U.PORTO

FACULDADE DE CIÊNCIAS
UNIVERSIDADE DO PORTO

MSc

2.º CICLO

FCUP
2021

U.PORTO

CT segmentation in the context of Prostate Cancer

Ana Sofia Oliveira Couto

FC

CT segmentation in the context of Prostate Cancer

Ana Sofia Oliveira Couto

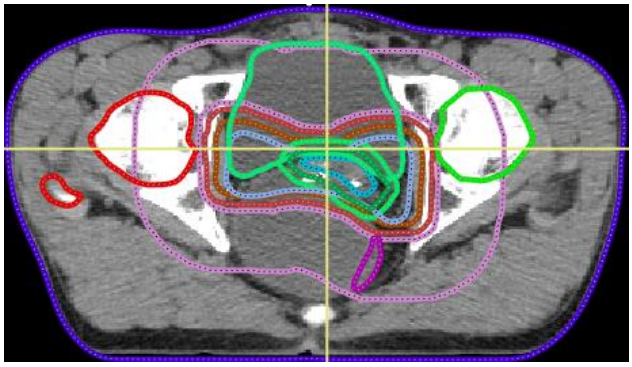
Dissertação de Mestrado apresentada à
Faculdade de Ciências da Universidade do Porto em
Engenharia Matemática

2021

U.PORTO

FACULDADE DE CIÊNCIAS
UNIVERSIDADE DO PORTO

MSc



CT segmentation in the context of Prostate Cancer

Ana Sofia Oliveira Couto
Mestrado de Engenharia Matemática
Departamento de Matemática
2021

Orientador
João Santos, Investigador, IPO

Coorientador
Inês Domingues, Investigadora, IPO

U.PORTO

FACULDADE DE CIÊNCIAS
UNIVERSIDADE DO PORTO

Todas as correções determinadas pelo júri, e só essas, foram efetuadas.

O Presidente do Júri,

Porto, _____/_____/_____

Resumo

A radioterapia desempenha um papel muito importante no tratamento do câncro. É responsável por 78% das curas de cancro não tratadas com cirurgia. Contudo, se afetar tecido saudável, aumenta os efeitos secundários. Por essa razão, segmentar os orgãos em risco é crucial. Atualmente, este processo é feito manualmente, o que é demorado e está sujeito a erro humano.

Com o objetivo de ajudar os especialistas e melhorar a precisão da segmentção, testaram-se alguns algoritmos para segmentar a bexiga em 47 Tomografias Computadorizadas de 47 pacientes com cancro da próstata, fornecidas pelo Instituto de Oncologia do Porto. Foram desenvolvidos dois pré-processamentos para detetar a região de interesse: um baseado nos valores de Hounsfiel Unit da bexiga e outro baseado na anatomia. Os algoritmos explorados foram *Clustering*, U-Net, *Active Contours* e *Graph Based*. O algoritmo *Clustering* usado foi o *K-means*. Aplicado o algoritmo, foram implementados dois pós-processamentos. No primeiro método, foi implementado *flood fill* para preencher regiões e buracos e no segundo foi escolhida a maior região conexa. O algoritmo U-Net usa Aprendizagem Profunda para prever a classe de cada pixel através de fragmentos locais que os rodeiam. Os pós-processamentos destes dois algoritmos com melhores resultados foram usados para definir a máscara inicial com os contornos iniciais a partir dos quais começa a evolução da segmentação pelo algoritmo *Active Contours*. Por último, no algoritmo *Graph Based* as máscaras que designam os píxeis nas imagens como sendo do primeiro plano (*foremask*) ou do plano de fundo (*backmask*) foram definidas com três abordagens: manualmente; definindo o resultado do pós-processamento do *clustering* como *foremask* e definindo a *backmask* ao implementar uma semente; definindo o resultado do pós processamento do U-Net como *foremask* e usando a *backmask* manual.

Os quatro algoritmos foram avaliados e o Dice obtido para o *Clustering*, U-Net, *Active Contours* e *Graph Based* foram 24%, 6%, 26% e 29%, respetivamente, no conjunto de dados com a máscara

baseada nos valores de HU. Para a máscara anatómica, os resultados obtidos ao aplicar a mesma métrica aos mesmos algoritmos foram 22%, 80%, 31% e 20%, respetivamente. O método que se destacou foi o U-Net usando a máscara anatómica e operações morfológicas, uma vez que foi a abordagem que identificou mais corretamente os píxeis correspondentes à bexiga.

No futuro, pretende-se usar técnicas de otimização de hiperparâmetros para escolher os melhores parâmetros em cada algoritmo, para se obter melhores resultados. Uma vez que o reto é um orgão em risco no contexto de cancro da próstata, planeia-se implementar os métodos anteriores, com o objetivo de estudar se o seu desempenho é semelhante.

Palavras-chave: Radioterapia, Orgãos em Risco, Bexiga, Tomografias Computadorizadas, Segmentação.

Abstract

Radiotherapy takes a very important role in cancer treatment. It is responsible for 78% of non-surgical cancer cures. However, when it affects health tissue, it increases the side effects. For that reason, segmenting the Organs at Risk is crucial. This process is currently done manually, which is time consuming and subject to human error.

With the goal of helping the specialists and improving segmentation accuracy, some algorithms were tested to segment the bladder in 47 Computerized Tomography scans from 47 patients with prostate cancer provided by the Institute of Oncology of Porto. Two pre-procedures were developed in order to detect the region of interest: one based on the Hounsfield Unit values of the bladder and another one based on the anatomy. The algorithms explored were Clustering, U-Net, Active Contours and Graph Based. The clustering algorithm used was K-means. After applying it, two different post-processing were implemented. In the first one, it was implemented flood fill to fill regions and holes. In the second approach, the biggest connect region was chosen. The U-Net algorithm uses deep learning to predict the class label of each pixel by providing a local patch around that pixel as input. The post-processing of these two algorithms with the best results were used to define the initial mask with the initial contours at which the evolution of segmentation begins by Active Contours algorithm. At last, for the Graph Based algorithm, the foremask and backmask (the masks that design pixels in the images as foreground and background, respectively) were defined with three approaches: manually; defining the clustering post-processing results as foremask and defining the backmask by implementing a seed; defining the U-Net post-processing results as foremask and using the manual backmask.

The four algorithms were evaluated and the Dice obtained for applying Clustering, U-Net, Active Contours and Graph Based were 24%, 6%, 26% and 29%, respectively, in the data set

with the mask based on HU values. For anatomic mask, the results obtained by applying the same metric to the same algorithms were 22%, 80%, 31% and 20%, respectively. The method that stood out was the U-Net using the anatomical mask and morphological operations, since it was the approach that most correctly identified the pixels corresponding to the bladder.

In the future, it is intended to use hyperparameter optimization techniques to choose the optimal parameters in each algorithm, to obtain better results. Since the rectum is an organ at risk in the context of prostate cancer, it is planned to implement the previous methods, with the aim of studying whether their performance is similar.

Key-words: Radiotherapy, Organs at Risk, Bladder, Computerized Tomography, Segmentation

Acknowledgements

A realização desta dissertação não teria seguido o mesmo rumo sem todos aqueles que, de alguma forma, foram essenciais para o seu desenvolvimento. A essas pessoas, a minha sincera gratidão:

À Doutora Inês Domingues, por toda a orientação, disponibilidade e dedicação. A minha evolução ao longo deste ano é fruto de todo o seu aconselhamento, rigor e conhecimento.

Ao Doutor João Santos, por todo o apoio e confiança.

Ao Doutor Matthew Davies, por toda a ajuda e disponibilidade.

Aos meus pais, por fazerem de tudo para que eu possa seguir os meus sonhos e por me darem força para correr atrás deles.

À minha irmã, por toda a motivação, força e preocupação.

Ao Hoff, por ser o meu suporte. Por todo o amor, apoio e compreensão.

À minha família e amigos, por todo o carinho e confiança.

Contents

List of Figures

List of Tables

Acronyms

3TMRI 3 Tesla MRI. 53

AAVE Average Absolute Volume Error. 44

AHD Average Hausdorff Distance. 57

AJI Average Jaccard Index. 56

AMD Average Minimum Distance. 56

ANN Artificial Neural Networks. 37

ART Adaptive Radiotherapy Plan. 54

ASD Average Surface Distance. 44

AVD Average Distance. 44

AVE Average Volume Error. 44

AVI Average Volume Intersection. 56

BCnet Boundary Coding Network. 52

BSP Binary Space Partitioning. 49

Btrfly-Nets Butterfly-type Networks. 53

CBCT Cone Beam Computed Tomography. 51

CC Cross Correlation. 48

CLAHE Contrast Limited Adaptive Histogram Equalization. 52

CNN Convolutional Neural Networks. 39

CT Computed Tomography. 19, 25

CTU Computerized Tomography Urogram. 50

CTV Clinical Target Volume. 23

CycleGAN Cycle Generative Adversarial Network. 50

DAUnet Deep Attention U-Net. 51

DICOM Digital Imaging and Communications in Medicine. 25

DL-CNN Deep Learning Convolutional Neural Network. 50

DSC Dice Similarity Coefficient. 43

dV-net Dense V-net. 53

EM Expectation Maximization. 33

FMS F-Measure. 43

FN False Negatives. 41

FNR False Negatives Rate. 42

FP False Positives. 41

FPR False Positives Rate. 42

GT Ground Truth. 41, 64, 67, 68, 73, 75, 78

GTV Gross Tumor Volume. 23

GVF Gradient Vector Flow. 31

HD Hausdorff Distance. 44

HU Hounsfield Unit. 25

JAC Jaccard Index. 43

LCR Local Contour Refinement. 56

MAP Maximum a Posteriori. 30, 49

MI Mutual Information. 48

MMSE Mini Mental State Examination. 49

MRF Markov Random Field. 30

MSL Marginal Space Learning. 49

MST Minimum Spanning Tree. 33

OR Organs at Risk. 24, 61

PCA Principal Component Analysis. 45

PGA Principal Geodesic Analysis. 47

POD Probability of Detection. 41

PPV Positive Predictive Value. 42

PRV Planning Organ at Risk Volume. 24

PTV Planning Target Volume. 23

PVO Percentage of Volume Overlap. 44

RoI Region of Interest. 20, 21

RUS Random Under-Sampling. 37

SDE Surface Distance Error. 44

SDSM Statistical Deformable Shape Model. 47

sMRI Synthetic Magnetic Resonance Imaging. 50

SSD Sum of Squared Differences. 48

std standard deviation. 26

SVM Support Vector Machine. 67

TN True Negatives. 41

TNR True Negatives Rate. 41

TP True Positives. 41

TPR True Positives Rate. 41

Chapter 1.

Introduction

Medical imaging is the process that leads to visualisations of the human body at a microscopic level (cells and tissues) and at a macroscopic level (organs and systems). These representations are bio-medical images. There are several techniques to capture them: using light (endoscopy, OCT), magnetism (MRI), radioactive pharmaceuticals (PET, SPECT), X-rays (CT scan) or sound (ultrasound) [1]. This work will focus on Computed Tomography (CT) scans. It consists of a beam of x-rays which is aimed at a patient and rotated around the body, producing signals that are processed by the computer to generate several images of the body named slices. These slices can be analysed individually or they can be digitally combined, forming a 3D image of the body [2].

