 3.4 have an interesting application on real-time segmentation of the prostate to be used during the biopsy procedure, described in [55].

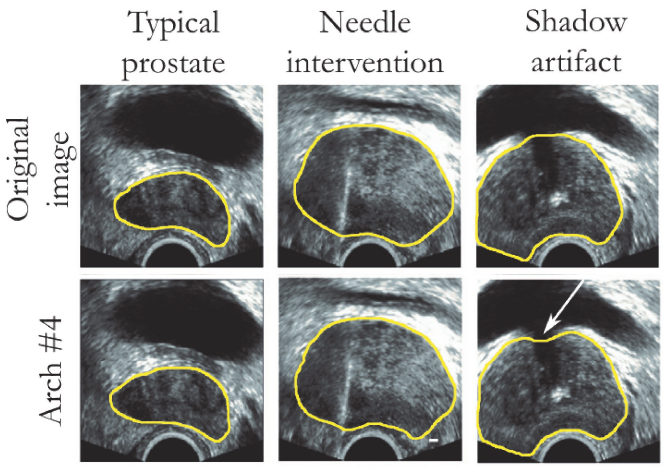


Figure 4.8: Example of the segmentation results in [55]. The method was tested in typical prostate images and prostate images with shadow artifacts and the presence of needles.

Both CNN (based on the architecture of the U-net) and Long short-term memory (LSTM) Network are used for spatial and temporal analysis, respectively. An improved version of the U-Net architecture for prostate segmentation is applied, which is called "ResU-Net". It is improved with the implementation of a residual neural net to improve training, and dilated convolutions to improve the local prediction of the network. Both extensions are applied within the expanding branch of the U-Net. The implementation of a convolutional variant of recurrent units is also done to incorporate temporal features in the analysis. The analysis of each 2D image acquired

is done separately, since the acquisition in the third dimension is very sparse. The weights of the convolutional layers are initialised with Gaussian-distributed small random numbers. For the upsampling layers, the initialisation of the weight matrix corresponds to a bilinear interpolation kernel. The training set was composed by 2238 images. The results (Figure 4.8) of the testing process, which was performed using 1017 images got a DSC of 91.84 $\pm$ 2.85. These results correspond to the stage where performance of the LSTM (time series analysis) is not taken into account, as it is not relevant in this context.

In [56] three DL methods - Patch-based LeNet, a U-Net, and a transfer learning approach with a pretrained FCN-AlexNet - are implemented, tested and compared, based on their ability to correctly detect lesions on breast US images. Two datasets were used, one with 306 images with a mean image size of 377$\times$396 pixels and other of 163 images with the mean size of 760$\times$570 pixels. The results obtained when using the same database are more accurate and reliable to assess their relative efficiency. On the other hand, having two different datasets is beneficial to verify the capacity of the method to adjust to images with different acquisition settings. The patch-based CNN, which uses image patches for training, has the drawback of the redundancy introduced by patch overlap, and it also results in a very time-consuming algorithm. In this case, the patches were 28$\times$28 pixels and stride 1. The U-Net avoids this redundancy and returns a pixel-wise segmentation rather than a single probability distribution per image. The transfer learning approach was develop as an effort to surpass of limitation of data deficiency in medical imaging research, by pre-training the model with non-medical images. The type of images used for training are not specified in the article. After this initial training, the classification function is changed to optimize the network to learn the specific features of breast US.

Overall, after a lesion-wise analysis (not pixel-wise), FCN-AlexNet got the best score for True Positive Rate (0.99 for dataset A and 0.93 for B) and the LeNet returned the lower amount of False Positives per image (0.07 for dataset A and 0.09 for dataset B). Some examples of the segmentations of each method are shown in Figure 4.9.

4.1.2 Segmentation of follicles

This section presents the state-of-the-art automatic methods for follicle segmentation. The reviewed methods are diverse, including both traditional image processing approaches such as Canny edge detection and energy minimization algorithms, and deep learning approaches which employ a Multilayer Perceptron and a U-Net.

A method for segmentation of follicles is described in [67]. Its methodology begins with a preprocessing step, by employing a Contourlet transform for denoising, and then histogram equalization. The Contourlet transform performs multi-resolution and multi-direction decomposition of the image by using the Laplacian pyramid filters and direction filters. For edge detection, the Canny operator is used. The Canny edge detection algorithm is composed by four stages: (i) noise reduction with a Gaussian filter; (ii) computation of the edge gradient and direction, using the first derivative of the image in both horizontal and vertical direction (obtained with Sobel kernels); (iii) non-maximum suppression in the direction of gradient (perpendicular to the edges),

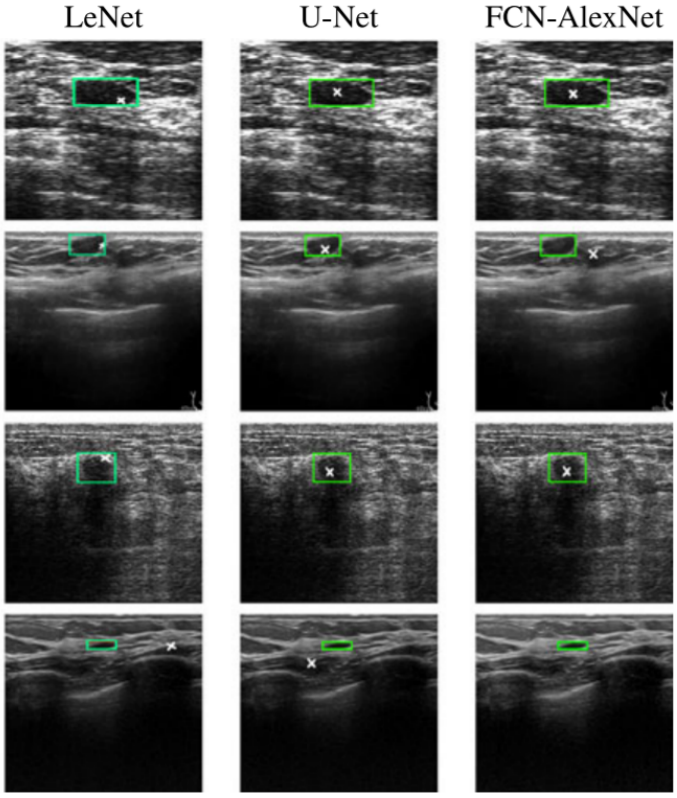


Figure 4.9: Examples of the segmentation results in [56]. Each row corresponds to a different image and each column to the respective result of each method. The green rectangles indicate the ground-truth and the white crosses indicate the lesion detected by the automatic method.

in order to set to 0 every pixel that may not constitute the edge; (iv) hysteresis thresholding, which consists on discarding previously detected edges with a low intensity gradient and that have no connections with edge lines with high intensity gradient[2]. After the edge detection step, two morphological operations are applied: morphological dilation to close broken edges (resulting in the initial segmented image) and erosion to get rid off spurious elements that are falsely detected due to remaining noise. Besides morphological erosion, a minimum size is picked also to discard small objects that are assumed to be irrelevant. Parameters such as the size of the kernels used in these operations and the size bellow which an object is considered irrelevant were chosen by trial and error. The remaining detected regions are filled, labeled and several geometric features are determined.

The ratio between the major axis length and minor axis length (Figure 4.10),r , of the elements is computed, and the mean ($\bar{r}$) and standard deviation (σ_r) of all of them are used to formulate the classification rule. The results of the analysed images were $\bar{r} = 1.8427$ and $\sigma_r = 0.5971$. Empirically, the authors found that the best rule would be to classify as follicles the objects with value of R comprehended between 4σ intervals around the mean of the ratio r. The testing phase was done using 25 images (with 93 follicles in total, according to the ground-truth segmentation

[2]https://docs.opencv.org/3.4/da/d22/tutorial_py_canny.html

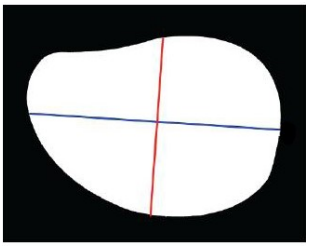


Figure 4.10: Illustration of the definition of major axis length (blue) and minor axis length (red) [68].

determined by a medical expert) and resulted in a follicle-wise true positive rate of 75,2%, with the false positive rate 22.5% and false negative rate 24.1%.

In [15], it is proposed a method based on two learning frameworks, namely probabilistic boosting-tree (PBT) and clustered Marginal Space Learning (cMSL), to perform the detection and segmentation of ovarian follicles in 3D volumes.

A probabilistic boosting-tree is a deep learning method in a tree form, where, in each node, a conditional posterior probability of a certain output (based on the input, which are the features of the input data) is computed based on a classifier. The outputs of each node are propagated until top node, which returns the overall posterior probability [69].

In this case, the probabilistic framework is used to estimate the volume and the localization of the ovary as an 3D ellipsoid, to reduce the search space. The parameters used to define this ellipsoid are center position, orientation and size. The probability measure is computed and so the parameters that most likely correspond to the ovary are inferred. To take into account the possibility of an ovary with irregular shape, the probability volume of the ovary is determined based on these parameters. This probability is used so that the confidence of the ovary detector is taken into consideration for the classification of objects that are near the initial ovary mask (inside the initial ovary mask this probability is 100%).

Then for the follicle detection, given the fact that the shape of a follicle is very variable and that they can be squeezed together, no common shape is assumed and the parameters to represent them are only the position of the center and the size of its the minimum bounding box in directions x, y and z. MSL is a algorithm for single object detection, but since this is a case of multiple object detection (an ovary may contain more than 30 follicles), to infer the parameters, clustered MSL is proposed (also using PBT to take into account the location of the ovary). With this, the estimation of the follicles' location is done and then the actual segmentation follows. For this step a database-guided graph-cut segmentation is performed.

To make the training more robust, a multi-scale follicle detector is trained with 3 levels (from 2 to 32 mm). Also the whole algorithm is applied in a multiscale manner, from coarse to fine, with voxels sequentially of $(2\text{mm})^3$, $(1\text{mm})^3$ and $(0.5\text{mm})^3$. The training and testing sets were composed by 6448 and 1660 follicles, respectively (Figure 4.11). For the testing set, the false positive rate (voxel-wise) was 20.7%, and a false negative rate (voxel-wise) was 5.9%. These values are quite high, unfortunately, and so the authors suggest that in future work it could be

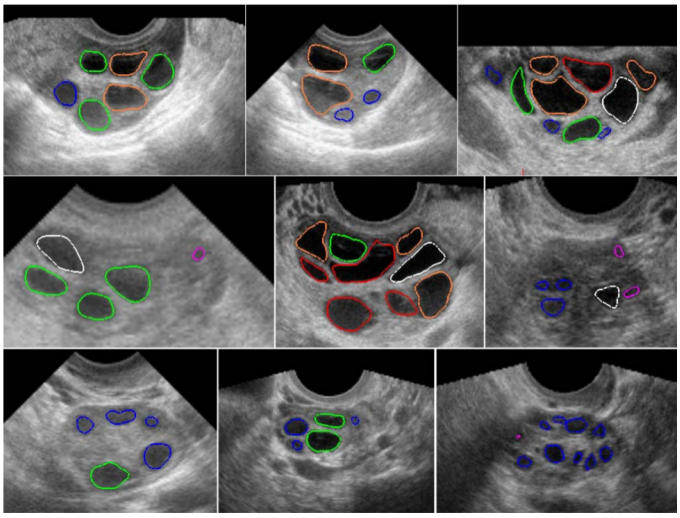


Figure 4.11: Examples of segmentation results of different volumes (shown in 2D) from [15].

improved by training a specific designed classifier to lower the amount of false positives and false negatives in the results.

The segmentation process in [70], focuses on the texture of the structures in the image, and it consists on a fully automatic method for the detection of ovarian follicles. It takes advantage of that fact that US images of the ovaries have different echo-texture patterns for follicles, artifacts, speckle noises and other tissues, to develop the segmentation method. The preprocessing step consists on denoising with an adaptive Wiener filter, negative transformation of the gray-scale image and histogram equalization to enhance the contrast. Morphological operations and a Sobel filtering operation are followed by the application of the active contour without edge method (which was developed to detect object boundaries even if they are not defined by a gradient), resulting in the definition of the ROI. These regions are the ones used in the next steps: Gray Level Co-occurrence Matrix (GLCM) technique is applied to extract second-order texture features (contrast, correlation, energy and homogeneity) and they are used for the learning process of a Multilayer Perceptron (MLP) with 3 hidden layers. MLP classifies the pixels in the regions of interest into follicles and non-follicles. Levenberg-Marquardt backpropagation algorithm (known as the damped least-squares method) was used to perform the training of the network.

The accuracy of the results (Figure 4.12) was 96%, sensitivity was 99% and specificity of 93% (follicle-wise).

Another variation of a DL method (similar to the previous method [70], but with additional steps) is proposed in [71], which is based on an hybrid algorithm that combines MLP with the Particle Swarm Optimization (PSO) algorithm. First of all, an adaptive filter, image negative transformation, histogram equalization and morphological operations are employed for image preprocessing. Then active contour without edges is performed to obtain an initial segmentation. The feature extraction is done very meticulously, including several geometric, first order texture and Gray Level Co-occurrence Matrix (GLCM) features for each area segmented. The next step is the selection of the best performing features using PSO. PSO is a computational method that it-

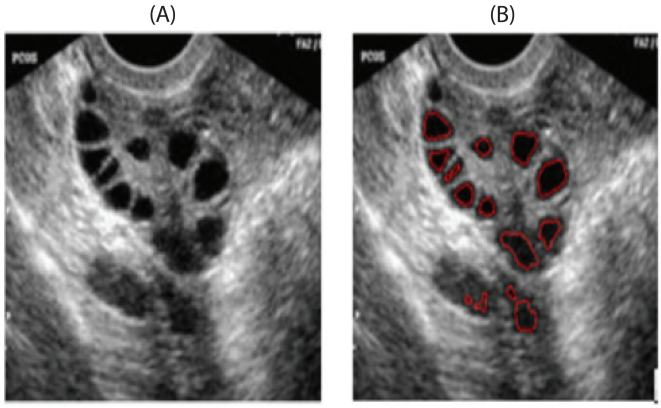


Figure 4.12: Example of segmentation results from [70]. (A) is the original image and (B) is the image superimposed with the segmentation boundaries (red).

eratively determines the best candidate solutions (or particles), which in this case are the optimal image features; the so called swarm is the complete set of candidates. The particles are placed in a search space and moved according to the algorithm until the global optimum is found. Then, the MLP is run with the features of that iteration to evaluate the training performance. For that, the fitness function employed is the Mean Squared Error. The algorithm runs until the maximum iterations or minimum error criteria is reached. For the configuration of the MLP, it has only one hidden layer and the number of neurons was chosen based on rule of thumbs which states that the number of hidden layer neurons should be about 2/3 of the size of the input layer. Levenberg–Marquardt algorithm is used for the training of the MLP.

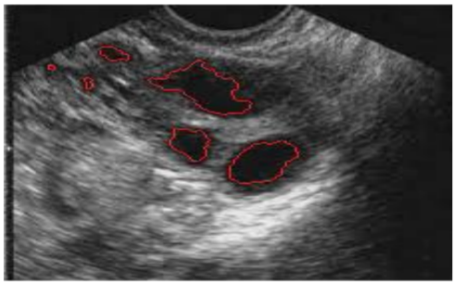


Figure 4.13: Example of image superimposed with the segmentation boundaries (red) from [71].

When the training performance is high enough, the network is saved as it is and the testing phase is performed. The results (Figure 4.13) of the final MLP, using a 25 image dataset, were 98.3% for accuracy, 100% for sensitivity and 96.8% for specificity (follicle-wise). Comparing these results to the results of [70], it is possible to conclude that considering other features besides GLCM features and select them using PSO improves the performance of the method.

The 3D segmentation method developed by Cigale *et al.* [72] is divided in two phases: detection, in which candidate follicular regions are determined, and recognition, in which the those candidates are classified as follicles or not. The detection is performed using a Wavelet transform that scans the entire volume (slide by slide) searching for objects whose inside is darker than the

outside. This scan is done in the three orthogonal directions x, y and z, and in multiples scales (defined adaptively), in order to detect the entire range of sizes of follicles. For the recognition phase, begins by performing optimum thresholding according to the mean gray-values of each region found in the previous phase, and then regions that were incorrectly fused are separated based on the convexity of the regions. Finally, the likelihood of the region actually correspond to a follicle is computed by means of a product of normalized features, namely gray-values found inside in relation of the gray-values of its surroundings, convexity, compactness (comparison to a spherical shape) and the concentration of the darker voxels around the centroid.

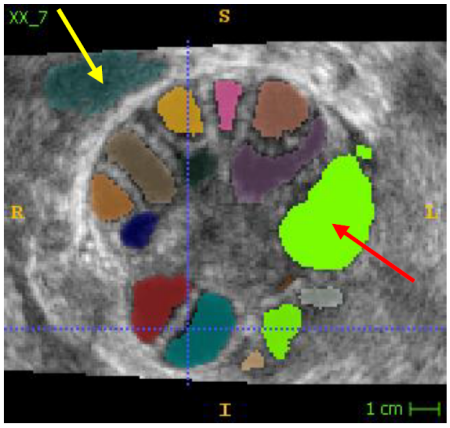


Figure 4.14: Example of image superimposed with the 3D segmentation from [72]. A higher opacity in the recognized follicles, means the likelihood of that follicle is higher. The red arrow point to the follicle with higher likelihood in this volume, and the yellow one to the follicle with lower likelihood.

This method was tested in 30 3D ultrasound volumes (with a total of 151 follicles) and compared to manually annotated volumes. A total of 21 equidistant likelihood thresholds were applied and the ROC curve was defined for each one of the 30 volumes separately. The optimum threshold was found for each volume and the mean of all of them was then applied to all of the volumes. The results of the mean likelihood threshold, led to 0.88 sensitivity and 0.84 specificity, follicle-wise (Figure 4.14). A measure was defined as the result of the product between the sensitivity and the precision of the results (voxel-wise). For this measure, which only takes into account the follicles that are a true positive in the follicle-wise analysis, the results led to a mean value of 0.73. It would be interesting to determine the likelihood threshold using only some of the images (comparable to the use of a training set in DL) and then test it in the images that were not used to determine that value and see it results in a similar performance.

In the work of Wanderley *et al.* [73], a fully convolutional neural network (fCNN) based on the U-Net architecture [53] is used to simultaneously obtain the segmentation of the follicles and the stroma of the ovary. Several regularization techniques are proposed, by means of a loss function with higher penalization for the prediction of follicles outside of the ovary area and between two follicles. The use of Contrast Limited Adaptive histogram equalization for pre-processing is also proposed.