Segmentation is the process that divides an image in different objects with common characteristics. In the medical domain, these objects can be organs or other regions of interest, such as tumours. CT segmentation is one of the most critical challenges of biomedical images since it is a difficult, labor-intensive, error-prone and very time-consuming procedure. With the purpose of addressing these challenges, different segmentation algorithms will be explored in this thesis.

1.1 Contextualization

In Portugal, prostate cancer is the most common cancer for 50+ year old men and is second after lung cancer in causing deaths [3]. According to *Associação Portuguesa de Urologia*, it is estimated that it affects 82 in 100000 inhabitants and it has a mortality of 33 in 100000 inhabitants. In 2016, it was responsible for the death of 76900 European men, which represents 3% of

all male deaths and 10% of all male cancer-related deaths [3, 4]. Surgery and radiotherapy are the most efficient and used treatments for cancer. More specifically, radiotherapy is responsible for 78% of non-surgical cancer cures [5]. 60% of the patients submitted to radiotherapy are treated with curative intent but it also has an important role in the reduction of symptoms [5]. It consists of applying a certain radiation dose to the Region of Interest (RoI), to try to save the healthy tissue. There are three important axioms to this procedure: an increase of radiotherapy dose to the tumor normally improves the probability of local control; improving local control in the context of a localised tumor achieves an improvement in the overall cure rate; sparing normal tissues improves the side effects of radiotherapy [5]. It means that an increase of dose will increase the probability of cure but also the side effects. In order to find this balance, it is important to observe in detail where the tumor ends and the surrounding organs begin. This procedure is made manually. In the context of prostate cancer, the prostate, the bladder, the femoral heads and the penile bulb are outlined. The risk of involvement of seminal vesicles and ganglia is calculated using a formula. If the risk is too high, they are outlined too. Usually, the treatment starts by outlining the prostate introducing three transrectal markers so that there is a triangulation. Then, a balloon is inserted into the rectum. Its contour is made by outlining the balloon wall. Once the urethra is collapsed, a catheter is introduced and the outlining is made after setting the brush to $3mm$. This approach is painless but may have some side effects. In the middle of the treatment, some urinary burning may appear and, after the treatment, less than 10% of the cases have some of these effects: varicose veins on the rectal wall, inflammation of the urethra or erectile dysfunction. The variability inter and intra-subjects is another problem [6, 7]. It refers not only to the variability of two different contours of two different specialists but also the variability of two different contours from one specialist. In addition, it is a very time-consuming process (the contour of each structure takes about 3 hours) and it is subject to human error.

1.2 Objectives

The main goal of this thesis is to study different segmentation algorithms to apply to prostate cancer patients' CT scans provided by the Portuguese Institute of Oncology of Porto (IPO). More specifically, it is planned to apply the algorithms to segment the bladder. By doing so, it is intended to reduce the time consuming of manual segmentation and improve its accuracy.

1.3 Contributions

The principal contributions of this thesis are:

- Development of two approaches to detect the RoI;
- Comparison between four segmentation algorithms;
- Application of two post-processing techniques;
- Validation in a real database of 47 CT scans from 47 patients with prostate cancer.

This work also resulted on the paper:

Ana Couto, Inês Domingues and João Santos. "Comparison of bladder segmentation techniques in CT scans". RecPad 2021 (attached in A).

1.4 Document Structure

This report is divided into six Chapters. Chapter 1 is a theme introduction and contextualization. In Chapter 2, the theoretical concepts used in this work are presented. Chapter 3 starts with a description of the different approaches used in segmentation of organs related with prostate cancer. Then, in Section 3.2, some works related with bladder segmentation are presented. The methods used in this work are explained in Chapter 4 and the respective obtained results can be analysed and compared in Chapter 5. Finally, in Chapter 6, conclusions about the work are drawn and some directions for future work are mentioned.

Chapter 2.

Background

For a better follow-up of all the segmentation processes that will be presented, there are some concepts that are important to know in advance. To better understand the radiotherapy process and the importance of segmentation, the concepts of target volumes and organs at risk are explained in Sections 2.1 and 2.2, respectively. Several image processing techniques are used to manipulate the CT scans. Those techniques are described in Section 2.3. Different segmentation algorithms will be studied throughout this work. For this reason, the main ones will be presented in section 2.4, followed by the main evaluation metrics used in segmentation, in Section 2.5.

2.1 Target Volumes for Radiotherapy

The definition of tumour and target volumes provide the best possible characterisation of their location and extent. Therefore, it is essential for radiotherapy planning.

There are three main target volumes (Figure 2.1). The first one represents the part of the tumour where tumour cell density is highest, known as the Gross Tumor Volume (GTV) [5]. Since the tumour control requires a higher dose if the initial cell number is larger, it may have some implications for choice of radiotherapy dose. Although the GTV is the easiest volume to define, in practice, the edges are not always clear. The second one is the Clinical Target Volume (CTV). It surrounds the GTV and describes microscopic tumor spread. Generally, the CTV margin cannot be fully imaged and it is difficult to find, since it requires a clinical assessment of risk and extent of spread. The third volume contains the CTV and the tissue needed to ensure that the radiotherapy prescription dose is delivered to the CTV, known as the Planning

Target Volume (PTV). Therefore, it allows uncertainties in planning or delivery [5].

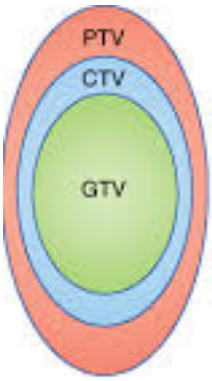


Figure 2.1: Target volumes for radiotherapy: GTV, CTV, PTV. Figure from [5]

2.2 Organs at Risk

The Organs at Risk (OR) are normal tissue that surround the target volumes and whose radiation sensitivity influences treatment planning or the prescribed radiation. They must always be considered in radiotherapy planning. It is helpful to create a Planning Organ at Risk Volume (PRV) around an OR whose damage is especially dangerous, particularly those whose loss of a small amount of normal tissue from radiation damage would produce a severe clinical manifestation [5].

The main OR typically considered in prostate cancer are the bladder, rectum and femoral heads (Figure 2.2).

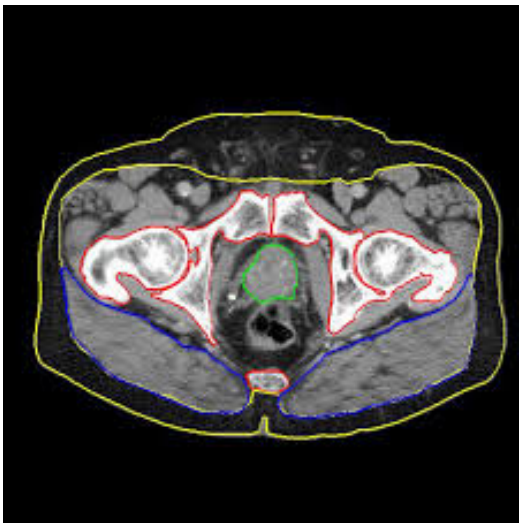


Figure 2.2: OR in context of prostate cancer: bladder (green), rectum (red, below the bladder) and femoral heads (red, on the right and the left side of the bladder).

2.3 Image Processing

It is common to use image processing techniques in the CT scans with the main goal of improving their quality and extracting information in a robust, efficient and accurate manner.

2.3.1 Conversion of CT data to HU

Hounsfield Unit (HU) is a dimensionless unit universally used in CT scanning to express CT numbers in a standardized and convenient form. To convert the CT data to HU, the following linear transformation is applied:

$$HU = RescaleSlope \times CT + RescaleIntercept \tag{2.1}$$

The *RescaleSlope* and the *RescaleIntercept* values are stored in the CT Digital Imaging and Communications in Medicine (DICOM) file. A DICOM file contains the scanned image and the data characteristics of the image and of the patients.

2.3.2 Contrast Stretching

This process changes the range of pixel values into a specified range using a linear transformation. Its usual purpose is to convert an input image into a more normal range to the senses.

Normalization by Scaling Between 0 and 1

A typical normalization scales the values between 0 and 1. Let X be the variable to be normalized. Its normalized value is calculated as:

$$normalized\_X_i = \frac{X_i - X_{\min}}{X_{\max} - X_{\min}} \tag{2.2}$$

Normalization with the mean and the standard deviation

Another common normalization is to set the mean variable value to 0 and the standard deviation (std) to 1. Let X be the variable to normalize. Its normalized value is calculated as:

$$normalized\_X_i = \frac{X_i - mean(X)}{std(X)} \tag{2.3}$$

2.3.3 Morphological Operations

Morphological operations can affect the form, structure or shape of an object. They are usually applied on binary images (black and white images).

Flat Morphological Structuring

Flat Morphological Structuring is a binary valued neighborhood in which the true pixels are included in the morphological computation and the false ones are not. This operation is represented in Figure 2.3, where the structuring elements are created in the decomposition of a sphere-shaped with radius 3. The center, called the origin, identifies the pixel in the image being processed [8].

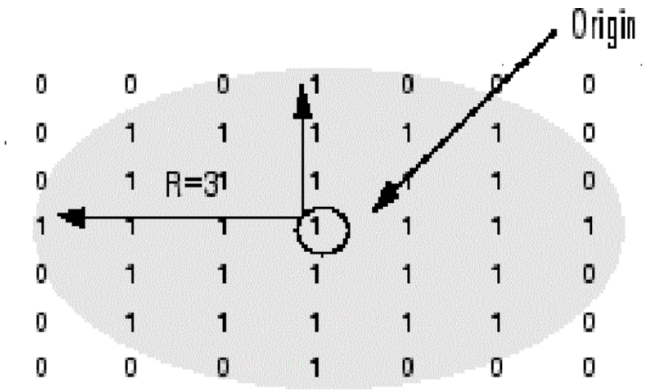


Figure 2.3: Creation of a nonflat structuring element. Figure from [8]

Dilation

Dilation is a process performed by laying the structuring element on the image and sliding it across the image in a manner similar to convolution. The dilation operator takes two pieces of data as inputs. The first is the image which is to be dilated. The second is a set of coordinated points known as a kernel. As shown in Figure 2.4, if the origin of the kernel coincides with

a foreground pixel (1's), all pixels from the image are covered by the structuring element as foreground pixels. However, if it coincides with a background pixel (0's), there is no change [9].

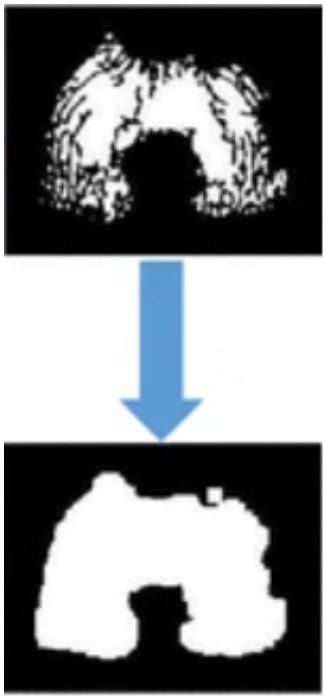


Figure 2.4: CT dilation. Figure from [10]

Flood Fill

This is an algorithm that determines and changes connected background pixels to foreground pixels, which finishes when it reaches object boundaries. In gray scale images, it brings the intensity values of dark areas that are surrounded by lighter areas up to the same intensity level as the surrounding pixels [11]. A common use of the flood fill operation is to fill holes in images, which is shown in Figure 2.5.

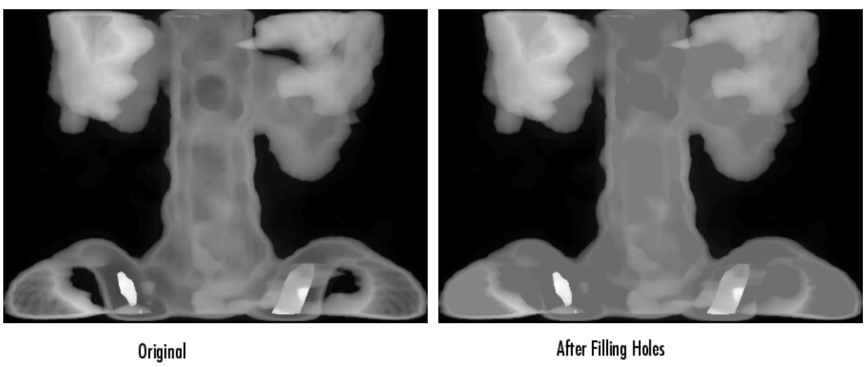


Figure 2.5: CT before and after filling holes. Figure from [11]

2.3.4 Rotation

Image rotation is a common image processing routine with applications in matching, alignment and other image-based algorithms. In this work, it will be used with the goal of data augmentation. It transforms the position (x_1, y_1) of an input image onto a position (x_2, y_2) by moving it around a certain point O with the coordinates (x_0, y_0) (rotation centre), until it has completed a given rotation angle θ. The transformations performed are:

$$x_2 = cos(\theta)(x_1 - x_0) - sin(\theta)(y_1 - y_0) + x_0 \tag{2.4}$$

$$y_2 = sin(\theta)(x_2 - x_0) + cos(\theta)(y_2 - y_0) + y_0 \tag{2.5}$$

2.3.5 Reflection

The reflection transforms an image so that the pixel values located at position (x_1, y_1) in the original image are reflected through an image axis or image point into a new position (x_2, y_2). Reflection is mainly used as an aid to image visualization. In this work, it will be combined with rotation to augment the data.

2.3.6 Crop

Cropping is the removal of undesired areas from an image by specifying the size and position of the cropping window.

2.4 Segmentation Algorithms

There are many segmentation techniques, based on different type of approaches: statistics, geometry, machine learning, anatomy, etc. In this section, the main segmentation algorithms will be presented: Thresholding, Region Growing, Deformable Models, Atlas Based, Markov Random Field Models, Active Contours, Graph Based, Clustering Algorithms and Classifier Based. In spite of being presented individually, the techniques are often used in conjunction [12].

2.4.1 Thresholding approaches

The thresholding approaches attempt to determine one or more intensity values, called a threshold. The pixels whose intensities are limited by the threshold are grouped into a class, defining a binary image. In some cases, it is possible to analyse this division through the histogram of intensity values that can also help to find the thresholds. This is a simple process, usually performed interactively and often used as an initial step.

Otsu's Method is the most used one. When the threshold value has to be manually specified, the experimentation of different values can be a tedious and time-consuming task. Otsu's method is a technique that automatically determines a single intensity threshold, separating pixels into two classes: foreground and background. This threshold is determined by minimizing or maximizing intra-class intensity variance.

2.4.2 Region Growing

Region growing is a procedure that extracts an image region that is connected based on some predefined criteria. The criteria could be based on intensity information or edges in the image [12]. An operator manually selects a seed point and the pixels connected with it are extracted.

The disadvantage of this approach is that it requires manual interaction to obtain the seed point. Therefore, a seed is needed for each region that needs to be outlined. It can also be sensitive to noise which can lead to some disconnected regions.

2.4.3 Deformable Models

Deformable models delineate region boundaries by combining geometry, physics and approximation theory. Geometry represents the object shape, physics imposes constraints on shape variations over space and time and optimal approximation theory provides the formal mechanisms for fitting the models to measured data [13].

A close curve or surface is placed near the desired boundary allowing an iterative relaxation process. Through the deformation energy, the curve or surface computes internal forces to keep

it smooth. Taking a physics-based view of classical optimal approximation, external potential energy functions are defined in order to fit the model [12, 13]. In some cases, the model fitting process is made in a probabilistic framework, which incorporates the prior model in terms of probability distributions.

These models' advantages are the possibility of generating closed parametric curves or surfaces and the constraint smoothness, providing a robustness to noise and edges. However, they require manual interaction to initialize the model and optimize the hyper parameters [12].

2.4.4 Atlas Based

Atlas based approaches are a powerful process, once an atlas is available. The atlas can be generated by manually segmenting an image or through a compilation of information about the anatomy of the structure to segment. The labels are also included. The process is initialized with an atlas warping [12]. It consists of finding a transformation that maps the atlas to the target image. The warping can be performed with linear transformations, but due to anatomical variability, a sequential application of linear and nonlinear transformations is often used.

The anatomical variability could be a problem when the structures are too complex. The use of probabilistic atlases helps with it, however it requires additional time and interaction to accumulate data.

2.4.5 Markov Random Field Models

Markov Random Field (MRF) is a statistical model used in segmentation methods, modeling spatial interactions between neighboring or nearby pixels [12]. It is a method often used in medical imaging, since the majority of the pixels belongs to the same class as their neighborhood. The goal of this model is to determine the optimal label of observation data Y, given a label field X [14]. This label x_{opt} is given by the Maximum a Posteriori (MAP) criterion:

$$x_{opt} = argmax P(X|Y) \tag{2.6}$$

where $P(X|Y)$ is the posterior probability of the label field X under the condition of observation data Y.

Spite of being useful in segmentation, MRF methods have some disadvantages. There is some difficulty to select the parameters that control the strength of spatial interactions. If the parameters are too high, it can result in a smooth segmentation, losing important structural details. Moreover, it usually requires computationally intensive algorithms [12].

2.4.6 Active Contours

Active contours algorithms use energy forces and constraints for segregation of the pixels of interest. It describes the object boundaries or other features of the image by forming a parametric curve. The curvature is determined with contour algorithms that use external and internal forces. Energy functional is associated with the curve defined in the image. By minimizing the energy functional, a contour fits the required image contour. Contour deformation is described by a collection of points that meet that contour [15].

The most common active contour models are the snake model, gradient vector flow model and balloon model.

Snake Model

The snake model uses prior knowledge about the target contour. It applies splines to minimize energy by various image forces. A spline is a set of polynomials to derive geometric figures [15]. Spline of minimizing energy directs the constraint forces using internal and external image forces based on appropriate contour features.

Gradient Vector Flow Model

This model is an extension of the snake model, since it makes use of the gradient vector flow field as energy constraint to define the contour flow. Gradient Vector Flow (GVF) is determined by detecting the edge mapping function from the image and a energy functional equation based on it. GVF is used to replace the energy constraints in the traditional snake model [15]. With this constraints, the computation of the curve flow occurs iteratively, defining the contour.

Balloon Model

The balloon model locates an area in the volume and places an icosahedron that contains no points. The starting point of the icosahedron is inserted manually. The icosahedron is expanded or subdivided so that it approximates the volume [15]. Each vertice is connected to its neighbors using the pressure inside the volume, developing a contour.

These methods have the advantage of capturing the local shape features. However, due to the fact that the contour is driven by a predefined segmentation energy functional, its convergence is local, causing sensitiveness to the initialization. Moreover, the shape of the object may not be well preserved [16].

2.4.7 Graph Based

Based on graph portioning, these methods see the image as a graph G, where the pixels are vertices and the weight is determined based on the vertices it relates as edges. Graph based segmentation is similar to finding a set of sub-graphs $SG_1, SG_2, \ldots, SG_n$ from the graph G, such that for all $k \in 1, 2, \ldots, n, \forall_{i,j}, i \neq j, v_i, v_j \in SG_k$ with walks between v_i and v_j [17]. Each sub-graph is composed by a collection of vertices with strong connections between them.

The graph based methods might be grouped as Graph Cut Based Methods, Interactive Methods, Minimum Spanning Tree Based Methods and Pyramid Based Methods.

Graph Cut Based Methods

This algorithm parts the graph into two disjointed components. Graph cut is the total weight of the discarded edges, which infers the degree of association between the two components. The segmentation is achieved by suitably and repeatedly partitioning the graph constructed from an image using the graph cut [17].

Interactive Methods

In situations where automatic segmentation is difficult and can not guarantee accuracy and precision, interactive methods are a good option. They consist of getting the user preferences

and generating an optimal solution according to it.

Minimum Spanning Tree Based Methods

A spanning tree of a connected undirected graph is a subgraph which links all the vertices of the graph, creating exactly a single path between any two vertices [17]. Minimum Spanning Tree (MST) is a spanning tree whose total weight of edges is less than or equal to the total weight of edges of every other spanning tree. In segmentation context, MST represents the possible weakest connections. By suitably removing the lowest weighted edges, the different partitions with stronger inherent connections can be found.

Pyramid Based Methods

From the original image graph, a set of graphs defined in multi-level of resolution is built, similar to a pyramid. The vertices and edges at level $L+1$ are computed from the reduction of vertices and edges at level L by a reduction function. A level of pyramid, called working level, is chosen as the one to provides the segmentation.

Due to the difficulty of defining a good partitioning, image segmentation with Graph Based algorithms is a challenging problem [17]. In spite of that, the graph based methods have the advantage of segment unique and continuous boundaries from an image [18].

2.4.8 Clustering Algorithms

Clustering algorithms are unsupervised methods, which means that it is a machine learning technique that discovers patterns and information on unlabelled images (training data). It iteractively alternates between segmenting the image and characterizing the properties of each class.

There are three clustering algorithms often used: the K-means algorithm, the Fuzzy C-means algorithm and the Expectation Maximization (EM) algorithm.

K-means

K-means starts by choosing K clusters and, consequently, selecting K random centroids. Each pixel of the image is assigned to the closest centroid. For each cluster, new centroids are computed based on the average of the pixels position. The process repeats until the centroids do not change.

Fuzzy C-means

The Fuzzy C-means algorithm is a generalization of the K-means algorithm. The algorithm classifies the image by grouping similar data points in the feature space into clusters. This clustering is achieved by iteratively minimizing a cost function that is dependent on the distance of the pixels to the cluster centers in the feature domain [19]. Unlike the K-means algorithm, where the data points exclusively belong to one cluster, in this case, the data points can belong to more than one cluster with a likelihood.

Expectation Maximization

This algorithm works with the same principles but there is the assumption that the data follows a Gaussian mixture model. The first mode attempts to estimate the missing or latent variables, called the estimation-step or E-step. The second mode attempts to optimize the parameters of the model, called the maximization-step or M-step [20]. It is the one with the greater sensitivity of the three.

These algorithms are simple to implement and computationally fast methods but they have the disadvantage of having to manually choose the number of clusters.

2.4.9 Classifier Based

Classifier methods are used to predict the class of a given pixel by approximating a mapping function. It consists in pattern recognition techniques that seek to a feature space derived from the image. A feature space is the range of any function of the image, the most common one is the image intensity itself [12]. These algorithms require the manually segmented structures to use as training data. Through the training phase, the parameters of the function are adjusted

and then, to evaluate the performance, the function is applied on the new data (testing phase).