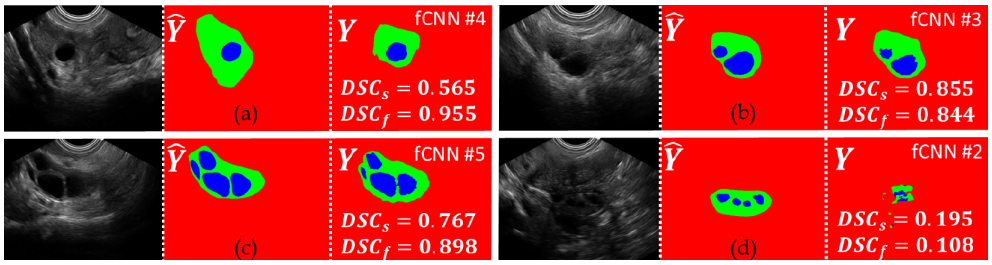


Figure 4.15: Examples of segmentation results from [73]. Background area in red, follicle area in blue and ovarian stroma area in green. (a) the best follicle DSC, (b) the best ovarian stroma DSC, (c) a standard case, (d) the worst image.

The method was tested on 15 images and it was concluded that the proposed pre-processing method and regularization techniques did not significantly improve the performance of the segmentation. The results of the test-set (Figure 4.15) of the fCNN configuration, without pre-processing and regularization techniques, got a DSC of 0.757 ± 0.207

On Table 4.1, it is given an overview of methods included in this section and the respective performance evaluation.

4.2 Concluding Remarks

There is a very diverse range of methods to perform segmentation of follicles and other hypoechoic anatomical structures in US images. The segmentation of follicles represents a particular challenge, when compared with the segmentations mentioned in Subsection 4.1.1, since it involves the segmentation of a high number of objects that may be very close to each other, in each image/volume.

It is important to remember that, even though the presented methods only reported the segmentation or the detection of the follicles, it is important to compute the results for both tasks, given that both the size and the number of the follicles is requested in clinical practice. To have a full understanding of the effectiveness and the limitations of a method is it essential to submit it to a follicle-wise and pixel/voxel-wise analysis and take from them several performance metrics.

From the analysed methods, it is possible to conclude that the best performing algorithms include a multiscale analysis, employ some form of DL method or perform a probability analysis based on *a-priori* knowledge of the features of the segmented structure.

The FPR is considerably high in most of the follicle segmentation methods as it is difficult to distinguish follicles from other hypoechoic structures in the ovary such as small vessels.

Table 4.1: Overview of the described methods for ovarian follicle segmentation and respective performance measures.

Reference	Brief method description	Performance evaluation
[67]	Canny edge detection and morphological operations.	Follicle-wise: - TPR 75,2%; - FPR 22.5% - FNR 24.1%
[15]	Probabilistic Boosting-Tree and Clustered Marginal Space Learning to estimate ovary and follicle localization (multiscale approach), followed by database-guided graph-cut segmentation.	Voxel-wise: - FPR 20.7% - FNR 5.9%
[70]	Active contour without edge method to define regions of interest. GLCM technique is applied to extract second--order texture features, which are used as input in a MLP.	Follicle-wise: - Accuracy 96%; - Sensitivity 99%; - Specificity 93%
[71]	Active contour without edges is to obtain an initial segmentation. Geometric, first order texture and GLCM features are extracted from the segmented regions. Particle Swarm Optimization is used to select the best performing features (using MSE as fitness function). The selected features are the input of the MLP.	Follicle-wise: - Accuracy 98.3%; - Sensitivity 100%; - Specificity 96.8%
[72]	Directional Wavelet Transform to determine candidate regions, in multiples scales. Features of candidate regions are used to determine their likelihood to correspond to actual follicles. The optimal likelihood threshold is determined based on the ROC-curve.	Follicle-wise: - Sensitivity 88%; - Specificity 84%
[73]	Fully convolutional neural network based on the U-Net architecture [53]	Pixel-wise: - DSC 0.757 ± 0.207

Chapter 5

Methodology

This chapter presents the methodology used to select the best performing method for the segmentation and detection of follicles and ovaries in US images. The main challenges of this work are discussed and the available dataset and the details of the proposed experiments are described.

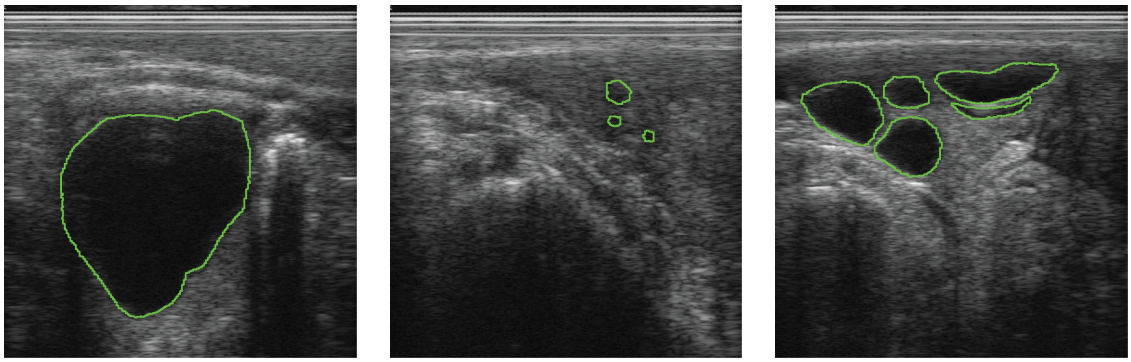
This work is developed based on a baseline method, to which modifications and additional steps are proposed in order to improve its performance. The baseline method uses B-mode images as input of a U-Net and it is optimized using a loss function based on DSC, to obtain the prediction of the follicle or ovary segmentation and detection. The proposed experiments include alterations regarding 1) input data, namely the type of data (B-mode and BRF), and pre-processing methods, 2) input size and U-Net architecture hyperparameters, 3) the implementation of a loss function based on DSC with higher penalization for the follicle borders and 4) post-processing methods.

5.1 Objectives and Challenges

As mentioned in Chapter 1, the main purpose of this dissertation is to develop an algorithm based on a U-Net which performs the automatic segmentation and detection of follicles and ovary in US images. The accomplishment of this aim entails the following objectives:

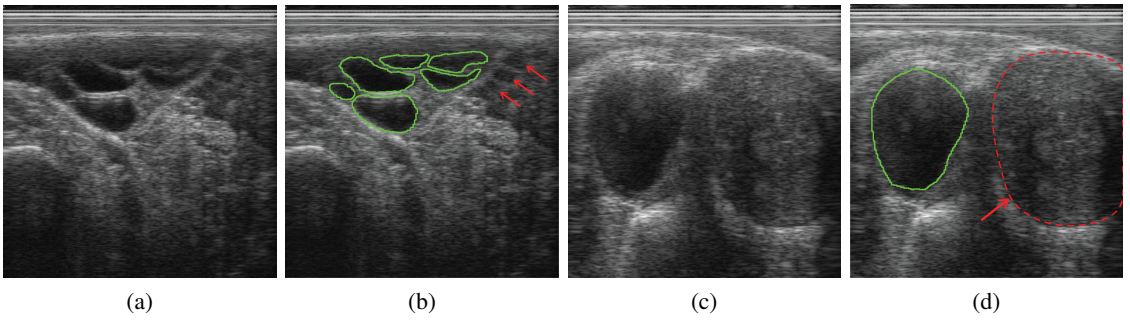
- Develop a baseline model for follicle and ovary detection and segmentation.
- Assess the influence of the type of input data and pre-processing methods on the model's performance.
- Assess the importance of the image context analysis on the model's performance.
- Develop a loss function to prevent the incorrect union of neighboring follicles in the segmentation predicted by the model.
- Assess the effect of post-processing methods on the model's performance.
- Establish and validate the final model for follicle and ovary detection and segmentation.

One of the main challenges of follicle segmentation is the high variability of the shape and size of the follicles (Figure 5.1), which hampers the identification of common features during the U-Net's training.



(a) (b) (c)

Figure 5.1: Examples of linear B-mode images of the dataset to demonstrate the size and shape variability of the follicles. GT delineations of the follicles are in green.



(a) (b) (c) (d)

(e) (f) (g) (h)

Figure 5.2: Examples of linear B-mode images of the dataset, with GT delineations of the follicles in green. (a-b) Proximate follicles and hypoechoic structures (red arrows); (c-d) Large hypoechoic structure similar to a follicle (red arrow); (e-f) Low signal-to-noise ratio and acoustic shadowing; (g-h) Follicle intercepted by a septum (yellow arrows) and hypoechoic structure similar to a follicle (red arrow).

The presence of hypoechoic structures besides follicles, such as vessels and the bladder, is problematic as it may result in a high number of false positive objects in the predictions (Figures 5.2.a to 5.2.d). The final model should be able to gather the global context of each image and to understand the relation between the position of the structures in order to distinguish between follicles and other hypoechoic structures. For instance, there is always ovarian stroma tissue around the follicles, so if the model is able to identify the texture of the tissue and perceive that requirement,

the number of of false positive objects in the predictions decreases.

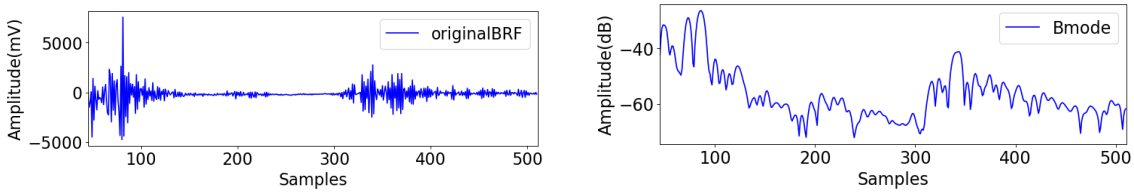
Another challenge is the low signal-to-noise ratio of the US data and shadowing effect, which makes it difficult to identify the borders of the follicles (Figures 5.2.e and 5.2.f). On the other hand, the septa with a high degree of echogenicity intercepting malignant follicles (Figures 5.2.g and 5.2.h), is visually similar to the separation between neighboring follicles. In the case of ovary segmentation, the low signal-to-noise ratio hampers the identification of the texture of the ovarian stroma tissue and, consequently, its borders.

5.2 Dataset

The dataset used in this dissertation is based on 99 original BRF transvaginal US images of the ovary. These images were acquired at Centro Hospitalar de São João, with consent of the patients, while these attended first time appointment for fertility treatment planning. Each image contains one ovary with single or multiple follicles. The original BRF data was acquired with an EC9-5/10 Endovaginal Microconvex transducer (frequency range 9-5 MHz and image field of 124°). The acquired BRF images are 21.7 mm wide, with a focal distance between 58.8 mm and 107.1 mm, with the original image dimensions of 192×[2000, 6200] px.

Two types of image were extracted from the original BRF data, namely, absBRF and B-mode.

AbsBRF (referred to as simply BRF in the rest of the document) is obtained by computing the absolute value of the original BRF data and normalizing the obtained intensity to a range of [0, 255].



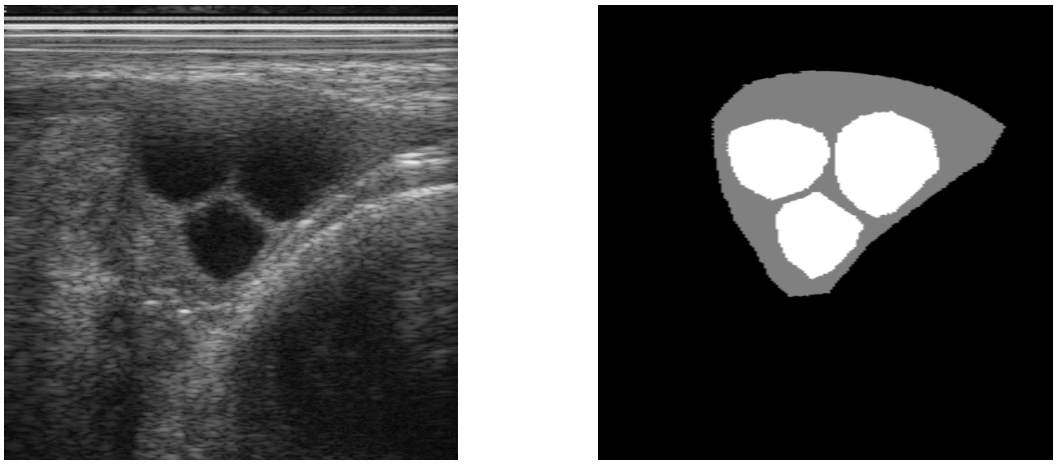
(a) (b)

Figure 5.3: Example of corresponding intensity profiles of one pixel column in a image in a) original BRF data; and b) B-mode data.

The B-mode images are attained by applying envelope detection and log compression to the signal obtained after quadrature signal demodulation and filtering with a Hamming window of the original BRF data (Figure 5.3). The parameters used during image acquisition were used during the data conversion processes.

The images were resized to 512×512, to normalize the resolution of the dataset. Moreover, resizing of the BRF data, along the depth direction, is a decimation step, used in the conversion of BRF to B-mode.

Figure 5.4.a represents a linear B-mode image (reconstructed from the original BRF data) and Figure 5.4.b represents the corresponding GT image. The GT contours of ovaries and follicles



(a) (b)

Figure 5.4: Example of (a) a linear B-mode image; (b) GT of a) with follicles represented in white and stroma in grey.

were drawn by a medical expert on the scan-converted B-mode images and mapped afterwards to the linear space.

For 8 of the 99 images, two different annotations were made by different medical experts, resulting in a total of 107 images for the dataset.

For this work, it was necessary to prepare two different dataset arrangements: one for the conducted experiments to perform model comparison, and another for the validation of the final model.

The dataset division used for model comparison was as follows: 92 images were used for 5-fold cross-validation and 15 for test. Due to the uneven division of the number of images for 5-fold cross-validation, three of the folds were composed by 74 images for training and 18 for validation and the other two folds were composed by 73 images for training and 19 for validation.

For the validation of the final model, it was necessary to obtain the results for every image of the dataset. Ten dataset arrangements were defined, so that every image was in the test-set of one of the arrangements. In these datasets, the number of images in the test-set is between 9 and 12, in the validation-set between 19 and 20, and in the training-set between 75 and 76. The difference between the image division in the dataset arrangements is due to patient restrictions and the uneven division of the number of images for 5-fold cross-validation. Dataset division used is described in more detail in Appendix A.

5.3 Equipment and tools

The equipment used for the development of this work has the following specifications:

- Operating System - Ubuntu®16.04.5 LTS;
- Operating System type - 64-bit

- Memory - 42 GB
- Processor - Intel®Core™i7-6700K CPU @ 4.00GHz × 8;
- Graphics - NVIDIA GeForce GTX 1080 (8GB)

The programming language used for the development of the algorithms is Python™. The main explored libraries are:

- Pytorch, for the implementation of the U-Net model;
- OpenCV, for the implementation of image processing techniques and analysis of the obtained predictions;
- TensorFlow, to record and visualize the progression of the model's results.

5.4 U-Net Architecture

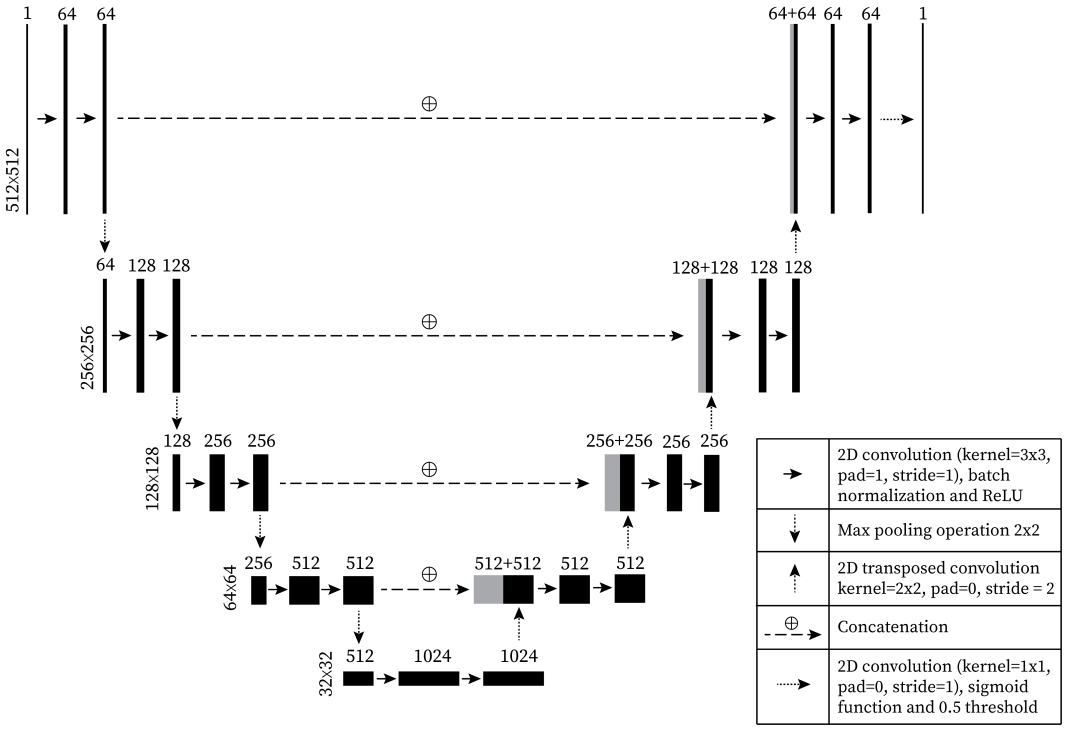


Figure 5.5: Illustration of the implemented U-Net architecture.

The implemented U-Net architecture is similar to the original [53], described in Subsection 3.4.

This architecture is composed by a contracting path and a expanding path. It begins with the contracting path, composed by a series of four contracting blocks, each block performs two sequences of one convolution (with 3×3 kernel, stride of 1 and padding of size 1), batch normalization and rectified linear unit, followed by a max-pooling operation with a 2×2 kernel and stride of 2. In each contracting block, the first convolution results in an increase of the number of feature

channels: the one in the first block goes from 1 to 64 features and the ones in the following blocks expand the number of features by a factor of 2.

The output of these four contracting blocks is passed to the block that connects the two paths, which consists of two sequences of one convolution (with 3×3 kernel, stride of 1 and padding of size 1), batch normalization and rectified linear unit. This block also results in an increase of the number of feature channels.

This outcome is then passed to the expanding path. This path is composed by 4 expanding blocks. Each one of these blocks corresponds to an up-sampling step with a scale factor of 2, performed by a transposed convolution, then a concatenation with the corresponding feature map from the contracting path (as described in Figure 5.5), and two sequences of one convolution (with 3×3 kernel, stride of 1 and padding of size 1), batch normalization and rectified linear unit. In this case, in each expanding block, the first convolution results in a reduction of the number of feature channels by a factor of 2. Lastly, a convolution is performed, with a 1×1 kernel and stride of 1, whose outcome is a 512×512 image with one feature channel.

The sigmoid function is applied to the pixel values, producing a pixel-wise probability distribution of it belonging to the region of interest (that being follicle or ovary). A 50% probability was chosen to threshold the probability distribution and obtain the final prediction. Taking into account that only one class was being segmented at a time (follicle vs. background or ovary vs. background) a 50% probability is a neutral parameter in terms of interference with the sensitivity and specificity of the model.