Classification Trees

Classification trees develop classification models in the form of a tree structure. It is based on an if-then rule set. The rules are learned sequentially using the training data one at a time until meeting a termination condition, as represented in Figure 2.6.

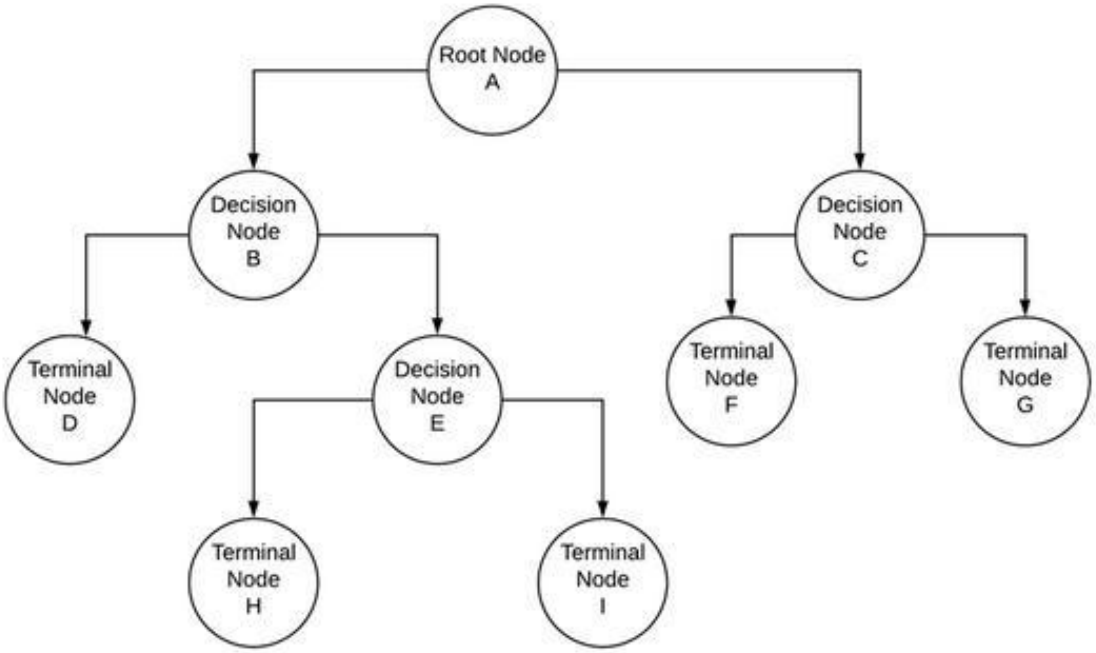


Figure 2.6: Structure of a classification tree. Figure from [21]

This classifier can be easily over-fitted, generating many branches and may reflect anomalies due to noise or outliers, witch takes to a poor performance on the unseen data even though it gives an impressive performance on training data. This can be avoided by pre-pruning, which halts tree construction early, or post-pruning, which removes branches from the fully grown tree.

Discriminant Analysis

Discriminant Analysis works based on the assumption that different classes generate data based on different Gaussian distributions. To train the classifier, the fitting function estimates the parameters of a Gaussian distribution for each class. If the purpose is to predict the classes of new data, the trained classifier finds the class with the smallest misclassification cost.

k-Nearest Neighbors

KNN algorithms classify new data points based on a distance function, assuming that closer points (neighbors) are similar. The data is assigned to the class which has nearest neighbors from the k clusters. As you increase the number of nearest neighbors (the k value) accuracy might increase.

This algorithm has a simple application. However, it has the disadvantage of needing to find the optimal k.

Naive Bayes

The Naive Bayes classifier is based on the simplifying assumption that the attribute values are conditionally independent given the target value [22]. This means that the probability of observing the conjunction is just the product of the probabilities for the individual attributes, given the target value of the instance. Therefore, this classifier ignores the possible dependencies, namely correlations, among the inputs.

Support Vector Machine

SVM classifies data by finding the best hyperplane that separates all data points of one class from those of the other class. The best hyperplane for an SVM means the one with the largest margin between the two classes. Margin means the maximal width of the slab parallel to the hyperplane that has no interior data points.

The support vectors are the data points that are closest to the separating hyperplane. These points are on the boundary of the slab. Figure 2.7 illustrates these definitions.

Ensembles Classifcation

Ensemble combines basic models in a strategic manner to achieve better accuracy rates. The key objective of the ensemble methods is to reduce bias and variance. Diversity, combination methods and selection topology are among the main factors to determine the ensemble performance. Consequently, it is a challenging task to design an efficient ensemble scheme.

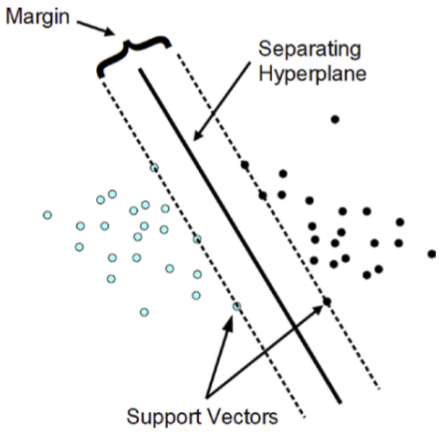


Figure 2.7: SVM representation: Figure from [23]

Classification Tree Ensembles

Classification Tree Ensembles, also known as Tree Bagger, bags an ensemble of decision trees. Bagging stands for bootstrap aggregation, which is a learning method that is commonly used to reduce variance within a noisy data set. Every tree in the ensemble are grown on an independently drawn bootstrap replica of input data. Observations not included in this replica are "out of bag" for this tree. Individual decision trees tend to overfit. To reduce the effects of overfitting and improve generalization, Classification Tree Ensembles combines the results of many decision trees.

RUS Boost

RUSBoost is an algorithm to handle class imbalance problem in data with discrete class labels. It uses a combination of Random Under-Sampling (RUS) and the standard boosting procedure AdaBoost to better model the minority class by removing majority class samples. This method results in a simpler algorithm with faster model training time.

Artificial Neural Networks

Artificial Neural Networks (ANN) are computational networks which attempt to simulate the decision process in networks of neurons of the human central nervous system (Figure 2.8). In the biological process, a dendrite first receives some sort of impulse. That impulse is then translated across an axon and finally released at an axon terminal. These three steps are similar

to the three types of layers that comprise an artificial neuron. The first one is the input layer in which the values used to predict are received. Then, there are one or more hidden layers, where is an activation function that will transform the input. The last one is the output layer that gives the output result in the expected format.

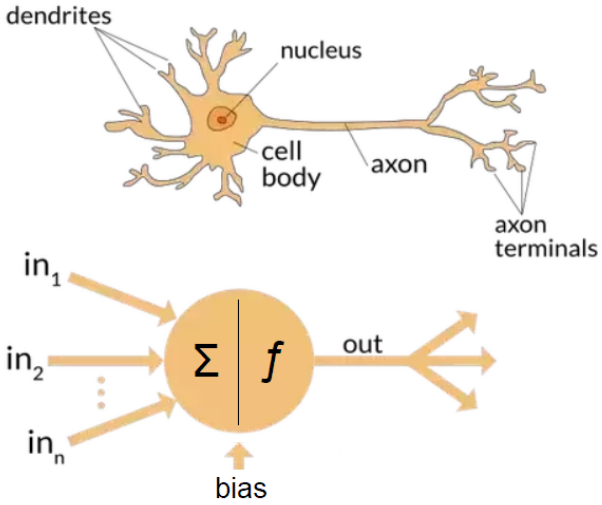


Figure 2.8: Artificial Neuron vs Biologic Neuron. Figure from [24]

The data input are features vectors assigned with a weight each. These weights will be adjusted through the entire process in order to minimize the error between the expected and the obtained output of the training samples. In the hidden layer, the weighted sum of all the features is passed through the activation function plus an optional bias. If there are other hidden layers, the output is fed to another neuron. If there are no more layers, the last one is the output layer and it is the network result. At last, the ANN is tested in a new data set.

Deep Learning

Due to the computer power increase and large amount of available data, it is now possible to use larger networks based on ANN. Deep learning models are a set of methods with multiple levels of representation that allows a machine to be fed with raw data and to automatically discover the representation needed for detection or classification [25]. Deep learning has become an increasingly popular approach in the last decade. This success came from its internal representation in the form of high-level features, which allows the modelling of complex problems and a smart initialization of some other deep structures [26].

The deep learning models enables computers to learn from experience and understand the

world in terms of a hierarchy of concepts [27]. Therefore, there is no need for a human computer operator formally to specify all of the knowledge that the computer needs. It works by training on a large of input data set and uncovering complex patterns in data. Due to today's computing hardware, it generates predictions quickly and accurately. However, the impact of deep learning in medicine is still limited due to its complexity and lack of interpretation. The scarcity of labelled medical images is another problem, since it can lead to overfitting and hard parametrization. To attenuate this, there are data geometric augmentation, transfer learning and fine-tuning [26].

Convolutional Neural Networks

The most popular deep learning model in medical image analysis are Convolutional Neural Networks (CNN). The architecture of a typical CNN is structured as a serie of stages (Figure 2.9). The first ones are composed of three types of layers: convolutional layers, pooling layers and fully connected layers. In convolutional layers, local features are detected in different positions in the image and then a series of convolution filters and kernels are applied, which output various features maps. By applying an activation function to these features, it is introduced the non-linearity property to detect non-linear features. In every convolutional layer, this process is repeated. The pooling layers are interspersed with the convolutional layers to reduce the feature maps dimensionality, generally using the max pooling or average pooling operations. At last, the fully connected layers combine the feature maps in feature vectors. The output is generated by a softmax function.

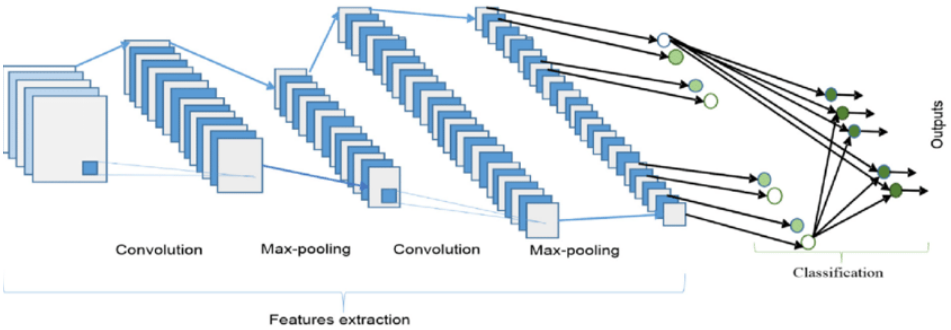


Figure 2.9: Convolutional neural networks architecture. Figure from [28]

By rewriting the fully connected layers as convolutions, the CNN can take input images larger

than it was trained on and produce a likelihood map, rather than an output for a single pixel [29].

U-Net

In biomedical image processing, the desired output should include localization, this means that a class label is supposed to be assigned to each pixel [30]. For that purpose, U-Net predicts the class label of each pixel by providing a local patch around that pixel as input.