5.5 Learning Framework

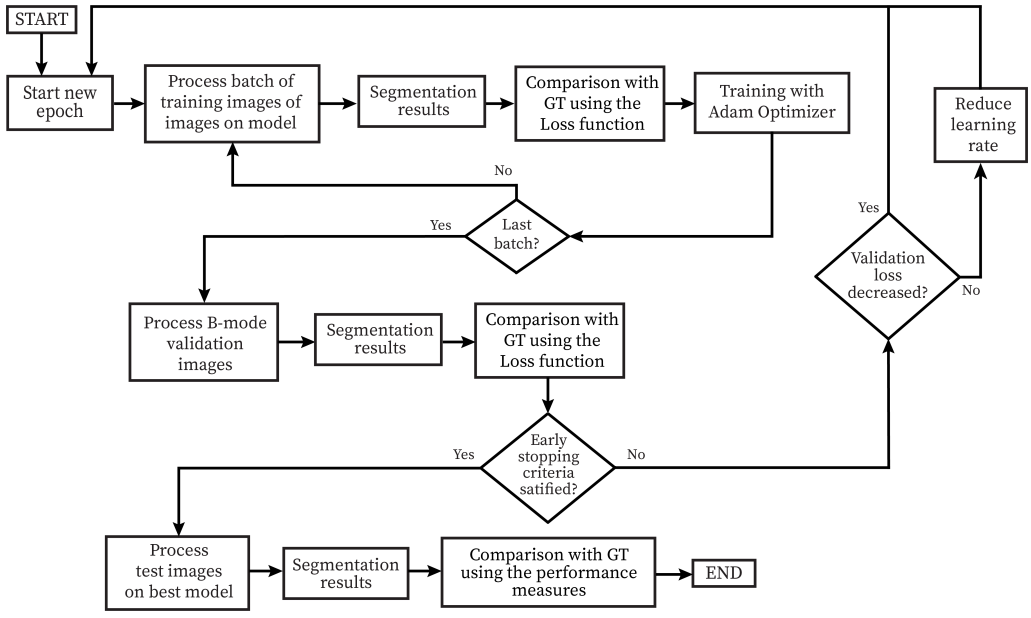


Figure 5.6: Schematic overview of the learning framework.

The method uses a learning framework as illustrated in Figure 5.6, based on the U-Net architecture. In the first epoch the weights of the U-Net model are initialized using a normal distribution. For each epoch, the optimization step is performed after the back-propagation of the loss of each batch, using the Adam (Adaptive Moment Estimation) Optimization Algorithm, mentioned in Section 3.2.2.

In the end of each epoch the validation images are processed by the resulting model and the early stopping criteria is verified, based on the loss obtained for those predictions. This criteria consists in stopping the training of the model after a certain number of epochs (patience parameter) with improvement of the validation loss of the model, from one epoch to the next, lower than the minimum delta parameter. This early stopping callback is used to decrease the duration of the training process after convergence, while still allowing for the U-Net model to train as many times as necessary to properly learn from the data.

If the early stopping criteria is not met, before the beginning of the next epoch, it is verified if the validation loss of the model has decreased when compared to the previous epoch. If the validation loss is the same or higher than in the previous epoch, then the learning rate is reduce by a factor (called factor of reduction of learning rate).

On the other hand, if the early stopping criteria is met, the model with best performance (measured by the loss function) in the validation set is applied to the test set and the corresponding validation is performed.

The sensitivity of the model to hyperparameters directly related to the learning rate was studied. To evaluate the sensitivity of the model to different initial learning rates (0.1, 0.01 and 0.001) and different factors of reduction of learning rate (5, 10 and 20), the evolution of the loss value for the validation-set over 100 epochs was assessed and all presented a similar convergence, as shown in Appendix B. Taking this into account, the initial learning rate of the optimizer was arbitrarily set to 0.001 and reduced by a factor of 10 after every epoch for which there was no improvement of the the loss for the validation set.

Due to memory limitation of the equipment, the batch size used is of 2 images. A higher value for the batch size (e.g. 4 images) would help the U-Net to have a better convergence given that each gradient would be computed with more data.

As shown in Appendix B, the maximum number of epochs was set to 100 which is high enough to ensure that the training is stopped by the early stopping criteria callback and not by reaching the maximum number of epochs. The patience parameter was set to 30 since it ensures that the model reaches convergence. The minimum delta parameter was set to 10^{-3}, given that an increase of 10^{-3} for the loss function is irrelevant for the model's performance.

Data augmentation was applied for the validation of the final model. It was not used for model comparison has it would have no impact in the conclusions in terms of the difference between models. The data augmentation mechanism consisted on horizontal flips, rotations, scaling and shear transformations, and also the introduction of image noise (changes at the level of low-pass filtering and cut-off frequencies used for the BRF to B-mode conversion).

5.6 Baseline Method

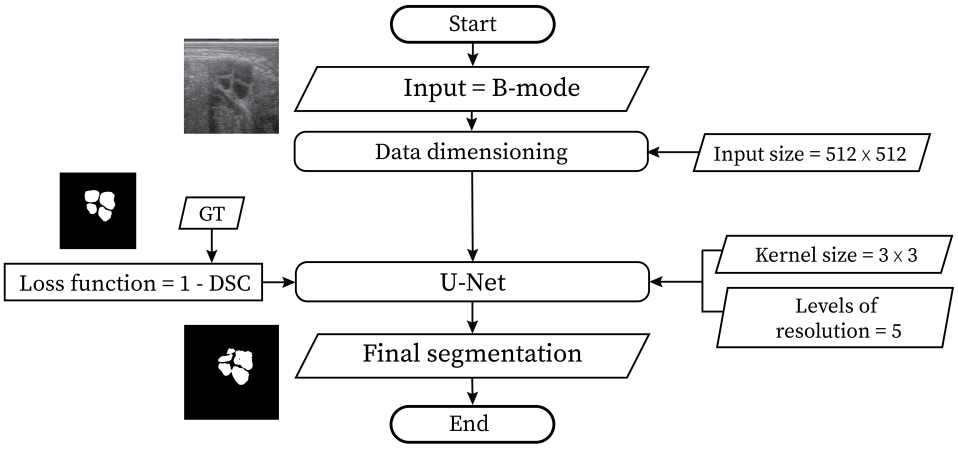


Figure 5.7: Schematic overview of the baseline method.

The baseline method for the automatic segmentation of the follicles and the ovary, was defined, as shown in Figure 5.7.

The baseline method receives as input B-mode data, with size 512×512 px and without pre-processing. The model uses the architecture of the U-Net, as described in [53], that is, with convolutions using a 3×3 kernel and 5 levels of resolution.

The loss function used for the baseline method is based on the Dice Similarity Coefficient (DSC) [52]. Given that the DSC is a ratio that ranges from 0 to 1, where 0 corresponds to no overlap and 1 corresponds to complete overlap, for training and validation it is used the corresponding difference to one, as follows:

$$\text{DSC Loss} = 1 - \frac{2\text{TP}}{2\text{TP} + \text{FP} + \text{FN}} \tag{5.1}$$

The output of the U-Net, both for follicle and ovary segmentation, consists on the final segmentation after threshold since no post-processing methods are applied.

5.7 Proposed Experiments

There are still opportunities for further enhancements to the aforementioned baseline method, namely regarding:

- Enhancement of image features through image pre-processing techniques, which may allow the model to perform the extraction of more relevant features, like edges, in detriment of features related to noise.
- Ability to perform an analysis of the global image context;

- Formulation of the loss function, as it entails low penalization for prediction errors that lead to inaccurate follicle count, namely the incorrect union of neighboring follicles and small false negative objects;
- Improvement of the predictions performed by the U-Net through post-processing methods, by taking advantage of *a priori* knowledge about the characteristics of the follicles and the ovary.

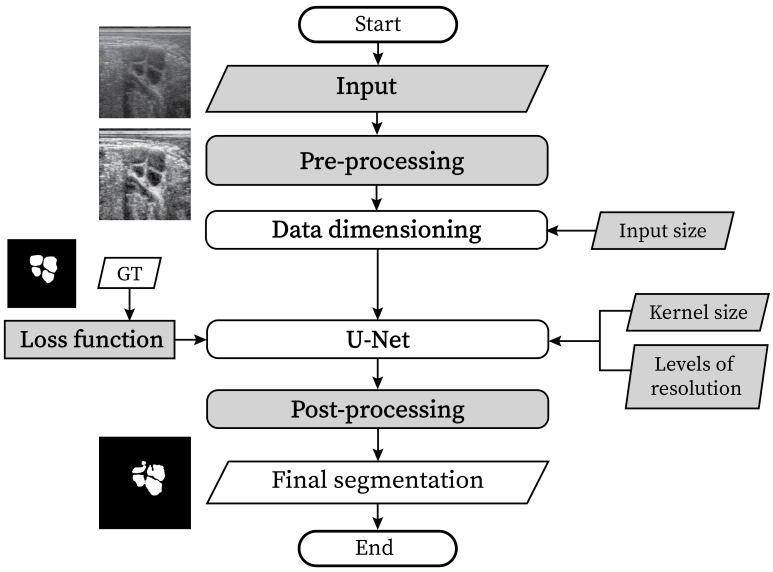


Figure 5.8: Schematic overview of the proposed methodology. The steps that comprise proposed experiments are presented in grey boxes.

With this in mind, changes in the several stages and additional steps are proposed in order to improve the performance of the baseline method. The overview of these additional options are outlined in Figure 5.8. The steps of this flowchart that comprise proposed experiments are the input data, pre-processing methods, data dimensioning (regarding the input size), U-Net architecture (regarding the kernel size and the number of levels of resolutions), the definition of the loss function and post-processing methods.

The proposed changes lead to a sequence of four experiments, described in the Sections 5.7.1 to 5.7.4.

5.7.1 Input Data and Pre-processing Methods

The input data and pre-processing stages are the coding blocks responsible for the type of information used by the U-Net for feature extraction.

The goal of the experiment is to assess the influence on the model's performance of the addition of BRF data and pre-processed B-mode data in the U-Net's input data.

To perform this experiment, the B-mode and BRF images are used as input data, as well as a combination of the two in the form of a 3D array.

The pre-processing methods are applied to the B-mode images. Two enhancement techniques (detailed in the following subsections) are applied to the B-mode images:

- Despeckling filter [74]
- Contrast Limited Adaptative Histogram Equalization (CLAHE)

Both techniques were applied to the 512×512 images separately and combined (the despeckling filter followed by CLAHE), resulting in three different pre-processing methods.

While BRF data is a more raw type of data, as it is not subjected to filtering, demodulation and logarithmic compression like B-mode data, pre-processed B-mode images consist on a more transformed type of data than the original B-mode images.

This way, the experiment tests data with a range of levels of processing, and so it is assessed the trade-off between the amount of noise in the image (more prevalent in BRF data) and data loss (possible consequence of BRF to B-mode conversion and of the application of pre-processing methods).

5.7.1.1 Despeckling Filter

Speckle noise is an artefact which affects US images, compromising the effectiveness of their analysis. The interference effects between returning echoes results in fine false structures whose apparent resolution is higher than the real acquisition of the imaging system [75].

A despeckling filter must be able to eliminate or fade the speckle, while still preserving relevant features such as edges. The implemented despeckling filter is described in [74]. This algorithm consists on a N×N kernel which slides over the image and uses the intensity of the pixels in that image patch (p_j for $j = 1,2,...,\mathrm{N}\times\mathrm{N}$) to update the central pixel (p_C). The first step is to compute a weight value for each pixel in the kernel:

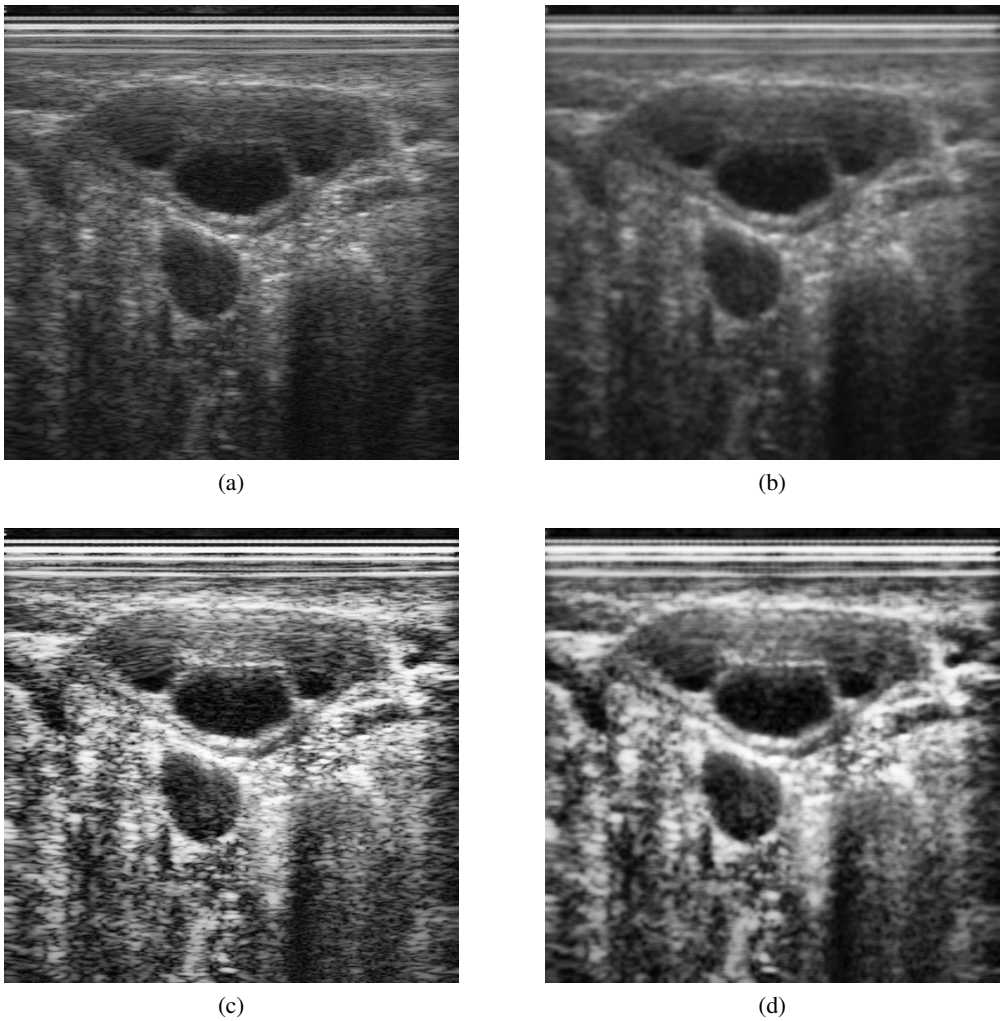
$$\begin{aligned} w_j &= \left(1-\left(\frac{i_{p_j}-i_{p_\mathrm{C}}}{\alpha\sigma}\right)^2\right)^2 \\ \text{if} \quad 1-\left(\frac{i_{p_j}-i_{p_\mathrm{C}}}{\alpha\sigma}\right)^2 &< 0, \qquad \text{then} \qquad w_j = 0 \end{aligned} \tag{5.2}$$

where i_{p_j} and i_{p_C} are the intensities of pixels p_j and p_C, respectively, σ is the standard deviation of the image intensity values, and α is a coefficient with an arbitrary value.

The updated intensity of p_C is given by the following:

$$i_{p_\mathrm{C}} = \frac{\sum_{j=1}^{\mathrm{N}\times\mathrm{N}} w_j i_{p_j}}{\sum_{j=1}^{\mathrm{N}\times\mathrm{N}} w_j} \tag{5.3}$$

For this method the filter hyperparameters (kernel size N and coefficient α) were determined empirically by testing the combination of a range of values for each. The tested values for N are the odd numbers ranging from 3 to 11 and for α were 1, 5, 10, 20, 30, 40. The kernel size



(a) (b)

(c) (d)

Figure 5.9: Example of effect of the despeckling filter and of CLAHE. (a) Original B-mode image; (b) Results of the despeckling filter; (c) Result of CLAHE; (d) Result of despeckling filter followed by CLAHE.

(N) was set to 7 and the coefficient α was set to 20 as they were the ones that, visually, led to a considerable noise reduction without losing the structure of the elements in the image such as edges. In the example of the result of the despeckling filter found in Figure 5.9.a and 5.9.b it is possible to verify the speckle noise is faded but the detail of the edges is preserved, which is specially noticeable for the separation between follicles and for the boundary of the ovary.

5.7.1.2 Contrast Limited Adaptive Histogram Equalization

Contrast Limited Adaptive Histogram Equalization (CLAHE) [76] is a contrast enhancement technique, used in this work to promote the detection of the structures' borders.

Histogram Equalization is a contrast enhancement algorithm which analyses the intensity histogram of an image in order to uniformly redistribute its pixel intensities and balance their proba-

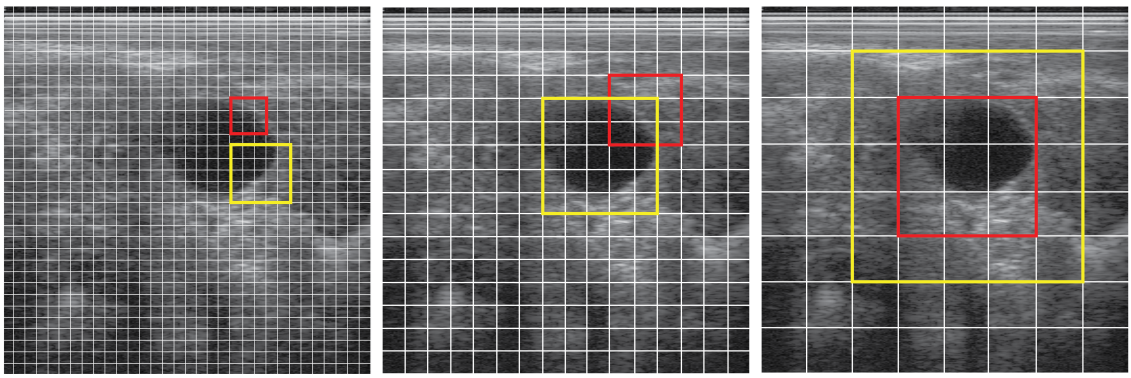
bility of occurrence. Due to the global analysis performed by the Histogram Equalization algorithm, if an image contains areas significantly lighter of darker than the majority of the image, the contrast in those areas is not sufficiently enhanced. Adaptive Histogram Equalization surpasses this limitation by performing a local analysis, that is, updating each pixel according only to their neighborhood and not the total image.

CLAHE is a more refined version of the Adaptive Histogram Equalization, which limits the amplification with a user-defined clip limit, to avoid overenhancement of noise and to reduce the edge-shadowing effect produced by unlimited Adaptive Histogram Equalization [76]. Hyperparameter combinations of several values for the clip limit (ranging from 2 to 7) and the grid size (ranging from 3 to 11) were tested to visually assess which combination produced the best trade-off between the enhancement of the edges and noise accentuation. After that, it was decided to set clip limit and grid size hyperparameters to 3 and 5×5, respectively.