The U-Net architecture consists of a contracting path followed by an expansive path (Figure 2.10). The contracting path has the same architecture of a CNN: convolutional layers interspersed with max-pooling layers. At each downsampling step, the number of feature channels is doubled. In the expansive path, an up-convolution operation is the first to be applied, which consists of an upsampling of the feature map followed by a 2×2 convolution, halving the number of feature channels. Then, there is a concatenation with the correspondingly cropped feature map from the contracting path. It goes through two 3×3 convolutions, followed by a ReLU each. At last, a final layer 1×1 convolution is used to map each component feature vector to the desired number of classes. This network has a total of 23 convolutional layers.

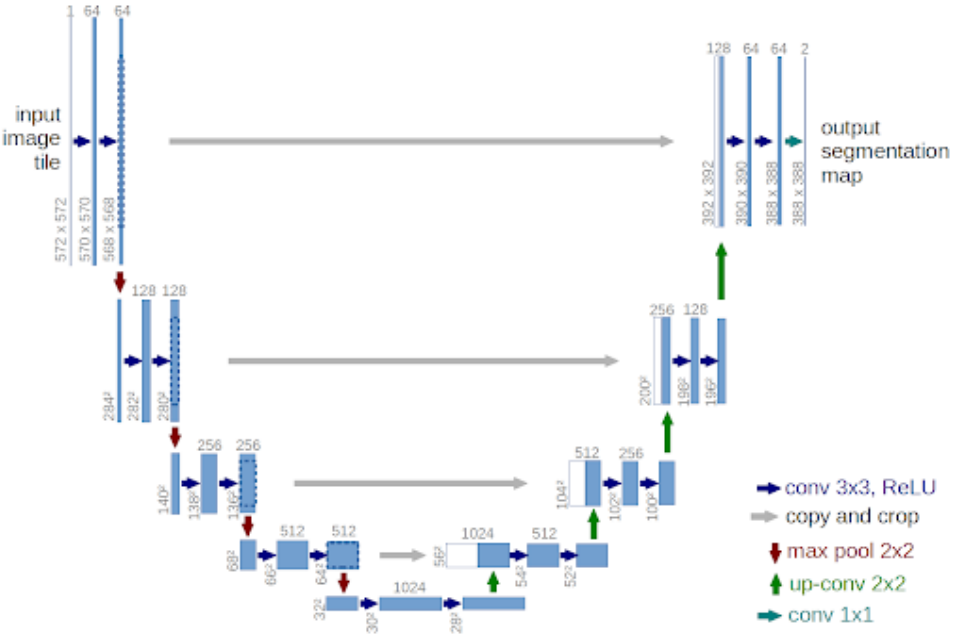


Figure 2.10: U-Net architecture. Figure from [31]

The two main advantages of this network are the capacity of localize and the training data in terms of patches is much larger than the number of training images. However, it also has two

disadvantages: it is quite slow, since the network must be run separately for each patch and there is a trade-off between localization accuracy and the use of context. It is due to the fact that larger patches require more max-pooling layers, which reduces the localization accuracy, while small patches allow the network to see only little context [30].

Classifiers are relatively computationally efficient. However, they have the disadvantage of requiring manual interaction to obtain the training data, what could be time consuming and laborious.

2.5 Evaluation metrics

There are several evaluation metrics with different approaches. In this study, overlap based metrics will be used. The main metrics from this category can be derived from the four basic cardinalities of the confusion matrix [32]: True Positives (TP), False Positives (FP), True Negatives (TN) and False Negatives (FN). TP and FP represent the voxels correctly and incorrectly classified as belonging to the Ground Truth (GT), respectively. TN and FN constitute the voxels correctly and incorrectly classified as not belonging to the GT, respectively.

2.5.1 True Positive Rate (TPR)

True Positives Rate (TPR), also called Sensitivity, Recall or Probability of Detection (POD), computes the portion of positive voxels in the ground truth classified as positive.

$$TPR = \frac{TP}{TP + FN} \tag{2.7}$$

2.5.2 True Negative Rate (TNR)

True Negatives Rate (TNR), also called Specificity, measures the portion of negative voxels (background) classified as negative.

$$TNR = \frac{TN}{TN + FP} \tag{2.8}$$

2.5.3 False Positive Rate (FPR)

False Positives Rate (FPR), also called Fallout, provides the portion of negative voxels classified as positive.

$$FPR = \frac{FP}{FP + TN} = 1 - TNR \tag{2.9}$$

2.5.4 False Negative Rate (FNR)

False Negatives Rate (FNR), also called Probability of False Detection, gives the portion of positive voxels classified as negative.

$$FNR = \frac{FN}{FN + TP} = 1 - TPR \tag{2.10}$$

2.5.5 Positive Predictive Value (PPV)

Positive Predictive Value (PPV), also called Precision, represents the portion of voxels correctly classified as positives.

$$PPV = \frac{TP}{TP + FP} \tag{2.11}$$

2.5.6 BF score

BF score, also called F-Measure (FMS) or harmonic mean, relates the PPV and the TPR:

$$BFscore = \frac{2PPV \times TPR}{PPV + TPR} \tag{2.12}$$

2.5.7 Dice Similarity Coefficient (DSC)

Dice Similarity Coefficient (DSC), also called Overlap Index, is the most used metric in validating medical volume segmentation [32]. It can be defined as:

$$DSC = \frac{2TP}{2TP + FP + FN} \tag{2.13}$$

2.5.8 Jaccard index (JAC)

Jaccard Index (JAC) is similiar to DSC but penalizes the wrong predictions more:

$$JAC = \frac{TP}{TP + FP + FN} \tag{2.14}$$

2.5.9 Accuracy

Accuracy is the proportion of true results among the total number of cases examined:

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \tag{2.15}$$

2.5.10 Other Evaluation Metrics

The following metrics were used by different authors in the State of the Art, to evaluate their models.

- **Hausdorff Distance (HD)** - maximum distance between two volumes;
- **Centroid Distance** - distance between two centroids (a pixel group's centroid of a binary image is the coordinate mean);
- **Average Distance (AVD)** - equivalent to HD averaged over all points;
- **Average Surface Distance (ASD)** - evaluates how much the surface changes between the segmentation and the ground truth, on average;
- **Surface Distance Error (SDE)** - difference between the surface of the segmentation and of the ground truth;
- **Average Error** - average of all the differences between the segmentation and the ground truth;
- **Average Volume Error (AVE)** - average of the differences between the volume of the segmentation and of the ground truth;
- **Average Absolute Volume Error (AAVE)** - difference between the segmentation and the ground truth;
- **Percentage of Volume Overlap (PVO)/ Percentage of Volume Intersection/ Mean Conformity Index** - equivalent to DSC, reveals the segmentation's accuracy through the percentage of overlap between the segmentation and the ground truth.

Chapter 3.

State of the Art

A search about the most common algorithms published between 2004 and 2021 was made. In Section 3.1, the segmentation algorithms applied to organs in context of prostate cancer are described. Then, in Section 3.2, the study is specifically for bladder segmentation algorithms. Some conclusion are described in Section 3.3.

3.1 Segmentation in Context of Prostate Cancer

In 2004, Tang et al. applied geometric model-based techniques that consists in exhibiting non-rigid deformation to segment the prostate and surrounding structures using 5 CT scans. The model learns the mean shape, regions of interest's local appearance and also the most typical object deformations. Principal Component Analysis (PCA), a statistical analysis tool that transforms a number of correlated variables into uncorrelated variables, was used. The next step is fit the created model to a new image using an algorithm that, for each landmark point of each training images, a $n \times n$ window samples the gray level values. Then, the computation time and the modeling accuracy are optimized to choose the neighborhood size n. This model gives excellent matches to the prostate and surrounding structure and convergence is declared after 10 iterations. The maximum and minimum distance is 3.6207 and 0.2150 pixels, respectively. The average distance between the real landmark points and the fitting results calculated is 1.3174 pixels and with variance 0.6713 pixels. In 256×256 images, the mean distance between the hand segmented and the automatically estimated contours are 1.5 pixels (2.44 mm), with variance about 0.6 pixels (1.24 mm).

One year later, Rousson et al. jointly segmented the prostate and bladder. They design an approach that, in a probabilistic way, includes a coupling between the organs and a shape model of the prostate. Using the posterior density probability of the bladder and prostate segmentation to apply the Bayesian theorem, they realise that the optimal solution should minimize the energy. The user clicks inside each organ and then the prostate and bladder boundaries are initialized as small spheres centered on these two points. Due to the fact that the intensity of each organ is relatively constant, its mean value should be guessed with a good confidence and there is not a big sensitivity to user inputs. Since the bladder's shape suffers some variations, the shape model is not adequate, so the coupling is necessary to the segmentation of this organ. Besides that, coupling makes the initialization and the image quality more robust. Applying this approach on 4 patients with 4 images each, the average of the probability of detection, the probability of false detection, the centroid distance and the ASD are 0.84, 0.21, 5.2, 4.2 respectively.

In order to improve the time-consuming of manual delineation, in 2008 McBain et al. presented a software tool named SCULPTER (Structure Creation Using Limited Point Topology Evidence in Radiotherapy). This software is connected to a clinical database. Using a range of magnification and windowing tools, it is possible to visualize scan images. Initially, while SCULPTER was being refined, the images were viewed in the transaxial plan, in a 3×3 mosaic. Using a higher magnification to display a single screen image, the detail was improved, allowing the visualization in other planes or in 3D. To create a structure, the image set and the most superior and inferior slices were examined. Then, three slices were selected: one near the superior aspect, another one near the middle and the last one near the inferior aspect. To complete the process, the support points (anatomically distinct boundary points) were defined on each of the three slices by mouse clicks. The users classified SCULPTER as a software easy to use and apply. It was possible to contour the bladder in 6 MR scans and 10 CT scans but not in a female patient who had a collapsed bladder. SCULPTER delineations closely reproduced manual contours with no significant volume differences detected but they were significantly quick to obtain ($p < 0.05$) in most cases.

In the same year, Merck et al. jointly estimate the best geometric model for any given image

and shape distribution for the entire population of training images to segment . A collection of CT scans from a single patient was used. To initialize the training process, a model was hand manipulated to represent the shape that is pretended to be trained. Optimizing the model parameters according to a metric that measures the goodness of a model fit to a given image, the best deformation of reference shape is computed. Then, the model with the greatest probability density is chosen. At last, Statistical Deformable Shape Model (SDSM) was used as the basis of numerical methods for shape discrimination, comparison and interpolation for longitudinal shape studies and for deformable image segmentation. M-reps and other representations with explicit orientational components are governed by a Principal Geodesic Analysis (PGA), that is a generalized PCA. PGA reduced the complexity of the optimization and allowed the reconstruction of over 90% of the shape variability using only five model parameters, instead of 100. The average error across all organs is less than a voxel. 95% of the anatomic variability was cover by 15 models of variance. Their training resulted in 95.3% volume overlap.