An example of the result of the contrast enhancement using CLAHE is shown in Figure 5.9.

5.7.2 Data Dimensioning and U-Net Architecture Hyperparameters

In the analysis performed by a CNN such as the U-Net, it is important to consider the amount of image context that the architecture is able to apprehend throughout the feature extraction process.



(a) (b) (c)

Figure 5.10: Illustrations of the impact of size of the feature map in the lowest level of resolution and kernel size of U-Net convolutions, on the amount of context that the U-Net is able to apprehend. The white grid represents the number of pixels of the feature map in the lowest level of resolution and the colored squares represent the kernels of the convolution.

For the U-Net architecture, the ability to capture image context can be quantified through the ratio between the kernel size of its convolutions and the size of the feature map in the lowest level of resolution. This ratio correlates with the U-Net's ability to recognize the spatial relation between the several patches of input image (Figure 5.10).

For this experiment, the goal is to assess the impact on the model's performance of the implementation of a U-Net model with a higher ability to capture image context.

In order to increase the image context that is captured, there are three hyperparameters that can be changed, which are:

- Reduction of the input size;
- Increase of the kernel size of the convolutions employed by the U-Net;
- Increase of the depth of the architecture, that is, the number of levels of resolution.

It is hypothesised that an increase in the image context that is captured by the U-Net will allow the model to better perceive that there can only be one ovary in the each image and that the shape of each follicle and ovary is approximately round and convex.

The increase of the kernel size and depth of the U-Net leads to a higher consumption of GPU memory while the reduction of the input size represents a lower consumption. Taking this into account, and that the GPU memory available is limited, the combinations of these hyperparameters that are evaluated are shown in Table 5.1.

Table 5.1: Combinations of hyperparameters to study the influence of the amount of image context captured by the U-Net.

Input size	**Levels of resolution**	**Kernel size**	**Size at lowest level of resolution**	**Ratio**
512×512	5	3×3	32	0.093
320×320	5	3×3	20	0.150
256×256	5	3×3	16	0.180
320×320	6	3×3	10	0.300
256×256	5	5×5	16	0.313
256×256	6	3×3	8	0.375

In the case of a input size lower than 512×512, the final step of the model is a resize operation from the input size to 512×512.

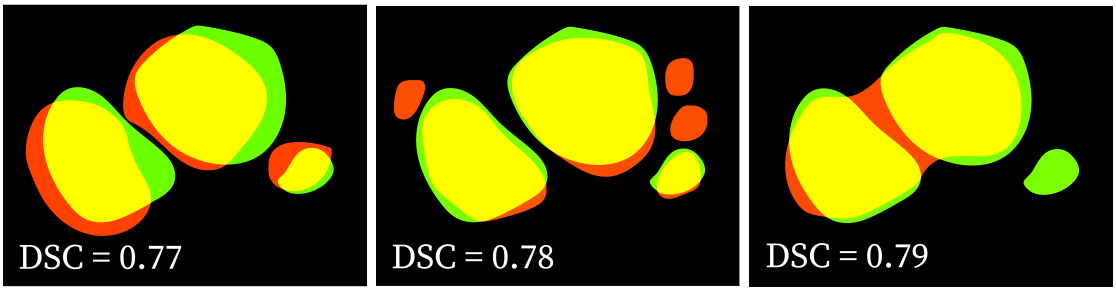
5.7.3 Loss Function

The loss function used for the baseline method is based on the DSC (DSC Loss $= 1 - \text{DSC}$). This is a metric of spatial overlap between two segmentations, defined as $\text{DSC}_{(\text{A,B})} = 2(\text{A} \cap \text{B})/(\text{A} + \text{B})$, where A and B are the two compared segmentations.

The disadvantage of using DSC to compute the loss in cases of multi-object segmentation is that, if two objects are very close to each other, the penalization of incorrectly predicting the background space between them is small. This is illustrated Figure 5.11, which shows how different segmentation errors influence the DSC.

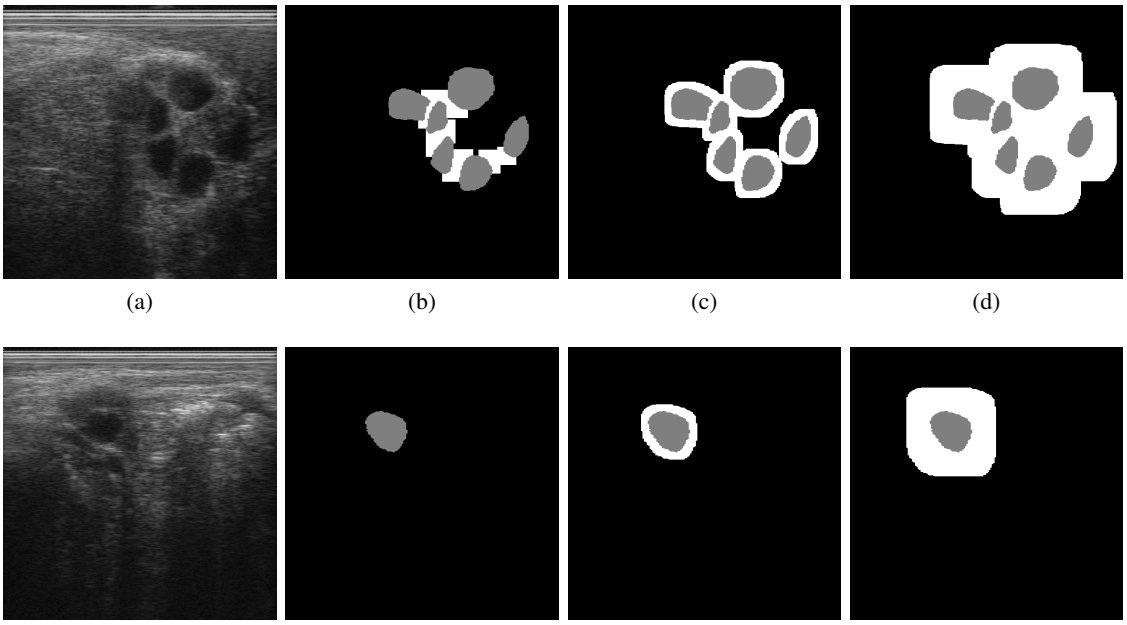
DSC is useful to penalize big FP areas, but a slightly higher TP area will compensate for a the FP area that corresponds to the incorrect union of two follicles. Also the penalization of small FN areas is low.

The goal of this experiment is to test a different definition of a DSC based loss function for follicle segmentation, which attributes a high penalization weight to the area near follicles.



(a) (b) (c)

Figure 5.11: Illustrations of the influence on DSC of different segmentation errors (compared with the same segmentation reference). (a) Correct object count, with considerable distance between borders; (b) Incorrect object count due to FP objects, but minor distance between borders; (c) Incorrect object count due to incorrect union of objects and FN object, but minor distance between borders. The TP area is in yellow, FP area is in red and FN area is in green.



(a) (b) (c) (d)

(e) (f) (g) (h)

Figure 5.12: Examples of the produced masks with the areas defined as in between or near follicles in white, and corresponding B-mode images. (a) and (e) B-mode images; (b) and (f) Loss1; (c) and (g) Loss2; (d) and (h) Loss3. The masks have three distinct areas: follicle area (gray), high penalization background area (white) and low penalization background area (black).

To compute the new loss function, it was necessary to produce three different masks for the definition of the area near follicles - masks Loss1, Loss2 and Loss3 (Figure 5.12).

The algorithm to produce mask Loss1 begins with a morphological closing operation with a kernel size of 50×50 applied to the original GT image, to the detect the areas between two neighbouring follicles, then the original GT of the follicles is subtracted. It is necessary to apply a morphological opening operation to the result of the subtraction in order to eliminate small

objects outlined by the closing operation that are detected, not because they belong to the area between follicles, but due to the concave shape of some follicles. Then, a dilation operation with a kernel size of 30×30 is applied to make the area between follicles more significant. This result is combined with the original follicle GT to produce the final mask.

Masks Loss2 and Loss3 result from the application of a morphological dilation on the GT images, one with a kernel size 30×30 and 100×100, respectively, and followed by the subtraction of the original GT, resulting in the high penalization areas around all follicles, including the isolated ones. Then the results of the morphological dilations are combined with the original follicle GT to produce the final masks. These two definitions for the area near follicles were tested to evaluate the influence of the size of the high penalization background area.

Loss1 was tested to analyse how the U-Net behaves in the case of a more inconsistent area definition, given that, by this definition, no area around isolated follicles is particularly penalized and in the follicles with close neighbours, only the area around the border that is close to the neighbour is included in the high penalization area. Although this definition is more accurate in the delineation of the area that is necessary to penalize to prevent the incorrect union of neighboring follicles, the penalized area is less consistent in terms of the extracted features that the model can associate with it. For instance, with Loss2 the model should easily associate the high penalization area with the areas near all borders of the follicles, which have features in common such as the intensity gradient, and consequently the model will retain that it is important to predict these areas as negative. On the other hand, with Loss1, these features are not always associated with the high penalization area.

For the computation of the loss function, it was made the distinction between the low penalization background area (*LB*), the high penalization background area (*HB*) and each one of the N follicles (*F*) present in the mask. A separate DSC score was computed for each area (*LB*, *HB* and N ×*F*) and a weight (*w*) was attributed to each. The definition of the loss function is as follows:

$$\begin{aligned}\text{New Loss} = 1 - \Big(&\frac{2\text{TN}_{LB}}{2\text{TN}_{LB}+\text{FP}_{LB}} * w_{LB} + \frac{2\text{TN}_{HB}}{2\text{TN}_{HB}+\text{FP}_{HB}} * w_{HB} + \\ &+ \sum_{i}^{\text{N}} \frac{2\text{TP}_{F_i}}{2\text{TP}_{F_i}+\text{FN}_{F_i}} * ((1 - w_{LB} - w_{HB})/\text{N})\Big)\end{aligned} \tag{5.4}$$

Two versions of the weight hyperparameters were defined - A and B:

- Parameters A: $w_{LB} = 0.54$ $w_{HB} = 0.26$;
- Parameters B: $w_{LB} = 0.63$ $w_{HB} = 0.17$.

These two versions are tested to assess the sensibility of this loss function to the weight hyperparameters. Both hyperparameter versions attribute 80% of weight to the overall background area, which is less than the mean percentage of background area on the overall image. This leads to a higher weight for the follicle area. The fact that for the 20% weight for the follicle area is equally distributed by the number of follicles in the images leads to a higher penalization in the case of missing small follicles than what would be penalized with the general DSC loss of the

baseline method. These two versions result in two different degrees of penalization, in which hyperparameters A are more penalizing for the incorrect predictions in the area between of near follicles.

For Loss3, in the cases of images only with isolated follicles, for which no HB is defined, the background weight is equal to the percentage of its area in the image (w_B) and the follicles' weight is equal to $(1 - w_B)/\mathrm{N}$.

Given that this definition of loss function attributes less weight to most of the background area (LB), it is expected that the produced predictions have a higher number of FP objects in that area. If the results verify the tendency for the separation of neighboring follicles, the predictions of this loss function can be used for a post-processing method, detailed in the following subsection.

5.7.4 Post-processing Methods

Post-processing methods can be applied to the outcome of the U-Net in order to minimize some of the errors made by the deep learning algorithm.

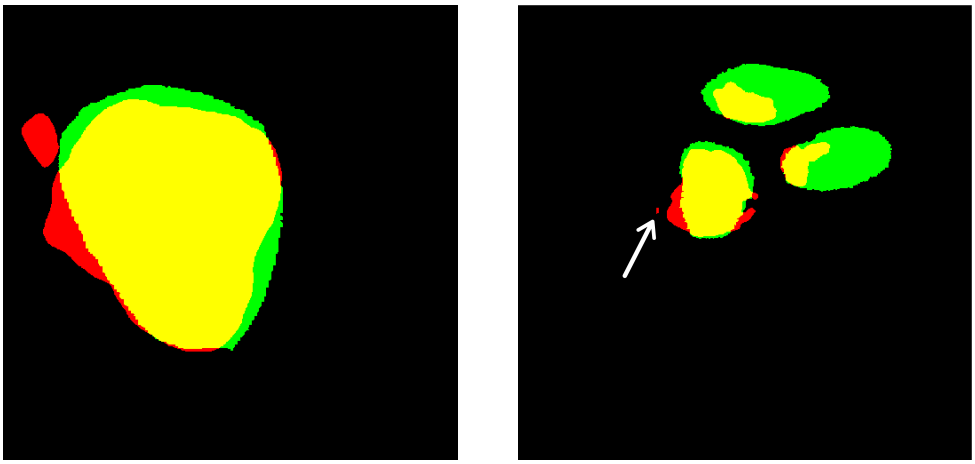
Two of the most common errors found in the outputs of the U-Net is the presence of small objects which correspond to false positives and the incorrect union of two neighboring follicles.

The goal of this experiment is to select the best post-processing methods or combination of methods, to improve the performance for ovary and follicle segmentation and detection by correcting the aforementioned errors.

To summarize, the proposed post-processing methods are:

- For ovary segmentation:
  - Method Selection - Selection of the largest segmented object
- For follicle segmentation:
  - Method Removal - Removal of objects smaller than 70 px;
  - Method Watershed 1 - Watershed algorithm, with a percentage of 30% for the selection of the threshold;
  - Method Watershed 2 - Watershed algorithm, with a percentage of 50% for the selection of the threshold;
  - Method Watershed 3 - Watershed algorithm, with a percentage of 70% for the selection of the threshold;
  - Method Out of Ovary - Removal of objects outside of the ovary area (using the prediction of the ovary after the implementation of Method Selection);
  - Method Separation - Use of the predictions of the model with the proposed loss function to separate incorrect unions of neighboring follicles.

The Method Selection, for ovary segmentation, is based on the *a priori* that only one object should be predicted. Given that in the case of the incorrect prediction of an additional object (Figure 5.13.a), its area is unlikely to surpass the area of the object that corresponds to the true ovary, it is proposed to select only the object with the bigger area as the predicted ovary.



(a) (b)

Figure 5.13: Example of appropriate cases for the application of (a) Method Selection; (b) Method Removal, with a small FP object with 52 pixels (white arrow). The TP area in yellow, FN area in green and FP in red)

The Method Removal, for follicle segmentation, the smallest object delineated by the annotators has an area of 87 pixels (in a GT image of 512×512), but in some of the U-Net predictions are found objects as small as 3 pixels (Figure 5.13.b). Based on this, it is proposed the removal of objects with areas bellow 70 pixels of area, in order to improve the results regarding follicle count.

As mentioned before, the U-Net optimized with the loss function of the baseline model tends to predict incorrectly the union of neighbouring follicles but in most of the cases it is possible to notice the delineation of each individual follicle. In the cases in which the union of two neighboring follicles is only small compared to the size of the follicles (as shown in Figure 5.14.b), the Watershed algorithm [77] is able to detect the separation between them.

For the Method Watershed, the implemented watershed algorithm begins with the computation of the distance transform for the prediction image, which results in a image representation of the distance to the closest boundary from each point (Figure 5.14.c). A threshold is applied to the distance transform to determine the area that will be defined as admitted to belong to the foreground. This threshold is a percentage of the maximum value of the distance transform of the image. In the case of images with big objects (high maximum value of the distance transform), the threshold would tend to discard smaller objects. In order to avoid that problem, the threshold is calculated separately for each object in the original prediction. The percentage used to select the threshold is a crucial hyperparameter and so different values are tested, namely 30%, 50% and 70% (Methods Watershed 1-3).

Then to determine the area of the image to consider as undoubtedly belonging to the background, two iterations of a morphological dilation with a 5×5 kernel are applied to the prediction and area outside of the resulting objects is selected. Besides the areas that are assumed to belong to the background or foreground, remains a so called 'unknown' area (Figure 5.14.d) whose medial axis will constitute the new borders of predicted segmentation.

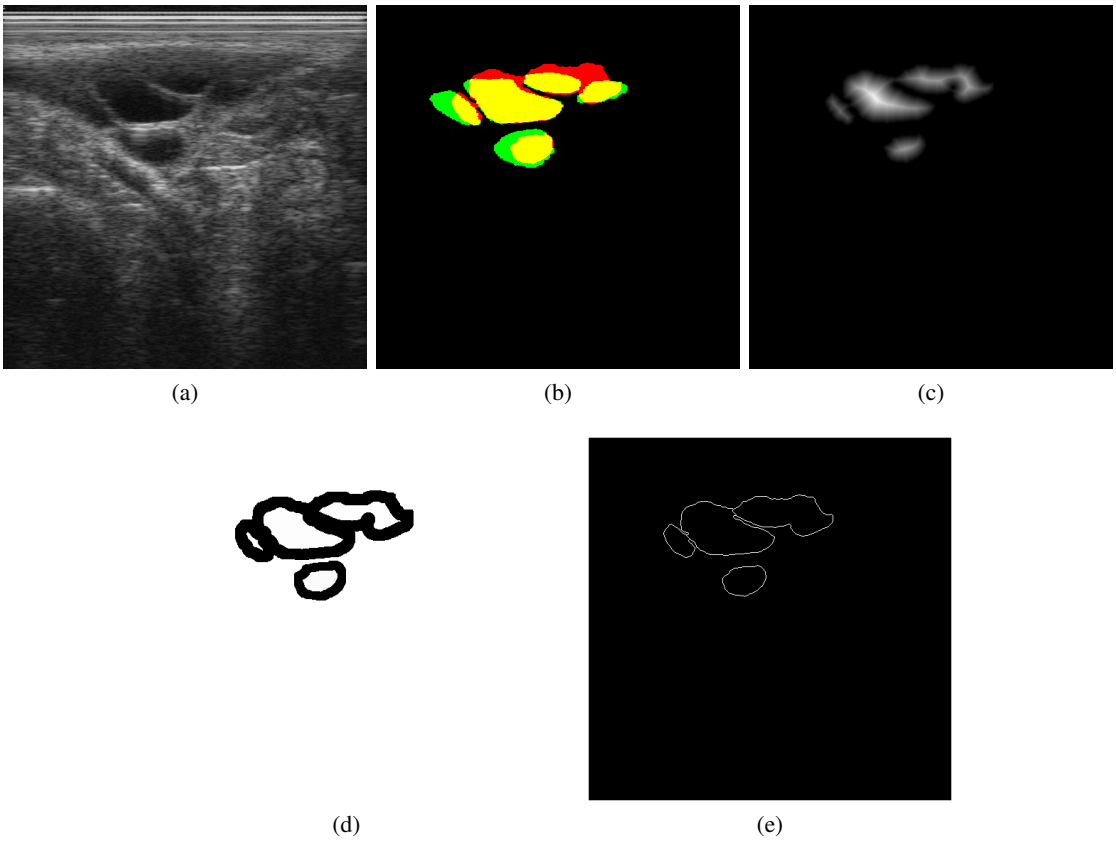


Figure 5.14: Example of the result of the Watershed algorithm. (a) B-mode image (b) Example of follicle segmentation, with TP area in yellow, FN area green and FP area in red; (c) Distance transform of the segmentation; (d) In white is the combination of the areas detected as being undoubtedly foreground and background and in black is the area called 'unknown'. (e) Delineation result of Watershed algorithm.