An auto-segmentation rule/logic-based algorithm was used by Huyskens et al., in the next year, to segment the prostate, the bladder, rectum and femur heads. The proposed algorithm consisted of three parts. The first one is the pre-segmentation, where the data is segmented in three classes: body, bone and air/lung equivalent tissue. In the second one it is the discovery of anatomic orientation using anatomical reference points, which provides the patient's position. The last part is the structure segmentation module. Since the filling of the bladder determines its shape and affects the prostate's position, their localisation and outlining were performed together in a six steps procedure: pre-defining the shape to create a boundary for the prostate, pre-filter this region using a noise filter, determine the intensity range by means of a sagittal line profile through the structures, generate an intermediate prostate outline by flood fill, repeat the last step for the bladder and, to get the final volumes, these intermediate outlines are subjected to 3D refinement. A different approach was used for the rectum. The outline was made slice by slice by flood filling the space inside surrounding fat tissue and neighboring structures. The femoral heads are modelled as spheres with a fixed radius. This algorithm was qualitative evaluated for 44 patients' CT scans. In 5 of them, the algorithm misses at least one structure. The clinicians evaluate the segmentations through scores. The mean scores for the prostate

were between good (45%) and acceptable (30%). For the bladder they were between excellent (36%) and good (42%). For the rectum they were between acceptable (27%) and not acceptable (45%). For the femoral heads, the opinions were divided to be good (27%) or not acceptable (54%).

In 2010, Acosta et al. investigated the use of atlas based methods to perform the non-rigid mapping and segmentation of the rectum, bladder and bones from 19 CT scans from patients treated for prostate cancer by radiotherapy. The training data is formed by the OR manually segmented by an expert. Through a number of affine and non rigid registration iterations, an average image (template) was obtained to represent the whole population. Probabilistic maps for each organ were generated by the amount of consensus between labels. The accuracy of the approach was validated by segmenting the organs using the training data in a leave one out scheme. A good agreement with this approach was obtained. The main cause of error in the automatic segmentation results is related to organ variation, particularly with the CTV1 (prostate and seminal vesicles), and the prostate. Besides that, obesity appears to be a source of error, as it induces some variability in the training data set. The average DSC of the bladder, rectum, right femur, left femur, CTV1 and CTV2 (prostate) have the following values, respectively: 0.636, 0.584, 0.851, 0.834, 0.564, 0.583.

One year later, Acosta et al. evaluated a multi-atlas-based approach to simultaneously segment prostate, bladder and rectum from CT scans of 24 patients treated from prostate cancer. First, affinely registered atlas are ranked on three different metrics: Sum of Squared Differences (SSD), Cross Correlation (CC) and Mutual Information (MI). Then, labels from the top n ranked atlas were propagated using the non-rigid registration to the individual's CT. At last, in order to obtain single segmentation for each organ, a decision rule was applied: a majority voting rule and, for the validation of the segmented organs, simultaneous truth and performance level estimation (STAPLE) algorithm. The non-rigid registration provides a significant improvement in the overlap: 23.2% for the prostate, 24.8% for the rectum and 35.0% for the bladder.

Li et al., in 2012, evaluated the performance of atlas-based auto-segmentation applied to high quality verification CT-imaging using a CT-on-rail system for prostate cancer. They use CT

scans from 7 prostate cancer patients with the bladder, rectum and prostate manually contoured. Three patient specific atlases were generated consisting of one, four and seven prior images and contour sets for each patient. Then, the auto and manual contours were compared geometrically and dosimetrically. The DSC for the bladder were above 91% with 1-image set atlas. For the rectum and the prostate, the values were greater than 81%. There was an improvement of the consistencies by including 4 images sets in the atlas. Patient specific atlas yielded more accurate contours when compared with non-patient specific atlas.

Later, in 2016, Lay et al. applied integration of both local and global information for multiple organ segmentation in two different data sets. First, their approach was tested on a set of lungs, heart, liver and kidneys in MR localizer scans. A total of 185 volumes was split into a training set of 135 and a test set of 50. In the second data set, the goal was to detect the prostate, bladder, rectum and femoral heads in CT scans. There were 145 volumes, of which 100 were selected for training and 45 for testing. They started by integrating local and global image context using a product rule into one posterior probability. Computing the expected landmark location through Mini Mental State Examination (MMSE) and MAP, it was possible to compute the nearest neighborhood from the training database. Then, it was taken a similar approach to local sensitive hashing and multiple hash indexes were built on the data. It was used a Binary Space Partitioning (BSP) tree, instead of a hash function. At each node of the BSP tree, it was chosen a random hyperplane to split the data. For the lungs, heart, liver and kidneys, it was noticed that MAP estimations were better than the MMSE's. The accuracy of the posterior global context suffers from sparse sampling. Even with dense sampling, its performance is worse than the Local + Global method. This method was compared with a learning-based approach to segment single organs in 3D volumes called Marginal Space Learning (MSL). For the prostate, bladder, rectum and femoral heads, it behaved similarly to or better than MSL, except for the bladder. These two approaches achieved accuracy very similar, with the exception of the rectum, since its shape varies a lot.

In the same year, Zhou and Xu presented different approaches for segmentation of multiple organs with significant shape and appearance variations and stringent accuracy and speed requirements. They used hundreds of CT scans from different clinical sites. The process was

initialized with multiple landmarks detection, then the organ was estimated and they finished by segmenting multiple organs. The accuracy increases as the segmentation's stage proceeds. The segmentations were evaluated by the metrics ASD and DSC and the results are noted in Table 3.1.

Table 3.1: The CT Multi-Organ Segmentation Results in terms of ASD and DSC. Based on the results from [42]

Structure	ASD	DSC
Liver	2.19	0.94
Spleen	2.37	0.90
Left lung	1.06	0.97
Right lung	1.17	0.97
Left kidney	1.36	0.93
Right kidney	1.43	0.93
Prostate	2.84	0.76
Bladder	2.39	0.87

In the next year, Gordon et al. applied Deep Learning Convolutional Neural Network (DL-CNN) to segment the bladder wall. Through 94 cases from the Institutional Review Board, they trained the DL-CNN in order to distinguish the areas with the bladder wall, using neighborhood information. The central pixel of the RoI is located between the manually outlined inner and outer bladder walls. Given that, half of the RoIs were extracted from inside the bladder wall and the other half outside the bladder wall. Then, to refine the contours, level sets were applied to the Computerized Tomography Urogram (CTU) scans and bladder wall likelihood. It was shown that the DL-CNN with level sets can segment the inner and outer wall of the bladder with efficiency. For the training set, the inner and the outer wall achieved a volume intersection of $90.0 \pm 8.7\%$ and $93.7 \pm 3.9\%$. For the test set, the results were $87.6 \pm 7.6\%$ for the inner wall and the $87.2 \pm 9.3\%$ for the outer wall.

In 2019, Dong et al. developed a method using cycle consistent deep attention network to segment the bladder, the rectum and the prostate. The Synthetic Magnetic Resonance Imaging (sMRI) yields soft-tissue information to aid the CT segmentation. The sMRI was estimated from CT images through a Cycle Generative Adversarial Network (CycleGAN). In order to

distinguish different organs, a Deep Attention U-Net (DAUnet) was used to train for auto-segmentation on sMRI and on the corresponding multi-organ contours. Therefore, it was possible to obtain the new patient's contours by subjecting the CT image to the CycleGAN, generating sMRI and then segmenting the organs using DAUnet. This approach was applied to 140 data set from patients with prostace cancer. To evaluate the model, the DSC and the mean surface distance were computed: the DSC for bladder, prostate and rectum were 0.95, 0.87 and 0.89, respectively; the mean surface distance for the same organs were 0.52, 0.93 and 0.92. The results showed that applying sMRI made the DSC increase and the mean surface distance decrease.

Schreier et al. in 2020 studied how to delineate organs automatically. Using 300 CT scans originating from 4 different clinics located in Europe and North America, they presented a novel deep neural network called BibNet that trains on CT and artificially generates pseudo Cone Beam Computed Tomography (CBCT) to segment bladder, prostate, rectum and seminal vesicles. CT scans use fan-shaped x-ray beams while CBCT uses a cone-shaped area detector that does not require patient movement. The results showed that the model's efficiency is either equally good (prostate and seminal vesicles) or better (bladder and rectum) than the structures from the routine clinical practice.

Also in 2020, Sartor et al. used manually segmented CTs to train and evaluate a CNN. The method starts by pre-processing the scans to remove the noisy pixels: the input values of HU were set in $[-1, 1]$. The largest connected component was kept. Then, it was applied a morphological hole filling to the resulting segmentation. Using negative likelihood, the CNN was trained on 226 manually segmented CT from 75 cervical cancers and 191 anorectal cancers receiving radiation therapy at Skåne University Hospital. For anorectal cancer, median DSC and mean surface distance scores were calculated: 0.91-0.92/1.93-1.86 for the femoral heads 0.94/2.07 for the bladder and 0.83/6.80 for the bowel bag. For cervical cancer the results were 0.93–0.94/1.42–1.49 for the femoral heads, 0.84/3.51 for the bladder, 0.88/5.80 for the bowel bag and 0.82/3.89 for the CTVNs.

In the same year, Rhee et al. also used a CNN to delineate structures divided in three groups:

bony structures, OR and CTVs. For train and validation data, it was used 2254 female pelvic CT scans from cancer patients from 2004 to 2018 at the University of Texas MD Anderson Cancer Center and 210 CT scans with kidney contours from the 2019 Kidney Tumor Segmentation Challenge. Initially, the classification architecture Inception-ResNet-V2 was trained to identify the structure's extent in the cranial-caudal direction. Then, the segmentation models were applied to the CT slices that contain the organ of interest. For the bony structures contours, it was applied a multi-atlas-based auto-contouring system. For the CTV and OR, there was the need for data augmentation techniques: resize the CT, crop the region around the segmented organ and re-segment the organ of interest on the cropped image. For these organs, it was used 3D V-Net segmentation models, with the exception for the spinal cord. This one was segmented by a 2D FCN-8s model. 80% of the CTVs, 97% of the OR, and 98% of the bony structure contours were clinically acceptable by a physician.

Also using a deep learning-based approach, Zhang et al. in 2020, segmented the bladder, the CTV and the rectum on CT scans from 91 patients with cervical cancer. For the pre-processing step, the CT scans were cropped in order to discard the regions depicting empty space or without labeled structures. Using linear interpolation, they were resized to an identical size. In order to improve the contrast, it was used a Contrast Limited Adaptive Histogram Equalization (CLAHE) algorithm. Then, a CNN architecture called DSD UNET was trained and tested. Applying it followed by a 3D skeletonization and a polynomial curve fitting, it was possible to reconstruct automatically the CTV and the OR. The results showed that DSD-UNET can perform a 3D U-Net in every structure. The mean DSC was 86.9% for the bladder, 82.9% for the CTV and 82.1% for the rectum.

In 2020, Wang et al. attempted to improve the organ boundaries. To that purpose, they presented a Boundary Coding Network (BCnet). This model is divided in two stages: boundary coding representation learning and organ segmentation. In the first one, two sub-networks supervised by the dilation and erosion masks transformed from the manually delineated organ mask were separately trained to learn the spatial-semantic context near the organ boundary. Given their predictions, the organ boundary was encoded and a multi-atlas based refinement strategy was designed. In the second one, the boundary coding was trained in order to achieve

the final segmentation network. This method was applied to 313 CT scans from 313 patients with prostate cancer from the North Carolina Cancer Hospital. To evaluate the quality of the model, the DSC, ASD, the PPV and sensitivity were computed for the prostate, bladder and rectum. In 2D, the DSC and the ASD presented the best results in comparison with previous methods, such as sensitivity, with the exception for the bladder. In 3D, the DSC and ASD had the second best results for prostate and rectum, and the sensitivity had the second best result for the bladder.