The algorithm developed for the Method Out of Ovary eliminates objects with more than 50% of their area outside the corresponding prediction of the ovary.

Lastly, Method Separation separates incorrect unions of neighboring follicles in a follicle prediction (A) using the predictions produced by the model with the loss function proposed in Section 5.7.3 (B). The implemented algorithm is the following: the objects in A that have any overlap with more than one object in B are deleted and replaced with the corresponding objects of B.

5.8 Performance Measures

This section presents the methodology used to compare the performance of the methods of the proposed changes to the baseline method and also to validate the final method.

For analysis of the results, measures regarding the segmentation (pixel-wise analysis) and detection (object-wise analysis) performance are computed, along with an adapted version of the

metric proposed by [78] which is a Partition Distance measure that combines pixel-wise and object-wise analysis.

Given that the implemented models use 5-fold cross-validation, it is detailed the methodology utilized to combine the results of each fold. Additionally, it is described the approach used for the comparison between the performance of the methods of the proposed experiments and the baseline method, along with the explanation of the interpretation of the boxplot, which is the graphic tool used to display the results of each method.

5.8.1 Segmentation Measure

The measure used for the pixel-wise analysis is the DSC. This measure is able to numerically represent the overall pixel-wise intersection between the prediction and the GT but does not report information about how exact is the object count of the prediction because, as mentioned before, DSC entails little penalization in the cases incorrect union of neighboring follicles and incorrect prediction of small objects.

5.8.2 Detection Measure

For the object-wise analysis, it is indicated the performance of the method regarding the object count.

The number of TP/FP/FN objects in each prediction is computed based on the method applied in [72]. The correspondence between the objects in the prediction and in the GT image, is determined based on a ratio (r) computed for a every pair of intersecting objects:

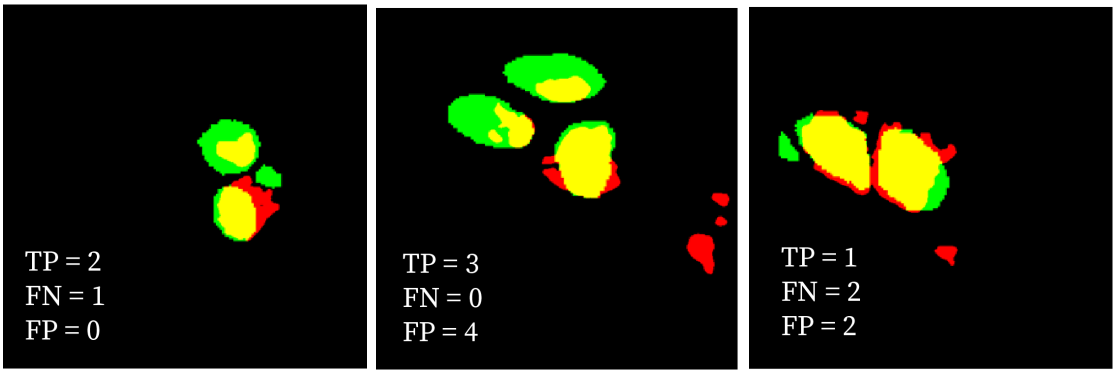
$$\mathrm{r} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}} \times \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}} \tag{5.5}$$

where TP, FN and FP values result from the pixel-wise comparison and are based only on the comparison of intersecting objects in question.

The pairs of intersecting objects with higher r are associated, resulting in the count of TP objects, while the remaining unpaired objects of the prediction result in the count of FP objects and the remaining unpaired objects of the GT image result in the count of FN objects. Examples of the count of TP, FN and FP objects are shown in Figure 5.15.

From the count of TP, FN and FP objects, the value of the Accuracy (ACC), Precision and False Negative Rate (FNR) for the objects-wise analysis are determined. It is important to note that for the computation of the ACC value, the TN objects are not taken into account, as it is equal to 1 in all the predictions:

$$\mathrm{ACC} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}+\mathrm{FP}} \tag{5.6}$$



(a) (b) (c)

Figure 5.15: Illustrations of the computation of the TP, FN and FP objects for the object-wise analysis.The TP area is in yellow, FP area is in red and FN area is in green..

For the case of ovary prediction, the Precision and FNR are not computed, given that it consists on a single-object problem and so ACC is sufficient to represent the errors of the prediction.

5.8.3 Partition Distance Measure

The Partition Distance measure (PD) proposed by Cardoso *et al.* in [78] compares the segmentation prediction and the GT as a partition problem and so it combines the penalization for incorrect object count and for the amount of overlap between corresponding objects in the GT and in the prediction.

The distance between two partitions (d) consists on the minimum number of pixels that need to be deleted for the two partitions to be equal. To normalize this distance, the author proposes to divide d by the total number of pixels in the image. Given that, for this work, the background area is significantly larger than the foreground area, the value of the normalized distance would be have a small magnitude. Hence, the implemented definition is the following:

$$\mathrm{PD} = \frac{d}{\mathrm{TP} + \mathrm{FN} + \mathrm{FP}} \tag{5.7}$$

5.8.4 Model Validation and Comparison

To combine the performance across the 5 folds of cross-validation for each method, regarding each of the aforementioned measures, it is computed the mean measure value of the 5 predictions for each image in the test-set.

The boxplot [79] is the graphic tool used to display the results of the compared methods (in the experiments regarding the proposed changes). This tool represents the results of a group or groups of numerical data through their medians, quartiles, and minimum and maximum values (Figure 5.16). In this representation the Interquartile Range (IQR) corresponds to the difference

[1] http://www.physics.csbsju.edu/stats/box2.html

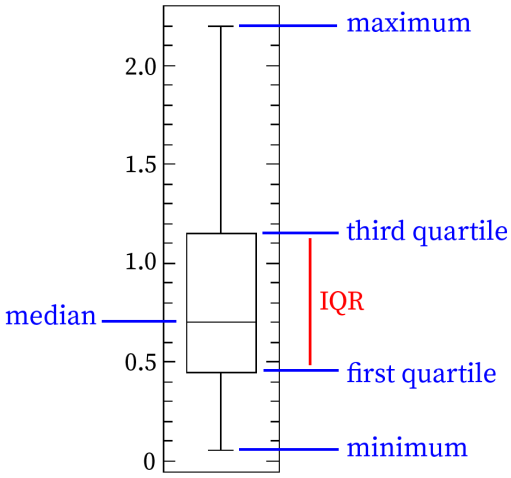


Figure 5.16: Illustration of boxplot interpretation[1].

between the third quartile and the first quartile. In the case of the presence of outliers, they are represented as small circles or dots above or below the plot. A data point is considered an outlier if its value is $1.5 \times$ IQR or more above the third quartile or if it is $1.5 \times$ IQR or more below the first quartile.

To assess the significance of the performance improvement in DSC (pixel-wise analysis) and ACC (object-wise analysis), a paired two tailed t-test analysis is performed, being the null hypothesis that the mean performance of the compared methods is equal. The alpha level for the tests was set at 0.05. This assessment is done for the experiments regarding the proposed changes used to select characteristics for the preliminary and the final method, namely the input data, preprocessing methods, input size, U-Net hyperparameters, the definition of the loss function and post-processing methods. The other performance metrics (Precision, FNR and PD) are used to complement the discussion of the performance analysis.

Chapter 6

Results and Discussion

This chapter presents the results for follicle and ovary segmentation and detection obtained through the conducted experiments (Sections 5.7.1 to 5.7.4) and for the final selected method.

The results of the experiments for the method comparison are presented and discussed, and based on that the characteristics for a preliminary method are selected (namely the input data, pre-processing methods, input size, U-Net hyperparameters and the definition of the loss function).

The results of the preliminary model are obtained for the total 107 images of the dataset, using data augmentation. The predictions of the preliminary model are then subjected to the proposed post-processing methods. According to the performance after post-processing, it is decided which post-processing methods to integrate in the preliminary method, completing the selection of characteristics for the final method.

The figures used in this chapter to illustrate the prediction images and their comparison with the GT use the following color scheme: the TP area is in yellow, FP area is in red and FN area is in green.

6.1 Input Data and Pre-processing Methods

The results of the models with different input data and that implement the proposed pre-processing methods (shown in Figure 6.1) and the correspondent t-test analysis show that there is no significant improvement for the proposed models, for follicle segmentation and detection when compared with the baseline model.

This leads to conclude that neither the additional information introduced by the use of BRF data or the data transformation resulting from the pre-processing methods are valuable for the feature extraction process in the case of follicles.

Hence, for the preliminary and the final model for follicle segmentation and detection, the input data will consist of B-mode images without the implementation of pre-processing methods.

On the other hand, the results of ovary segmentation and detection of the models with different input data and that implement the proposed pre-processing methods (shown in Figure 6.2) and the correspondent t-test analysis show significant improvement in ACC for the models using B-mode

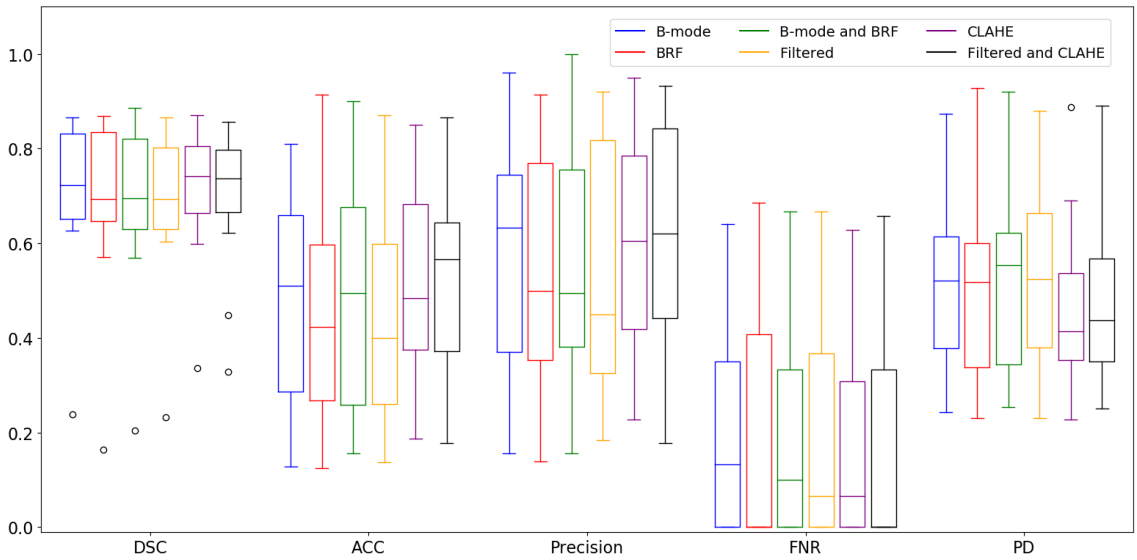


Figure 6.1: Boxplot of the results of the experiment regarding input data and pre-processing methods for follicle segmentation.

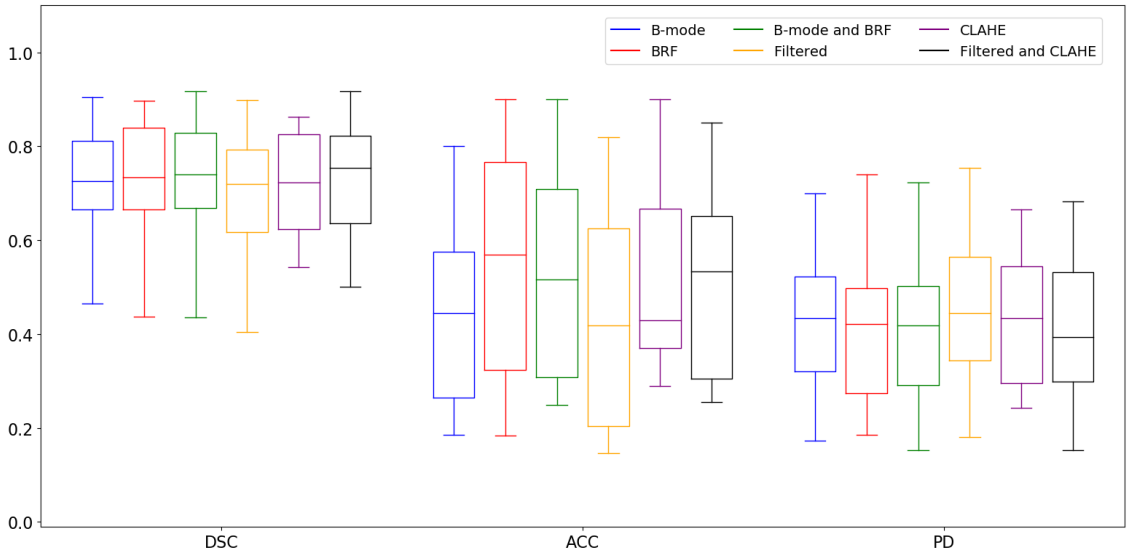


Figure 6.2: Boxplot of the results of the experiment regarding input data and pre-processing methods for ovary segmentation.

combined with BRF and also models using B-mode images pre-processed with CLAHE, when compared with the baseline model. The ovary segmentation and detection results regarding DSC show no significant improvement for ovary segmentation.

Given that the stroma tissue reveals its texture in US images (unlike the follicles which correspond to hypoechoic areas), the additional information introduced by the use of BRF data and the contrast enhancement resulting from the CLAHE are valuable for the feature extraction process in the case of the ovary.

For the preliminary and the final model of ovary segmentation, the input data will consist of

B-mode images pre-processed using CLAHE. CLAHE was chosen rather than B-mode and BRF due to the lower complexity of the model when using only one type of image as input which is advantageous as entails a lower memory cost.

6.2 Data Dimensioning and U-Net Architecture Hyperparameters

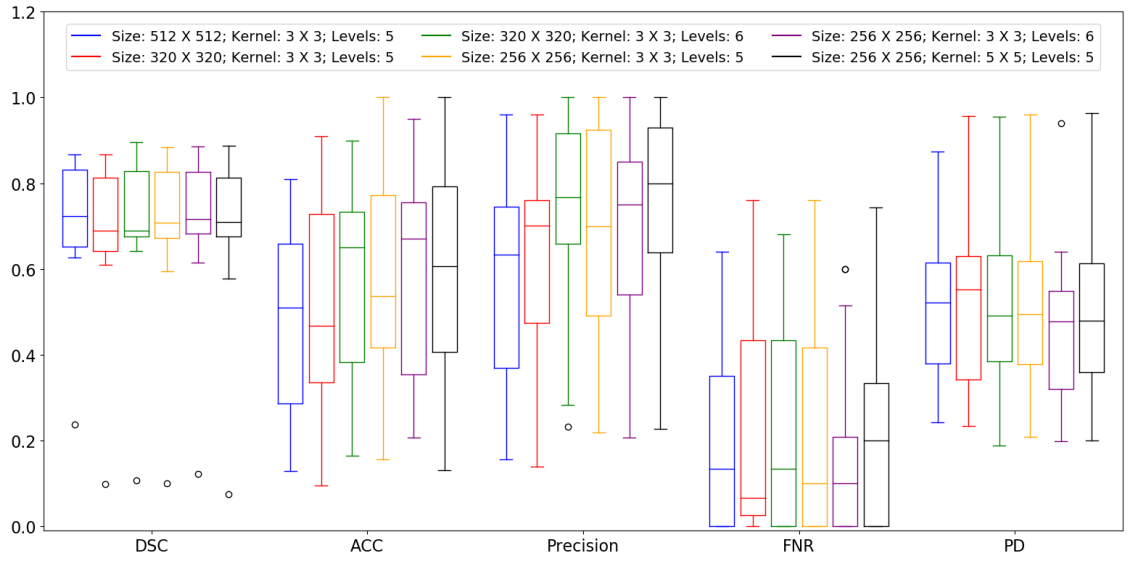


Figure 6.3: Boxplot of the results of the experiment regarding input size and U-Net hyperparameters, for follicle segmentation.

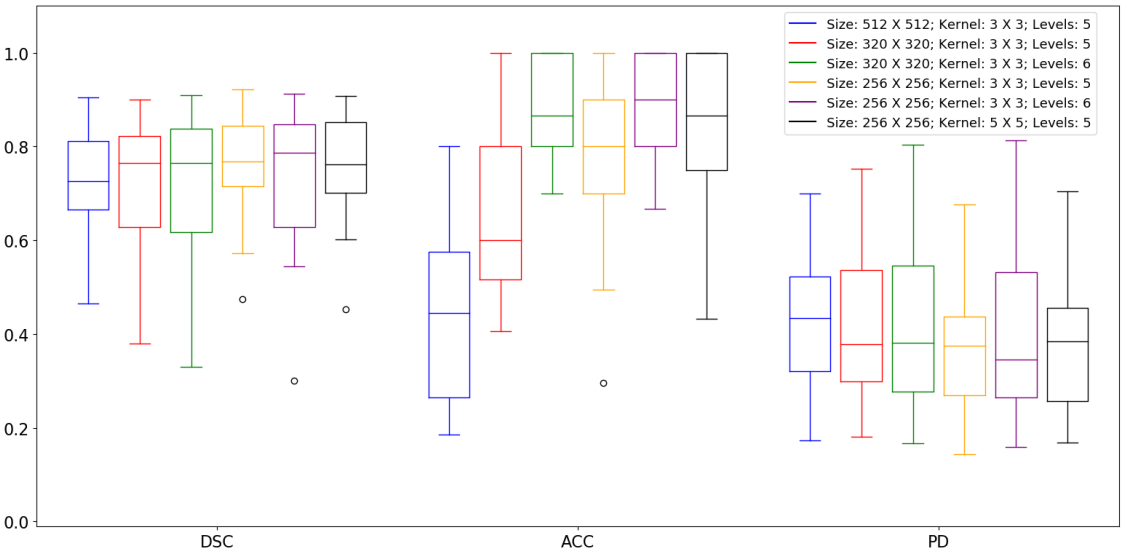


Figure 6.4: Boxplot of the results of the experiment regarding input size and U-Net hyperparameters, for ovary segmentation.

The results of the models with different values for the hyperparameters related to image context analysis (Figures 6.3 and 6.4) show that there is significant improvement for some of the proposed

models when compared to the baseline model, for ovary and follicle segmentation. It is important to recall that the hyperparameters of the baseline model are:

- Input size: 512×512; Kernel: 3×3; Levels of resolution: 5;

In the case of follicle segmentation, there was no significant improvement in performance for DSC. For the performance in follicle detection, the ACC performance showed significant improvement when using the following models:

- Input size: 320×320; Kernel: 3×3; Levels of resolution: 6;
- Input size: 256×256; Kernel: 3×3; Levels of resolution: 5;
- Input size: 256×256; Kernel: 3×3; Levels of resolution: 6;
- Input size: 256×256; Kernel: 5×5; Levels of resolution: 5;

For the same models, the Precision also increases while the FNR and PD are similar to the baseline model. This shows that the improvement in ACC is attributable to the decrease in the number of FP objects in the predictions.

In the case of ovary segmentation, the comparison results are similar to follicle segmentation, as there was no significant improvement in performance for DSC, except for model - Input size: 256×256; Kernel: 3×3; Levels of resolution: 5. Regarding ACC values, for the performance in ovary detection, there was significant improvement when using all of the proposed models. The improvement of ACC values for ovary segmentation is due to the decrease in FP objects, given that for all of the predictions the number of TP objects was equal to 1 and the number of FN objects was equal to 0.

The improvement in object detection measures, both for follicles and ovary, when the ratio between kernel size of the U-Net's convolutions and the size of the feature map in the lowest level of resolution increases leads to conclude that the amount of context that the U-Net is able to capture throughout its scanning is important. By analysing the image more globally during training the model better recognizes the difference between follicles and other hypoechoic structures and also tends to better apprehend the fact that the ovary prediction should consist of only one object.

Hence, the final model for follicle and ovary segmentation will have the following hyperparameters regarding data dimensioning and U-Net architecture:

- Input size: 256×256; Kernel: 3×3; Levels of resolution: 6;

This set of hyperparameters was chosen rather than the others which also showed improved performance because it is the one that has the higher ratio between the kernel size of the U-Net's convolutions and the size of the feature map in the lowest level of resolution, which entails a higher amount of context captured by the U-Net, as explained in Section 5.7.2. Also, this model showed a particular improvement in the FNR for follicle segmentation and in ACC for ovary segmentation.

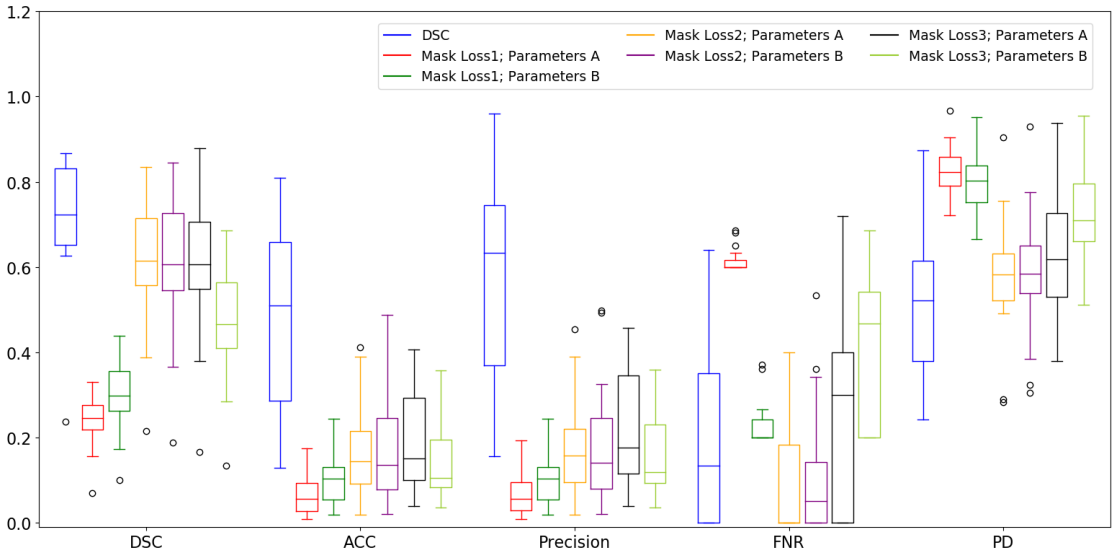


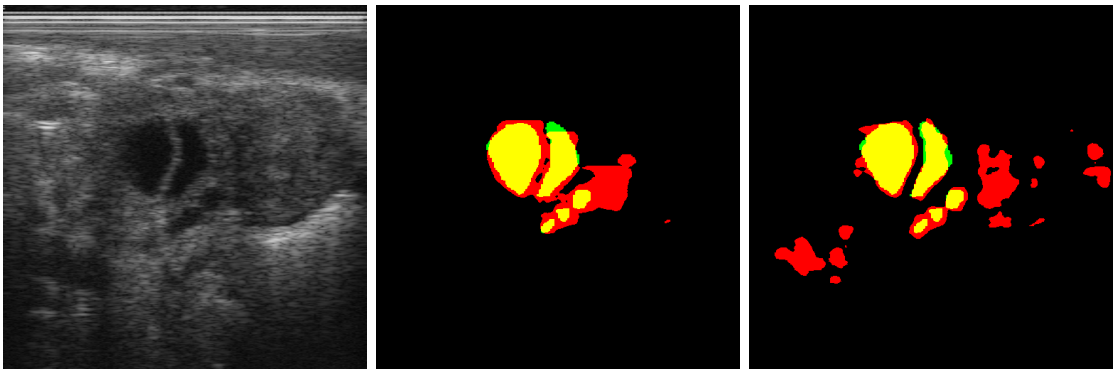
Figure 6.5: Boxplot of the results of the experiment regarding the proposed definition for the loss function, for follicle segmentation. DSC in the legend of the figure corresponds the results of the baseline model.

6.3 Loss Function

The results of the experiment regarding the implementation of the proposed loss function definition and its comparison with the baseline model are presented in Figure 6.5. The results of the models implementing each combination of types of masks for the definition of the area with higher penalization (Loss1, Loss2 and Loss3) and of the weight hyperparameters (A and B) are compared with the results of the baseline model.

These results show inferior performance resulting from the implementation of the proposed loss function definition. Therefore, it demonstrates that the proposed loss function is not suitable for follicle segmentation. Despite this, among the models implementing the proposed loss function, the ones using mask Loss2 were the ones that produced better results, both in terms of follicle segmentation and detection.

The proposed loss function attributes less weight to the background area of the mask (*LB*) than the baseline loss function, and so it leads to a higher number of FP objecs in that area. However, the improvement in FNR and a qualitative analysis of the predictions produced by the proposed loss function (with Mask Loss2), confirm that this loss results in the intended tendency for the separation of neighboring follicles, as shown in Figure 6.6. In this Figure, is it possible to verify that, although the FP area is bigger, the delineations around the follicles are more precise. For this reason, the predictions from the model using the proposed loss function with Masks Loss2 and Parameters A will be used in the post-processing method detailed in Subsection 5.7.4. The choice between Parameters A and B is arbitrary as there are no significant difference between the results of the model "Mask Loss2; Parameters A" and of the model "Mask Loss2; Parameters B".



(a) (b) (c)

Figure 6.6: Example of the results of the experiment regarding the proposed loss function. (a) B-mode image (b) Result of the baseline model; (c) Result of the model using the proposed loss function with Mask Loss2 and Parameters A.

6.4 Preliminary Model and Post-processing Methods

Based on the previous experiments, the selected preliminary model for the segmentation of the follicles and ovary has the following characteristics:

- Input data and pre-processing methods:
  - For follicle segmentation: B-mode;
  - For ovary segmentation: B-mode pre-processed using CLAHE;
- Input size: 256×256;
- U-Net hyperparameters:
  - Kernel: 3×3;
  - Levels of resolution: 6;
- Loss function: $\text{Loss} = 1 - \text{DSC}$

To the results of this model are applied the post-processing methods detailed in Subsection 5.7.4. The model used to obtain the predictions necessary for the implementation of Method Separation (with Mask Loss2 and Parameters A) has the same aforementioned characteristics.

In Figure 6.7, the results of the proposed post-processing methods are shown.

The results lead to conclude that the watershed algorithm (Method Watershed1-3) leads to a significant improvement in ACC, due to the separation of incorrect unions of neighboring follicles and consequent decrease in the number of FN objects. This is particularly evident with a percentage of 50% for the selection of the threshold (Method Watershed2), indicating that 50% is the most appropriate value for this parameter.

A similar effect is produced by the implementation of Method Separation, for the same reason: separation of incorrect union of neighboring follicles and consequent decrease in the number of FN objects. An example of the application of this method is shown in Figure 6.8.

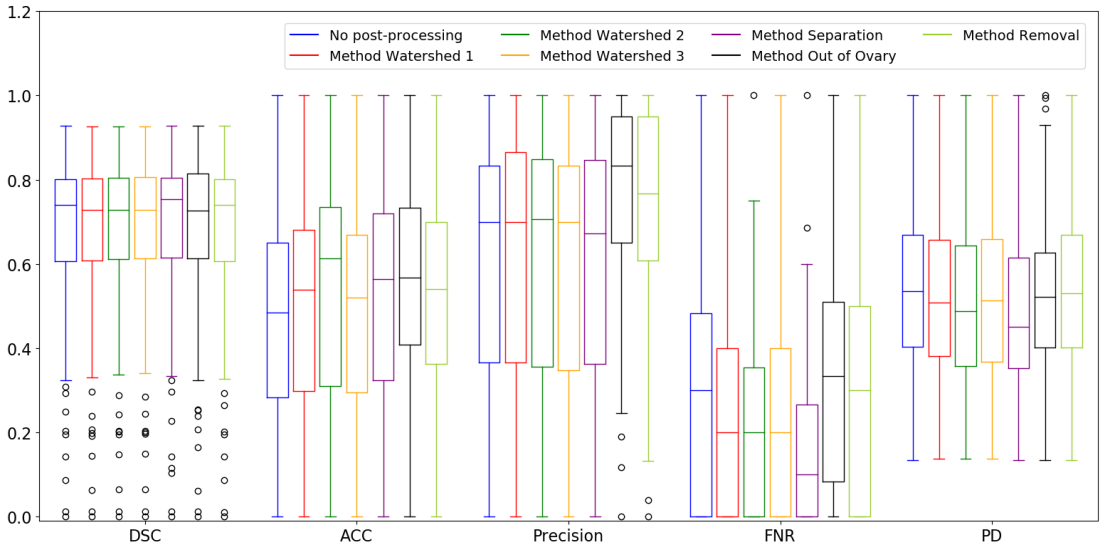
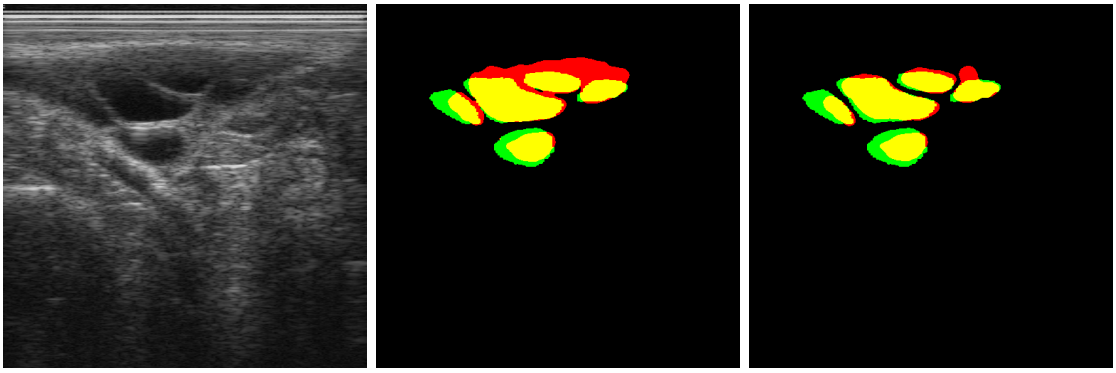


Figure 6.7: Boxplot of the results of the experiment regarding the post-processing methods, for follicle segmentation.



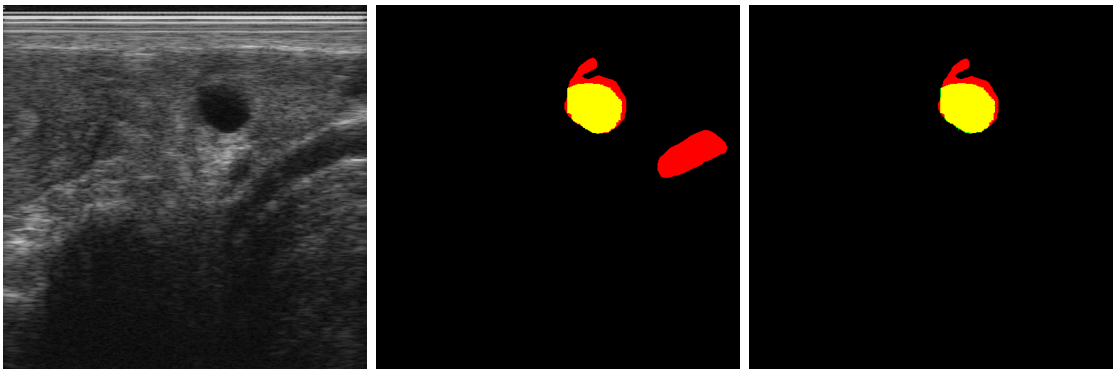
(a) (b) (c)

Figure 6.8: Example of the results of the implementation of the Method Separation. (a) B-mode image; (b) Result from the baseline model; (c) Results after implementation of the Method Separation.

The Method Out of Ovary results in a significant improvement in ACC, which leads to conclude that the removal of objects outside of the corresponding prediction of the ovary does decrease the number of false positive objects in the follicle prediction. An example of the application of this method is shown in Figure 6.9.

The Method Removal produces the same effect but in this case by eliminating objects that corresponded to errors in the prediction.

Finally, the results of the implementation of Method Selection are shown in Figure 6.10. The ACC significantly improves after the implementation of this method, as expected. The success of this method has only exceptions for the predictions of five images (overlaped outliers in ACC in Figure 6.10) because, in the prediction of one of their 5 folds, the biggest predicted object did not



(a) (b) (c)

Figure 6.9: Example of the results of the implementation of the Method Out of Ovary. (a) B-mode image; (b) Result from the baseline model; (c) Results after implementation of the Method Out of Ovary.

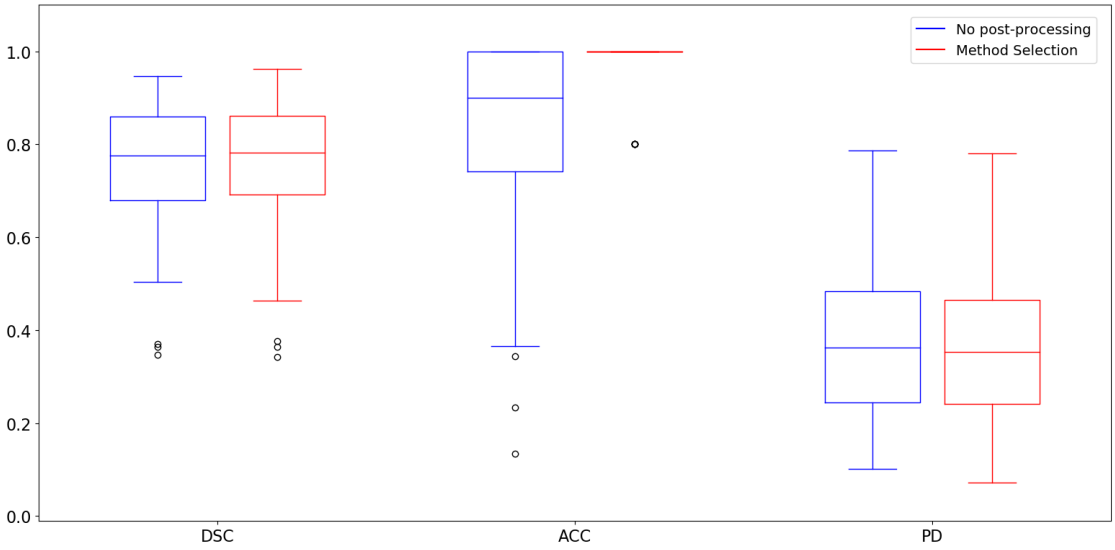
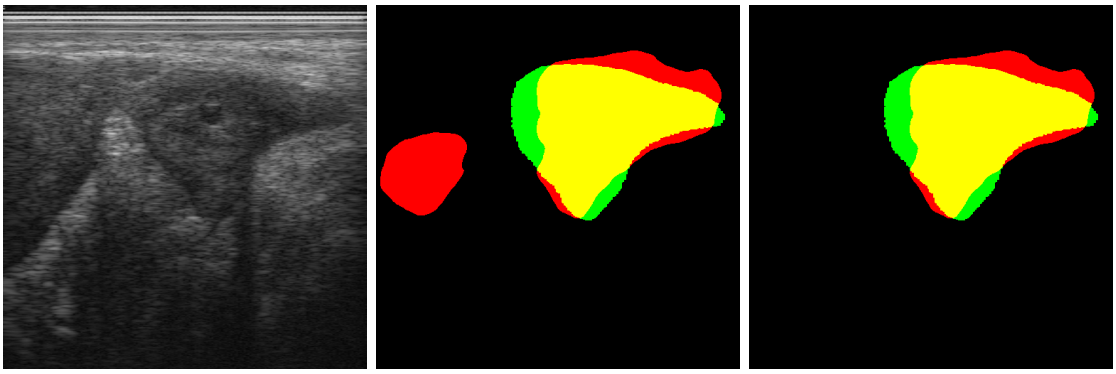


Figure 6.10: Boxplot of the results of the experiment regarding the post-processing method, for ovary segmentation.

correspond to the ovary. An example of the application of this method is shown in Figure 6.11.

From these results, it was decided to implement Method Selection to the ovary predictions of the final model. To the follicle predictions of the final model, it was decided to implement a sequential combination of these methods: Method Watershed2, Method Separation, Method Removal and Method Out of Ovary. A combination of the methods was selected as they are complementary to each other.



(a) (b) (c)

Figure 6.11: Example of the results of the implementation of the Method Selection. (a) B-mode image; (b) Result from the baseline model; (c) Results after implementation of the Method Selection.

6.5 Final Model

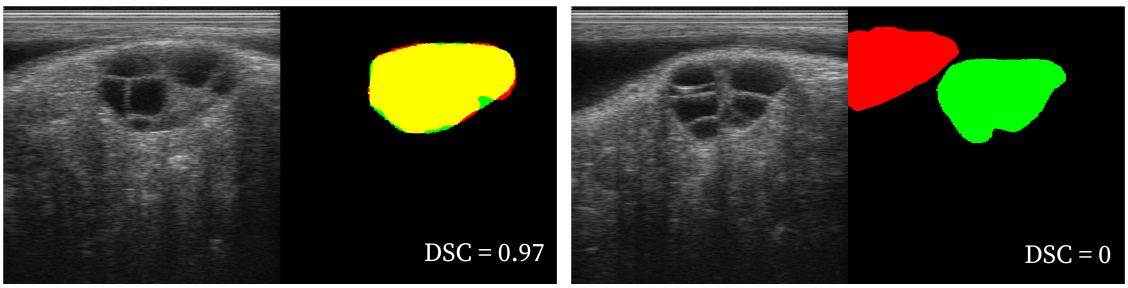
Based on the results of all the previous experiments, the final model for the segmentation of the follicles and ovary has the following characteristics:

- Input data and pre-processing methods:
  - For follicle segmentation: B-mode;
  - For ovary segmentation: B-mode pre-processed using CLAHE;
- Input size: 256×256
- U-Net hyperparameters:
  - Kernel: 3×3
  - Levels of resolution: 6
- Loss function: Loss $= 1 - \text{DSC}$
- Post-processing methods:
  - Ovary: Method Selection
  - Follicles: sequential combination of Method Watershed 2, Method Separation, Method Removal and Method Out of Ovary

The results of the final model regarding ovary segmentation are:

- DSC: 0.76 ± 0.13
- ACC: 0.99 ± 0.04
- PD: 0.37 ± 0.16

The images in Figure 6.12 are examples of the best and worst predictions for the ovary segmentation, and actually the two images are from the same patient. In the worst prediction there



(a) (b)

Figure 6.12: Examples of (a) the best DSC results and (b) the worst DSC results, for ovary segmentation.