In 2020, Savenije et al. studied the clinical use of deep learning-based automatic OR (bladder, rectum and right and left femur) delineation on MRI. For all 150 patients with prostate cancer, it was acquired 3 Tesla MRI (3TMRI). It has a stronger magnet and creates better images of organs and soft tissue than other types of MRI do. The first 48 patients were included in a study that trains two 3D convolutional networks called DeepMedic and Dense V-net (dV-net). Using three-fold-cross-validation, the patiens were divided: 32 for training and 16 for validation. DeepMedic performed patch-based training and dV-net was responsible for the training on whole volumes. The first one, used a fully connected network to combine the low, medium and high-resolution pathways and post-processing. The OR were equally sampled. The second one is a 3D U-Net with a sequence of three downsampling and upsampling features to propagate higher resolution information to the final segmentation. Then, the performance of these two networks was compared with an atlas-based approach called advanced medical imaging registration engine (ADMIRE). For the bladder, significant differences were observed between the ADMIRE and the two networks. For the rectum, no differences were observed. DeepMedic had a better performance with respect to the femurs. For the DeepMedic, dV-net and ADMIRE, the percentages of delineation that were acceptable or needed small adjustment were 81%, 59% and 3%, respectively. After retraining DeepMedic and testing on the successive patients, the performances improved.

Aoki et al., in the same year, examined the utility of the deep learning-based algorithm in detecting bone metastases in patients with prostate cancer using bone scintigraphy, a technique that is often used to evaluate bone metastases. 139 patients were studied. The algorithm segments the skeleton and detects the hotspot extraction. The Butterfly-type Networks (Btrfly-

Nets) was employed and two U-Nets were fused into a single one that simultaneously processes anterior and posterior images. Every examination was evaluated by specialists and by the algorithm but there were no big differences between them. The sensitivity, specificity and accuracy were 100%, 94.9% and 97.1%, respectively, by the specialists. The same measures, made with the software were 91.7%, 87.3% and 89.2%.

3.2 Bladder Segmentation

A search of articles about bladder segmentation is useful to understand the most common and efficient techniques to segment this specific organ. Therefore, studies from 2009 to 2019 about this topic were analysed.

In 2009, Shi, Yang, and Zhu, by applying a hybrid scheme of automatic bladder segmentation, combined mean shift clustering, line by line scanning and rolling-ball filter. Mean shift algorithm was used to obtain a clustered image containing the rough contour of the bladder. Later, it was extracted by applying a region-growing algorithm with the initial seed point selected from a line by line scanning process. Applying the rolling ball algorithm, the bladder contour was refined more accurately. These steps are then extended to segment the bladder volume in a slice by slice manner. Seven training data sets of 98 CT scans were selected to determine the optimal values of parameters and the model was evaluated in 15 data sets containing 214 CTs. The average values of sensitivity, specificity, positive predictive value, negative predictive value and Hausdorff distance were 86.5%, 96.3%, 90.5%, 96.5% and 2.8 pixels, respectively.

Later, Chai et al. in 2012 presented an automatic bladder segmentation approach suitable for CBCT scans of 23 bladder cancer patients and test its ability to select the appropriate plan from such an Adaptive Radiotherapy Plan (ART) product. Each patient received one planning CT scan and 7-20 (average 11.6) CBCT scans. For each patient, all CBCT were matched to the planning CT on bony anatomy. Bladder contours were manually delineated for each planning CT and CBCT. A patient-specific bladder deformation model was built from the training data set. To model it, PCA was applied to the training data. The number of PCA modes for each

patient was chosen such that the bladder shapes in the training set could be represented by them with less than 0.1cm mean residual error. The automatic segmentation started from the bladder shape of a reference CBCT, which was adjusted by changing the weight of each PCA mode. Then, the segmentation contour was deformed consistently with the training set to fit the bladder in the validation image. To measure its goodness, a cost function was defined by the absolute difference between the directional gradient field of reference CBCT sampled on the directional gradient field of validation CBCT sampled on the segmentation contour candidate. It was minimized using a simplex optimizer. For each validation CBCT image, the segmentation was done five times using a different CBCT and the one with the lowest cost function was selected as the final bladder segmentation. It was observed that two of four PCA modes were needed to represent the bladder shape variation with less than 0.1cm average residual error for the training data of each patient. The automatically segmented bladders had a 78.5% mean conformity index with the manual delineations. The mean surface distance of the local residual error over all patients was 0.24cm.

Van De Schoot et al. in 2014 developed and validated a generic method for automatic bladder segmentation on CT and CBCT scans of 20 patients treated for tumors in the pelvic region, independent of gender and treatment position (prone or supine), using only pre-treated imaging data. The full and empty bladder contour were used to generate a patient-specific bladder shape model. The reference bladder contour was deformed iteractively by maximizing the cross-correlation between directional grey value gradients over the reference and CBCT bladder edge. Automatic adaptations were implemented in order to overcome incorrect segmentations caused by CBCT image artifacts. Locally incorrect segmentations could be adapted manually. After each adapted segmentation, the bladder shape model was expanded and new shape patterns were calculated for following segmentations. All available CBCTs were used to validate the segmentation algorithm. The mean DSC, mean SDE and mean SD of contour-to-contour distances between segmentations and manual delineations were computed and are represented in Table 3.2. Manual local adaptations and expanding the shape model improved the segmentation results significantly ($p < 0.01$), both based on DSC and SD of contour-to-contour distances.

In 2016, Cha et al. introduced a computerized system for bladder segmentation in 173 patients

Table 3.2: Mean DSC, mean SDE and mean SD of contour-to-contour distances between segmentations and manual delineations. Based on results from [54]

Patient	DSC	SDE	SD
Female prone	0.87	0.27	0.22
Female supine	0.85	0.28	0.22
Male prone	0.88	0.23	0.17
Male supine	0.89	0.21	0.17

undergoing CTU as critical component for computer-aided detection of bladder cancer. These 173 patients subsequently underwent cystoscopy and biopsy. 81 of these cases were designed as the training set and the other 92 cases as the test set. Of the 81 bladders, 42 contained focal mass-like lesions, 21 had wall thickening and 18 were normal. Of the 92 cases, 43 contained focal mass-like lesions, 36 had wall thickening and 13 were normal. Initially, a DL-CNN was trained to distinguish between the inside and the outside of the bladder. Then, the DL-CNN was used to estimate the likelihood of a RoI being inside the bladder for RoIs centered at each voxel in a CTU case, resulting in a likelihood map. To generate the initial contour for the bladder, thresholding and hole-filling were applied to the map. At last, it was refined by 3D and 2D level sets. The results of this approach were compared with the Haar-feature-based likelihood map and level set and with the CLASS with Local Contour Refinement (LCR) method by computing the Average Volume Intersection (AVI), AAVE, AVE, Average Minimum Distance (AMD) and Average Jaccard Index (AJI). They can be analysed in Table 3.3. DL-CNN achieved better segmentation performance while using a single input.

Table 3.3: AVI, AAVE, AVE, AMD and AJI of the DL-CNN, CLASS with LCR and Haar-feature based likelihood map. Based on the results of [55]

Model	AVI	AAVE	AVE	AMD	AJI
DL-CNN	81.9% ± 12.1%	10.2% ± 16.2%	14,0% ± 13.0%	3.6 ± 2.0 mm	76.2% ± 11.8%
CLASS with LCR	78.0% ± 14.7%	16.5% ± 16.8%	18.2% ± 15.0%	3.8 ± 2.3 mm	73.9% ± 13.5%
Haar-feature based	74.3% ± 12.7%	13.0% ± 22.3%	20.5% ± 15.7%	5.7 ± 2.6mm	66.7% ± 13.5%

Xu, Zhou, and Liu in 2018 generated a deep learning-based approach involving a convolutional neural network segmentation and a pre-processing method, called dual-channel pre-processing, to further advance the segmentation performance. The data set used consists of 124 CT image volumes from several different CT scanners from one hospital. 100 image volumes were chosen

randomly and used for training. The remaining 24 volumes compose the validations set. Each CT contains a complete anatomic structure of bladder and a segmentation mask corresponding to it delineated by radiation oncologists. The proposed pre-processing method was applied on the input CT image and obtained a dual-channel image. It consists of the CT image and a enhanced bladder density map. A CNN was exploited to predict a coarse voxel-wise bladder score map on this dual-channel image. At last, a 3D fully connected CRF-RNN refines the coarse bladder score map and produce final fine-localized segmentation result. This approach was compared with a V-net. The DSC of the model was 92.24%, which is 8.12% higher than the V-net's. The bladder's probability maps performed by their approach present sharper boundaries and more accurate localization compared with that of the V-net.

One year later, Ma et al. developed a U-Net based deep learning approach (U-DL) for bladder segmentation in 173 CTU as part of computer-assisted bladder cancer detection and treatment response assessment pipeline. Of those 173 cases, it included 81 in the training/validation and 92 in the test set. Of the 81 cases, 42 presented masses, 21 suffer from wall thickening and 18 had normal bladders. Of the 92 cases, 43 had masses, 36 presented wall thickening and 13 were normal. Keras with Tensorflow beckend were used to implement the neural network. To obtain the best structure for the bladder segmentation task, the structure and some parameters of U-Net were modified and adjusted. Different U-DL were designed and compared to segment the bladder: 2D U-DL and U-DL using 2D CT slices and 3D CT volumes, respectively, as input, U-DLs using CT images of different resolutions as input and U-DLs with and without automated cropping the bladder as an image pre-processing step. The AVI, AVE, AAVE, AMD, Average Hausdorff Distance (AHD) and AJI values of the best 2D and 3D U-DL model and with the baseline method, all for the same test set, are represented in Table 3.4.

Table 3.4: AVI, AVE, AAVE, AMD, AHD and AJI of the 2D and 3D U-DL models and the baseline method. Based on the results of [57]

Model	AVI	AVE	AAVE	AMD	AHD	AJI
2D U-DL	93.4 ± 9.5%	-4.2 ± 14.2%	9.2 ± 11.5%	2.7 ± 2.5mm	9.7 ± 7.6mm	85.0 ± 11.3%
3D U-DL	90.6 ± 11.9%	-2.3 ± 21.7 %	11.5 ± 18.5 %	3.1 ± 3.2mm	11.4 ± 10.0mm	82.6 ± 14.2%
baseline	81.9% ± 12.1%	10.2 ± 16.2 %	14.0 ± 13.0 %	3.6 ± 2.0mm	12.8 ± 6.1mm	76.2 ± 11.8%

With the goal of comparing the different bladder segmentation approaches, all the results are summarised in Table 3.5.