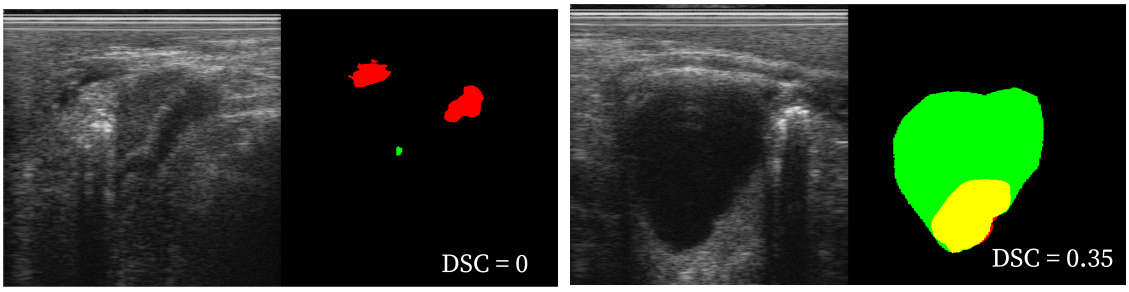
is a considerably large hypoechoic structure besides the follicles, and although the model also segmented the correct object as ovary, the object selected by the post-processing method corresponds to the hypoechoic structure and not the ovary. However, in most cases the model is able to correctly identify the area of the ovary.

Table 6.1: Results of the final model for the total dataset and for each class division, for follicle segmentation.

Measure	**All**	**Class 1**	**Class 2**	**Class 3**	**Class 4**
Number of images	107	19	46	31	11
DSC	0.69±0.20	0.42±0.24	0.71±0.13	**0.82 ± 0.07**	0.68±0.11
ACC	0.65±0.22	0.39±0.22	0.63±0.17	0.78±0.13	**0.84 ± 0.17**
Precision	0.78±0.22	0.54±0.29	0.77±0.18	**0.90 ± 0.10**	0.88±0.10
FNR	0.20±0.23	0.40±0.31	0.19±0.18	0.14±0.15	**0.07 ± 0.16**
PD	0.47±0.19	0.74±0.16	0.45±0.15	**0.35 ± 0.12**	0.48±0.11

Regarding follicle segmentation, to better understand the limitations of the model, the results were grouped according to the mean area of the follicles in the corresponding GT image (in pixels, in a 512×512 px image) - Class 1 for mean areas lower than 1500 px, Class 2 for mean areas between 1500 and 5000 px, Class 3 for mean areas between 5000 and 10000 px and Class 4 for mean areas higher than 10000 px.

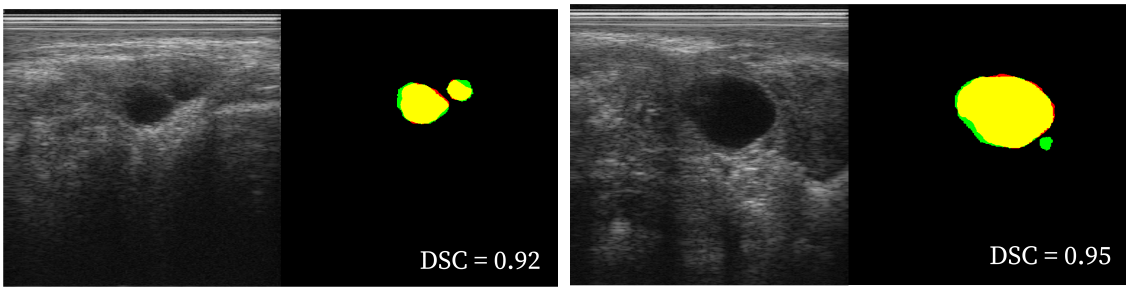
The results in Table 6.1 show that the lower values of DSC correspond to Class 1 and 4. These two classes include the most extreme cases in terms of follicle size and are the most uncommon cases in the dataset. The results from Class 1 report the higher FNR as those images contain follicles that are easily missed due to their small area and are difficult to detect even by the human eye (Figure 6.13.a). The results from Class 4 report the higher ACC as they are more easily detectable due to their large object area but, in terms of DSC, the performance is lower as in most cases only a part of the object is detected (Figure 6.13.b). This is due to the imbalance in the database, as most of the images in Class 4 are from the same patient, they were grouped together in a test-set for the validation of the final method. Consequently, the model used to obtain these



(a) (b)

Figure 6.13: Examples of the worst DSC results (a) Class 1 images (b) Class 4 images.

results had not been properly trained to predict objects with areas higher than 10000 px.



(a) (b)

Figure 6.14: Examples of the best DSC results (a) Class 2 images (b) Class 3 images.

On the other hand, Class 2 and 3, which are the most common cases in the dataset reported overall better performance. The follicles with stronger follicle edges result in the best predictions, as shown in Figure 6.14.

6.6 Comparison with State-of-the-art

This section presents the comparison of the perfomance of the proposed method regarding follicle segmentation and detection with the performance of the state-of-the-art methods.

Only methods that perform the analysis of 2D data are considered as they are more comparable with the proposed method. It is important to note that the accuracy metric presented in this section corresponds to the formulation that includes TN number (the definition presented in Section 3.3).

The overview of the performance of state-of-the-art methods and of the proposed method are shown in Table 6.2 and 6.3, respectively.

The results of these methods are not fully comparable given that they were computed for different datasets. Despite this, this comparison allows a general understanding of the range of values of performance that are expected for the proposed method.

Regarding the comparison of detection results (follicle-wise analysis), the method presented in [67] produced similar results to the proposed method for Classes 2, 3 and 4. As stated in the

Table 6.2: Overview of the performance of the state-of-the-art methods for follicle segmentation and detection.

Reference	Brief method description.	Performance evaluation
[67]	Canny edge detection and morphological operations.	Follicle-wise: - Sensitivity 75,2%; - FPR 22.5% - FNR 24.1%
[70]	Active contour without edges to define ROIs. GLCM technique is applied to extract second-order texture features, which are used as input in a MLP.	Follicle-wise: - Accuracy 96%; - Sensitivity 99%; - Specificity 93%
[71]	Active contour without edges to define ROIs. PSO is used to select the best performing features, among geometric, first order texture and GLCM features. The selected features are the input of the MLP.	Follicle-wise: - Accuracy 98.3%; - Sensitivity 100%; - Specificity 96.8%
[73]	Fully convolutional neural network based on the U-Net architecture [53]	Pixel-wise: - DSC 0.76 $\pm$ 0.2

Table 6.3: Results for follicle segmentation of the proposed method for the total dataset and for each class division (regarding follicle size).

Measure	**All**	**Class 1**	**Class 2**	**Class 3**	**Class 4**
Number of images	107	19	46	31	11
DSC	0.69±0.20	0.42±0.24	0.71±0.13	0.82±0.07	0.68±0.11
Accuracy	73%	51%	71%	83%	89%
Sensitivity	80%	60%	81%	86%	93%
Specificity	73%	53%	69%	86%	90%
FNR	20%	40%	19%	14%	7%
FPR	27%	47%	31%	14%	10%

previous section, Class 1 images have a higher FNR for the proposed method. Although the size of the follicles in the dataset used in [67] is not stated, from the images shown in the article, their work is implemented mostly on US images containing dominant follicle, which do not correspond to Class 1.

The results reported by [70] and [71] show higher value for accuracy, sensitivity and specificity. This suggests that an initial step of ROI selection to complement a deep learning methodology (such as the one proposed in this dissertation) leads to the improvement of the final results.

In regards to the comparison of segmentation results (pixel-wise), the performance of [73] is equivalent to the performance of the proposed method. Although the reported DSC of 0.76±0.2 is higher than the DSC of the proposed method, it is likely due to the size of the test-set (15 images) used by [73] and its different balance between the number of images of each follicle size class. Also, the database of [73] is composed by B-mode images directly acquired using Ultrasonix

SonixTouch Q+ software, which performs the conversion of BRF to B-mode data with an specific algorithm and more refined filtering techniques.

Chapter 7

Conclusion

The segmentation and detection of ovarian follicles are tasks frequently requested in clinical practice, and they are mostly done manually. An automatic and reliable algorithm to perform these tasks is required, given that establishing the delineations of such structures is extremely time-consuming and dependent on the experience of the clinicians.

The most challenging characteristics of the follicle segmentation and detection is the high variability of the shape and size of the follicles, which hampers the identification of common features during the optimization of the CNN. Besides that, the presence of hypoechoic structures besides follicles, such as vessels and the bladder, is problematic as it may result in a high number of false positives. Lastly, the low signal-to-noise ratio of the US data and the presence of malignant follicles with septa with a high degree of echogenicity intercepting them, makes it difficult to distinguish the separation between two neighboring follicles.

The main contribution of this work is the detailed analysis of the potential of a CNN based on the U-Net architecture to perform the segmentation and detection of the ovarian follicles and of the ovary, based on the study of the influence of the input data, pre-processing methods and several hyperparameters of the model. It is also shown how the final predictions of the U-Net model can be improved by the implementation of post-processing methods.

It was concluded that the type of input data and pre-processing methods proposed lead to significant improvement of the model's performance for ovary detection measures, when compared with the use of the original B-mode images, in the case of the use of radio-frequency data as complement to B-mode data and of B-mode data pre-processed using CLAHE. Besides that, the results of the study on the importance of the image context analysis on the model's performance showed that a model achieves better performance with a better ability to perform image context analysis (achieved through the decreased size of the input data, increase in the size of the kernel of U-Net's convolutions and additional level of resolution of the U-Net). A model that is able to analyse a more extended area of the image can more easily apprehend some key characteristics of the follicle and ovary segmentation, such as the difference between follicle and other hypoechoic structures and that the ovary prediction should consist in only one object.

Although these studies allowed for the improvement of the performance of the U-Net, regarding the detection of the structures, still some errors were produced. Hence the implementation of post-processing methods, for both follicle and ovary predictions. For the ovary predictions, the selection of the biggest object of the prediction leads to a higher accuracy on object detection results. For follicle predictions, the removal of small objects and of objects outside of the predicted ovary area decreased significantly the number of false positive objects, while the implementation of the watershed algorithm decreased the number of false negative objects. Additionally, the model using the proposed loss function is not suitable for follicle segmentation as it produces a higher number of false positive objects, but its predictions showed a much lower tendency for the incorrect union of neighboring follicles. Hence, those predictions were used in the computation of a post-processing method which improved the performance of the model in terms of the number of false negative objects.

The results of the final model showed that it is suitable to perform the follicle and ovary segmentation and detection tasks but its predictions still present some limitations that do not allow for its implementation in clinical practice. Despite this, the conclusions taken from this work are relevant for future investigation in the area.

In the future, the proposed loss function can be refined in order to decrease the number of false positive objects so that it can be implemented to obtain the direct predictions of the follicle, instead of being used in a post-processing method. It would also be beneficial to perform a global analysis combining the 4 experiments conducted, to verify that there is no interference between tested changes in the baseline method. Additionally, the US radio-frequency data can be explored using signal-processing algorithms instead of image-processing algorithms so that the method can utilize its complete information, without the data loss that the conversion from radio-frequency to B-mode entails. Lastly, an increase in the size of the dataset and a more balanced variety of follicle size in the dataset would also improve the model's performance.

Appendix A

Dataset Division

This Appendix presents the dataset division used for the validation of the final model.

Ten dataset arrangements are defined so that each one of the 107 images is in the test-set of one of the arrangements and the results for segmentation and detection are obtained for every image of the dataset.

Considering that multiple images in the dataset belong to the same patient, and that images from the same patient tend to be similar, the dataset division must ensure that those images are not simultaneously in the test-set and in training/validation-sets. Due to this patient restriction, for one of the dataset arrangements the test-set consists of 9 images, for five of the dataset arrangements consists of 10 images and for four of the dataset arrangements consists of 12 images, as described in Table A.1.

For each dataset arrangement, after the definition of the test-set, the remaining images are used for 5-fold cross-validation, which implies that each of those images is included in the validation-set in one of the five folds. In the case of the dataset arragements with a test-set composed by 9 and 10 images, the number of remaining images for 5-fold cross-validation (98 and 97, respectively) is not divisible by 5 and so the composition of the validation and training set varies between folds, as also shown in Table A.1.

Table A.1: Dataset division for the validation of the final model.

Number of datasets	**Number of images in**			**Number of folds**
	Test-set	**Validation-set**	**Training-set**	
1	9	19	76	2
		20	75	3
5	10	19	76	3
		20	75	2
4	12	19	76	5

Appendix B

U-Net Convergence - Learning Rate Hyperparameters

This Appendix presents the loss value progression for the validation set for the U-Net, using different initial learning rates (0.1, 0.01 and 0.001), for factors of reduction of learning rate of 5, 10 and 20, in Figures B.1, B.2 and B.3, respectively.

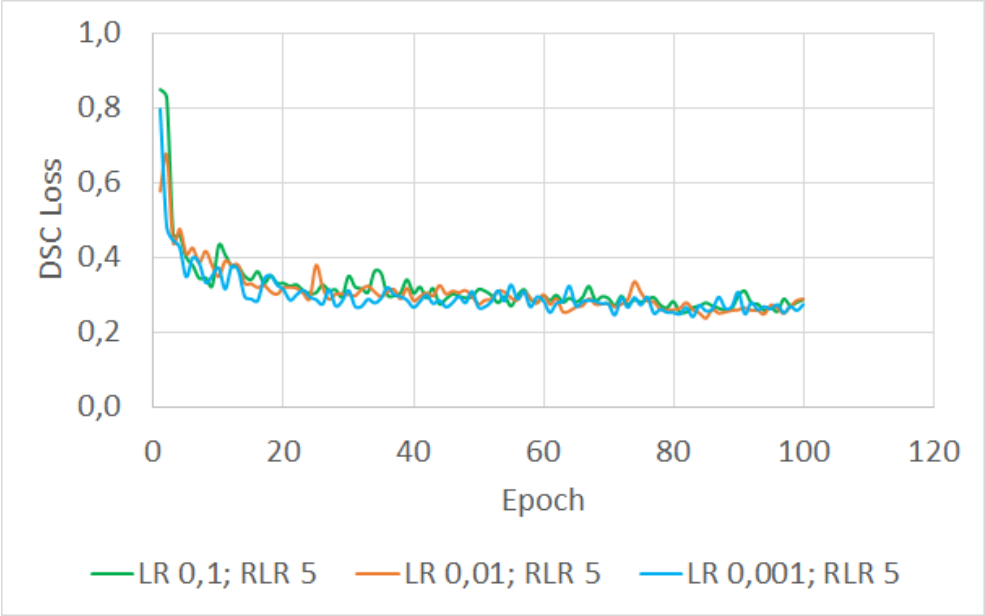


Figure B.1: Loss value progression for the validation set with different initial learning rates, for a factor of reduction of learning rate of 5.

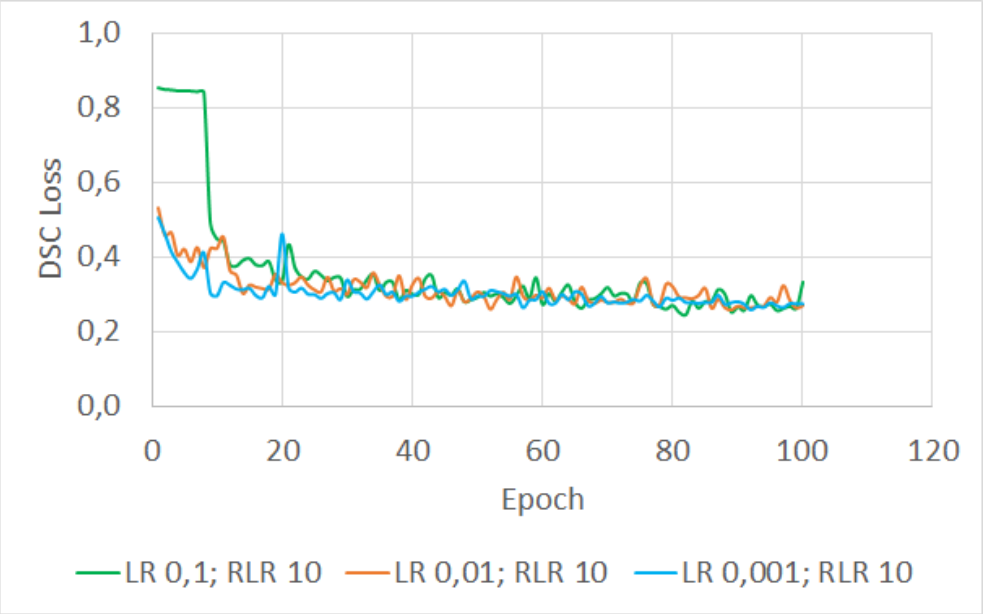


Figure B.2: Loss value progression for the validation set with different initial learning rates, for a factor of reduction of learning rate of 10.

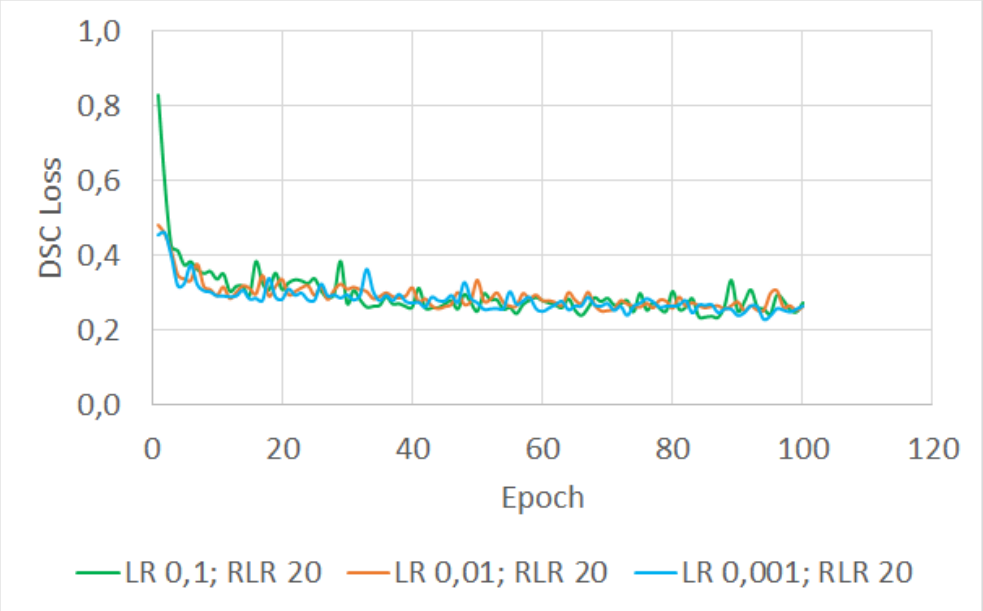


Figure B.3: Loss value progression for the validation set with different initial learning rates, for a factor of reduction of learning rate of 20.

References

[1] A. D. Dongare, R. R. Kharde, and A. D. Kachare, "Introduction to Artificial Neural Network," *International Journal of Engineering and Innovative Technology (IJEIT)*, vol. 2, no. 1, pp. 189–194, 2012.

[2] H. Feneis and W. Dauber, *Pocket Atlas of Human Anatomy*. New York: Thieme, 4 ed., 2000.

[3] P. B. Clement, "Anatomy and Histology of the Ovary," in *Blaustein's Pathology of the Female Genital Tract*, pp. 438–470, Springer, Boston, MA, 1987.

[4] B. Pansky, *Review of Medical Embryology*. New York: Macmillan, 6 ed., 1982.

[5] M. A. Coelho Neto, A. Ludwin, A. Borrell, B. Benacerraf, D. Dewailly, F. da Silva Costa, G. Condous, J. L. Alcazar, L. Jokubkiene, S. Guerriero, T. Van den Bosch, and W. P. Martins, "Counting ovarian antral follicles by ultrasound: a practical guide," *Ultrasound in Obstetrics and Gynecology*, vol. 51, no. 1, pp. 10–20, 2018.

[6] M. Pöhl, M. Hohlagschwandtner, A. Obruca, G. Poschalko, M. Weigert, and W. Feichtinger, "Number and size of antral follicles as predictive factors in vitro fertilization and embryo transfer," *Journal of Assisted Reproduction and Genetics*, vol. 17, no. 6, pp. 315–318, 2000.

[7] B. Potočnik, B. Cigale, and D. Zazula, "Computerized detection and recognition of follicles in ovarian ultrasound images: A review," *Medical and Biological Engineering and Computing*, vol. 50, no. 12, pp. 1201–1212, 2012.

[8] J. L. Frattarelli, D. F. Lauria-Costab, B. T. Miller, P. A. Bergh, and R. T. Scott, "Basal antral follicle number and mean ovarian diameter predict cycle cancellation and ovarian responsiveness in assisted reproductive technology cycles," *Fertility and Sterility*, vol. 74, no. 3, pp. 512 – 517, 2000.

[9] M. Y. Chang, C. H. Chiang, T. T. Hsieh, Y. K. Soong, and K. H. Hsu, "Use of the antral follicle count to predict the outcome of assisted reproductive technologies.," *Fertility and sterility*, vol. 69, no. 3, pp. 505–10, 1998.

[10] M. Christianson, G. Shoham, K. Tobler, Y. Zhao, C. Cordeiro, M. Leong, and Z. Shoham, "Measurement of antral follicle count in patients undergoing in vitro fertilization treatment: Results of a worldwide web-based survey," *Journal of Assisted Reproduction and Genetics*, vol. 32, no. 10, pp. 1435–1440, 2015.

[11] M. Monga, B. Alexandrescu, S. E. Katz, M. Stein, and T. Ganiats, "Impact of infertility on quality of life, marital adjustment, and sexual function," *Urology*, vol. 63, no. 1, pp. 126 – 130, 2004.

[12] M. Furuya, "Ovarian cancer stroma: pathophysiology and the roles in cancer development.," *Cancers*, vol. 4, no. 3, pp. 701–24, 2012.

[13] J. S. Richards, "From follicular development and ovulation to ovarian cancers: An unexpected journey," in *Vitamins and Hormones*, pp. 453–472, Elsevier, 2018.

[14] J. A. Rauh-Hain, T. C. Krivak, M. G. Del Carmen, and A. B. Olawaiye, "Ovarian cancer screening and early detection in the general population.," *Reviews in obstetrics and gynecology*, vol. 4, no. 1, pp. 15–21, 2011.

[15] T. Chen, W. Zhang, S. Good, S. K. Zhou, and D. Comaniciu, "Automatic ovarian follicle quantification from 3D ultrasound data using global/local context with database guided segmentation," *2009 IEEE 12th International Conference on Computer Vision*, pp. 795–802, 2009.

[16] V. Chan and A. Perlas, "Basics of ultrasound imaging," *Atlas of Ultrasound-Guided Procedures in Interventional Pain Management*, pp. 13–19, 11 2011.

[17] T. L. Szabo, "Chapter 1 - Introduction," in *Diagnostic Ultrasound Imaging: Inside Out (Second Edition)* (T. L. Szabo, ed.), pp. 1 – 37, Boston: Academic Press, second edition ed., 2014.

[18] R. S. Moorthy, "Transvaginal Sonography," *Medical Journal Armed Forces India*, vol. 56, no. 3, pp. 181–183, 2000.

[19] A. Carovac, F. Smajlovic, and D. Junuzovic, "Application of Ultrasound in Medicine," *Acta Informatica Medica*, vol. 19, no. 3, pp. 168 – 171, 2011.

[20] R. Rocha, "Image segmentation and reconstruction of 3D surfaces from carotid ultrasound images," *Engineering*, 2007.

[21] J. Wang, Y. Zhou, N. Koizumi, N. Kubota, T. Asano, K. Yuhashi, T. Mitake, K. Itani, T. Takahashi, S. Takeishi, S. Sasaki, T. Kadowaki, I. Sakuma, and H. Liao, "Automatic distance measurement of abdominal aorta for ultrasonography-based visceral fat estimation," in *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 6486–6489, July 2013.

[22] A. S. Rohr, S. L. Spector, S. C. Siegel, R. M. Katz, and G. S. Rachelefsky, "Correlation between A-mode ultrasound and radiography in the diagnosis of maxillary sinusitis," *The Journal of Allergy and Clinical Immunology*, vol. 78, no. 1 Part 1, pp. 58–61, 1986.

[23] J. M. Thijssen, "Ultrasonic tissue characterization," in *Acoustical Imaging* (M. Halliwell and P. N. T. Wells, eds.), pp. 9–25, Boston, MA: Springer US, 2002.

[24] K. Martin, "Introduction to B-mode imaging," in *Diagnostic Ultrasound: Physics and Equipment* (P. Hoskins, K. Martin, and A. Thrush, eds.), pp. 1–10, Cambridge University Press, 2010.

[25] A. C. Yu and J. A. Hossack, "Methods and protocols: Engine for the next wave of biomedical ultrasound innovations," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 64, no. 1, pp. 6–10, 2017.

[26] V. Krishna, F. Sammartino, and A. Rezai, "A review of the current therapies, challenges, and future directions of transcranial focused ultrasound technology advances in diagnosis and treatment," *JAMA Neurology*, vol. 75, no. 2, pp. 246–254, 2018.

[27] Y. Lin, L. Lin, M. Cheng, L. Jin, L. Du, T. Han, L. Xu, A. C. Yu, and P. Qin, "Effect of acoustic parameters on the cavitation behavior of sonovue microbubbles induced by pulsed ultrasound," *Ultrasonics Sonochemistry*, vol. 35, pp. 176 – 184, 2017.

[28] Y. Wang, W.-K. Bai, E. Shen, and B. Hu, "Sonoporation by low-frequency and low-power ultrasound enhances chemotherapeutic efficacy in prostate cancer cells in vitro," *Oncology Letters*, vol. 6, no. 2, pp. 495–498, 2013.

[29] M. Wang, Y. Zhang, C. Cai, J. Tu, X. Guo, and D. Zhang, "Sonoporation-induced cell membrane permeabilization and cytoskeleton disassembly at varied acoustic and microbubble-cell parameters," *Scientific Reports*, vol. 8, no. 1, pp. 1–12, 2018.

[30] E. Roux, "2D sparse array optimization and operating strategy for real-time 3D ultrasound imaging," *Acoustics. Université de Lyon*, 2016.

[31] B. Hussain, "Methods for Improved Estimation of Low Blood Velocities Using Vector Doppler Ultrasound," *Electronic Thesis and Dissertation Repository*, 2017.

[32] J. H. Chang, H. H. Kim, and K. K. Shung, "Frequency Compounded Imaging with a High-Frequency Dual Element Transducer," *Changes*, vol. 38, no. 3, pp. 319–335, 2010.

[33] J. F. Synnevåg, A. Austeng, and S. Holm, "Adaptive beamforming applied to medical ultrasound imaging," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 54, no. 8, pp. 1606–1613, 2007.

[34] A. Yu and L. Løvstakken, "Eigen-based clutter filter design for ultrasound color flow imaging: A review," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 57, pp. 1096–111, 05 2010.

[35] K. G. Kim, "Book Review: Deep Learning," *Healthcare Informatics Research*, vol. 22, no. 4, p. 351, 2016.

[36] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 8, pp. 1798–1828, 2013.

[37] Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[38] E. P. Xing, H. Gifford, and P. Karuturi, "Deep Learning and Graphical Models," *Spring*, pp. 1–9, 2015.

[39] S. Dreiseitl and L. Ohno-Machado, "Logistic regression and artificial neural network classification models: A methodology review," *Journal of Biomedical Informatics*, vol. 35, no. 5-6, pp. 352–359, 2002.

[40] R. De Oliveira, R. C. Fernandes Araújo, F. Barros, A. Paranhos Segundo, R. Zampolo, W. Fonseca, V. Dmitriev, and F. S. Brasil, "A system based on artificial neural networks for automatic classification of hydro-generator stator windings partial discharges," *Journal of Microwaves, Optoelectronics and Electromagnetic Applications*, vol. 16, pp. 628–645, 09 2017.

[41] A. Voulodimos, N. Doulamis, A. Doulamis, and E. Protopapadakis, "Deep learning for computer vision: A brief review," *Computational Intelligence and Neuroscience*, vol. 2018, pp. 1–13, 02 2018.

[42] J. Gu, Z. Wang, J. Kuen, L. Ma, A. Shahroudy, B. Shuai, T. Liu, X. Wang, G. Wang, J. Cai, and T. Chen, "Recent advances in convolutional neural networks," *Pattern Recognition*, vol. 77, pp. 354 – 377, 2018.

[43] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems 25* (F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, eds.), pp. 1097–1105, Curran Associates, Inc., 2012.

[44] D. Scherer, A. Müller, and S. Behnke, "Evaluation of pooling operations in convolutional architectures for object recognition," in *Proceedings of the 20th International Conference on Artificial Neural Networks: Part III*, ICANN'10, (Berlin, Heidelberg), pp. 92–101, Springer-Verlag, 2010.

[45] V. Suárez-Paniagua and I. Segura-Bedmar, "Evaluation of pooling operations in convolutional architectures for drug-drug interaction extraction," *BMC Bioinformatics*, vol. 19, p. 209, Jun 2018.

[46] J. Koutník, J. Schmidhuber, and F. Gomez, "Evolving deep unsupervised convolutional networks for vision-based reinforcement learning," *GECCO 2014 - Proceedings of the 2014 Genetic and Evolutionary Computation Conference*, 07 2014.

[47] C. Wachinger, M. Reuter, and T. Klein, "Deepnat: Deep convolutional neural network for segmenting neuroanatomy," *NeuroImage*, vol. 170, pp. 434 – 445, 2018. Segmenting the Brain.

[48] O. K. Ernst, "Stochastic Gradient Descent Learning and the Backpropagation Algorithm," *n.d.*, no. 3, pp. 1–9, n.d.

[49] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *Computing Research Repository (CoRR)*, vol. abs/1412.6980, 2015.

[50] F. Monteiro and A. Campilho, "Performance evaluation of image segmentation," *Image Analysis and Recognition: Third International Conference*, vol. 4141, pp. 248–259, 09 2006.

[51] V. H. Masías, M. Valle, C. Morselli, F. Crespo, A. Vargas Schüler, and S. Laengle, "Modeling verdict outcomes using social network measures: The watergate and caviar network cases," *PloS one*, vol. 11, p. e0147248, 01 2016.

[52] K. Zou, S. Warfield, A. Bharatha, C. Tempany, M. Kaus, S. J Haker, W. Wells, F. Jolesz, and R. Kikinis, "Statistical validation of image segmentation quality based on a spatial overlap index," *Academic radiology*, vol. 11, pp. 178–89, 02 2004.

[53] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 9351, pp. 234–241, 2015.

[54] F. Milletari, N. Navab, and S. A. Ahmadi, "V-Net: Fully convolutional neural networks for volumetric medical image segmentation," *Proceedings - 2016 4th International Conference on 3D Vision*, pp. 565–571, 2016.

[55] E. M. A. Anas, P. Mousavi, and P. Abolmaesumi, "A deep learning approach for real time prostate segmentation in freehand ultrasound guided biopsy," *Medical Image Analysis*, vol. 48, pp. 107–116, 2018.

[56] M. H. Yap, G. Pons, J. Martí, S. Ganau, M. Sentís, R. Zwiggelaar, A. K. Davison, and R. Martí, "Automated Breast Ultrasound Lesions Detection Using Convolutional Neural Networks," *IEEE Journal of Biomedical and Health Informatics*, vol. 22, no. 4, pp. 1218–1226, 2018.

[57] K. Horsch, M. L. Giger, L. A. Venta, and C. J. Vyborny, "Automatic segmentation of breast lesions on ultrasound," *Medical Physics*, vol. 28, no. 8, pp. 1652–1659, 2001.

[58] A. Mishra, P. Dutta, and M. Ghosh, "A ga based approach for boundary detection of left ventricle with echocardiographic image sequences," *Image and Vision Computing*, vol. 21, no. 11, pp. 967 – 976, 2003.

[59] B. Sahiner, H. P. Chan, M. A. Roubidoux, M. A. Helvie, L. M. Hadjiiski, A. Ramachandran, C. Paramagul, G. L. LeCarpentier, A. Nees, and C. Blane, "Computerized characterization of breast masses on three-dimensional ultrasound volumes," *Medical Physics*, vol. 31, no. 4, pp. 744–754, 2004.

[60] A. Madabhushi and D. N. Metaxas, "Combining low-, high-level and empirical domain knowledge for automated segmentation of ultrasonic breast lesions," *IEEE Transactions on Medical Imaging*, vol. 22, pp. 155–169, Feb 2003.

[61] L. D. Cohen and I. Cohen, "A finite element method applied to new active contour models and 3D reconstruction from cross sections," *Proceedings of the Third International Conference on Computer Vision*, pp. 587–591, 1990.

[62] N. Lin, W. Yu, and J. S. Duncan, "Combinative multi-scale level set framework for echocardiographic image segmentation," *Medical Image Analysis*, vol. 7, no. 4, pp. 529 – 537, 2003. Medical Image Computing and Computer Assisted Intervention.

[63] M. Mignotte, J. Meunier, and J. C. Tardif, "Endocardial boundary estimation and tracking in echocardiographic images using deformable templates and Markov Random Fields," *Pattern Analysis and Applications*, vol. 4, no. 4, pp. 256–271, 2001.

[64] D. Boukerroui, A. Baskurt, J. A. Noble, and O. Basset, "Segmentation of ultrasound images-multiresolution 2D and 3D algorithm based on global and local statistics," *Pattern Recognition Letters*, vol. 24, no. 4-5, pp. 779–790, 2003.

[65] F. Heitz, P. Perez, and P. Bouthemy, "Multiscale minimization of global energy functions in some visual recovery problems," *CVGIP: Image Understanding*, vol. 59, no. 1, pp. 125 – 134, 1994.

[66] Y. L. Huang, Y. R. Jiang, D. R. Chen, and W. K. Moon, "Watershed segmentation for breast tumor in 2-D sonography," *International Journal of Computer Assisted Radiology and Surgery*, vol. 1, no. SUPPL. 7, pp. 63–65, 2004.

[67] P. Hiremath and J. Tegnoor, "Automatic detection of follicles in ultrasound images of ovaries using edge based method," *International Journal of Computer Applications*, 01 2010.

[68] D. Moreno Bermudez, D. Ballesteros Padilla, and G. Sanchez Torres, "Image analysis for automatic feature estimation of the Mangifera indica fruit," *Ingeniería y Desarrollo*, vol. 31, pp. 84 – 104, 07 2013.

[69] Z. Tu, "Probabilistic boosting-tree: learning discriminative models for classification, recognition, and clustering," in *Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1*, vol. 2, pp. 1589–1596 Vol. 2, Oct 2005.

[70] A. D. Usman, O. R. Isah, and A. M. S. Tekanyi, "Application of Artificial Neural Network and Texture Features for Follicle Detection," *African Journal of Computing & ICT*, vol. 8, no. 4, pp. 2–9, 2015.

[71] O. R. Isah, A. D. Usman, and A. M. Tekanyi, "A hybrid model of PSO algorithm and artificial neural network for automatic follicle classification," *International Journal Bioautomation*, vol. 21, no. 1, pp. 43–58, 2017.

[72] B. Cigale, D. Zazula, and S. Member, "Directional 3D wavelet transform based on gaussian mixtures for the analysis of 3D ultrasound ovarian volumes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, pp. 64–77, Jan 2019.

[73] D. S. Wanderley, C. B. Carvalho, A. Domingues, C. Peixoto, D. Pignatelli, J. Beires, J. Silva, and A. Campilho, "End-to-end ovarian structures segmentation," in *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications* (R. Vera-Rodriguez, J. Fierrez, and A. Morales, eds.), (Cham), pp. 681–689, Springer International Publishing, 2019.

[74] M. Nii, Y. Kato, M. Morimoto, S. Kobashi, N. Kamiura, Y. Hata, S. Imawaki, T. Ishikawa, and H. Matsubayashi, "Ovarian follicle classification using numerical and b-mode image features from ultrasound scanning devices," in *2018 Joint 7th International Conference on Informatics, Electronics Vision (ICIEV) and 2018 2nd International Conference on Imaging, Vision Pattern Recognition (icIVPR)*, pp. 222–227, June 2018.

[75] R. G. Dantas, E. T. Costa, and S. Leeman, "Ultrasound speckle and equivalent scatterers," *Ultrasonics*, vol. 43, no. 6, pp. 405 – 420, 2005.

[76] E. D. Pisano, S. Zong, B. M. Hemminger, M. DeLuca, R. E. Johnston, K. Muller, M. P. Braeuning, and S. M. Pizer, "Contrast limited adaptive histogram equalization image processing to improve the detection of simulated spiculations in dense mammograms," *Journal of Digital Imaging*, vol. 11, p. 193, Nov 1998.

[77] L. Najman and M. Schmitt, "Watershed of a continuous function," *Signal Processing*, vol. 38, pp. 99–112, 07 1994.

[78] J. Cardoso and L. Corte-Real, "Toward a generic evaluation of image segmentation," *Image Processing, IEEE Transactions on*, vol. 14, pp. 1773 – 1782, 12 2005.

[79] Y. Liu, "Box plots: use and interpretation," *Transfusion*, vol. 48, pp. 2279–2280, Nov. 2008.