Table 3.5: Performance of the bladder segmentation works

Publication	Type of Image	Type of Segmentation	Performance
Shi, Yang, and Zhu, 2009 [52]	CT scans	2D	sensivity=86.5% specificity=96.3% PPV=90.5% NPV=96.5% HD=2.8
Chai et al., 2012 [53]	CT and CBCT scans	3D	PVO=78.5 % SD= 0.24cm
Van De Schoot et al., 2014 [54]	CT and CBCT scans	3D	DSC=0.87 % SDE=0.25 SD=0.20
Cha et al., 2016 [55]	CTU	2D and 3D	AVI=81.9% AAVE=10.2% AVE=14% AMD=3.6mm AJI=76.2%
Xu, Zhou, and Liu, 2018 [56]	CT scans	3D	DSC=92.24%
Ma et al., 2019 [57]	CTU	2D	AVI=93.4% AVE=-4.2% AAVE=9.2% AMD=2.7mm AIID=9.7mm AJI=85%
		3D	AVI=90.4% AVE=-2.3% AAVE=11.5% AMD=3.1mm AIID=11.4mm AJI=82.6%

3.3 Conclusion

In general, the obtained results with the different approaches were reasonable with the exception of the rectum segmentation by Lay et al. (2016) and Huyskens et al. (2009). The organs

variability is the principal cause of errors. According to Acosta et al. (2010), the obesity could also be a problem. This study shows that the most used approaches recently are neural networks. However, they have the disadvantage of the obligatory training phase.

Chapter 4.

Methods

This study is focused on the developing of different methods to segment the bladder, since it is an OR in context of prostate cancer. With these methods, the goal is to improve the challenges of manual segmentation.

This chapter starts with a description of the database in Section 4.1. Then, the algorithms explored are explained in Section 4.2 and the methodology used to evaluate them is presented in Section 4.3.

4.1 Database

The database was composed by 47 CT scans from 47 patients with prostate cancer provided by the Institute of Oncology of Porto (IPO). Each CT has the structures manually segmented by specialists, as it is represented in Figure 4.1. Since most of each CTs have more than one manually segmentation of the bladder, there are 71 bladders manually segmented. These manual segmentations are used as ground truth. The quantity of the structures manually segmented is given in Table 4.1.

For some methods, it was needed to split the data set into train, validation and test data sets. The train data set consists of 19 patients, the validation has 4 and the remaining 24 went to test data set. This leads to 37% of the structures to segment in the train data set, 8% in validation and 55% in the test.

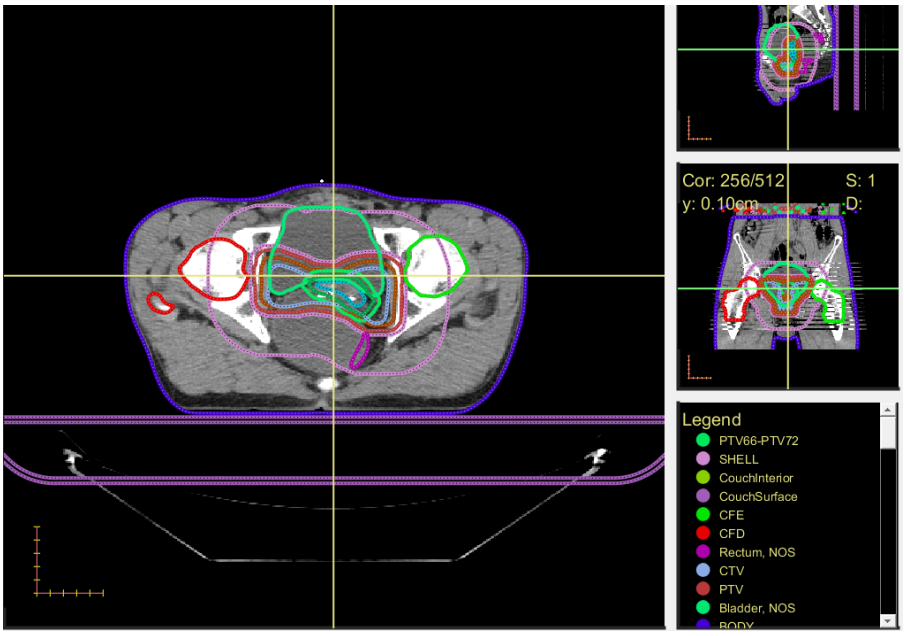


Figure 4.1: CT scan from one patient with structures manually segmented

Table 4.1: Quantity of the manually segmented structures

Structure	Quantity
PTV	111
CTV	59
GTV	8
Bladder	**71**
Rectum	67
Femoral Head Right	60
Femoral Head Left	60
Prostate	19
Bowel	10
Pelvis ganglia	10
Seminal Vesicle	7
Bone	5
Pelvis	2
Spinal cord	1
Urethra	1

4.2 Segmentation Techniques

In order to detect the RoI, two pre-procedures were developed: one by thresholding the CT scans based on its HU values and the other based in the anatomy. These pre-procedures are explained in Subsections 4.2.1 and 4.2.2, respectively. Every image went to these pre-procedures, creating two different masks. Once the RoI is detected, the algorithms were applied to both of the masks in order to segment the bladder. The algorithms studied and developed were Clustering, which is presented in Subsection 4.2.3, U-Net in Subsection 4.2.4, Active Contours in Subsection 4.2.5 and Graph Based in Subsection 4.2.6.

4.2.1 Thresholding

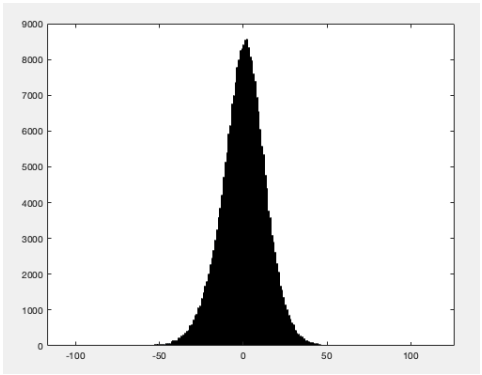
As a pre-procedure, it is intended to apply a threshold to each CT in order to get the RoI that the bladder belongs to, using its HU value. The bladder is typically full of water, whose HU value is zero. The goal is then to transform the CT data in HU and then apply a minimum and maximum threshold.

The CTs were converted to HU values using the linear transformation in Equation 2.1.

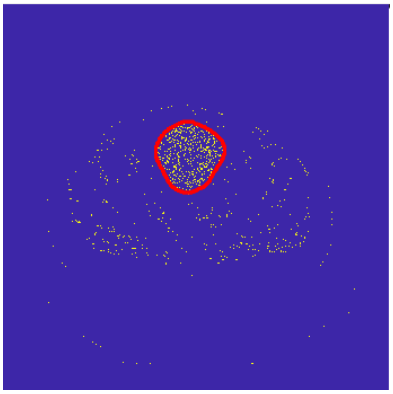
It begins to study the ideal thresholds by analysing the histogram of intensity values. Initially and ideally, they were centered in zero and the amplitude was four, but it only worked for 53% of the cases. For that reason, the mean and the standard deviation of the HU values of the ground truth was computed, and individual thresholds were obtained based on them. The mask obtained with these two approaches can be compared in Figure 4.2.

Using the individual thresholds, the mean of the minimum and maximum thresholds were computed. The final values were -3 and 45. Those were the values used to threshold each case.

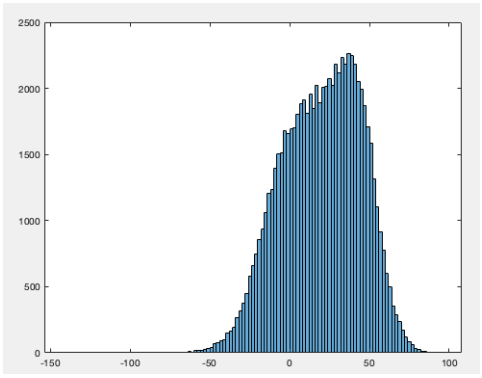
From each region with bladder patches, the biggest connected region was chosen. Each one of these regions suffered morphological operations. Using flat morphological structuring and dilation, a 3-D spherical structuring element was created whose radius is 15 pixels and multiple dilations of the most compact component were performed, using each structuring element in succession. In Figure 4.3, it is represented a slice with the final mask applied and an isosurface



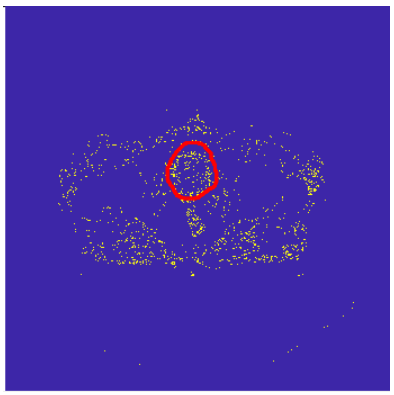
Histogram with m=0 and s=15



Thresholding centered in 0



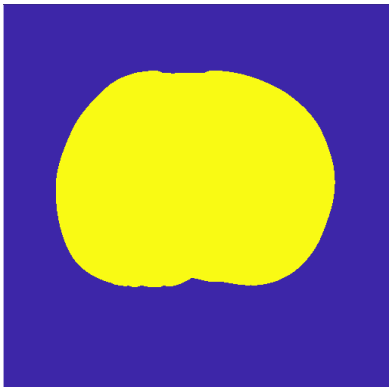
Histogram with m=19 and s=5



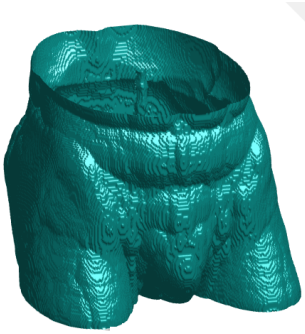
Thresholding based on GT

Figure 4.2: Creating mask based on HU values. m is the mean and s is the standard deviation. The red line is the GT

representation.



Slice with the final mask

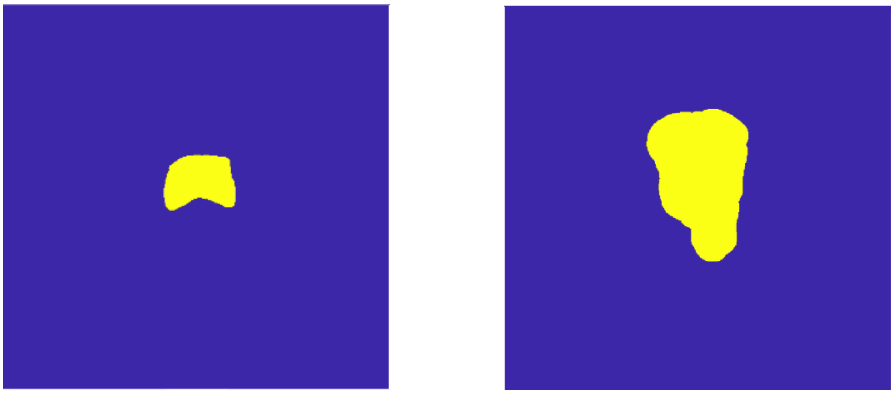


Final mask (isosurface representation)

Figure 4.3: Mask based on HU values. The image on the right is empty since the isosurface without the interior was drawn

4.2.2 Anatomic Mask

Another process to detect the RoI was approached. First, the sizes of each volume were obtained and a mask was created with the biggest dimensions. The sizes of each ground truth were found and defined in the center. Since it was applied to each case, the RoI was constantly updated originating a final one in which every contour is contained in. The procedure is represented in Figure 4.4.



GT of one case defined in the mask GT of every cases defined in the mask

Figure 4.4: Creating mask based on anatomy

At last, each mask was adapted to the respective size of the volume. The mask applied to each volume is represented in Figure 4.5.

Slice with the final mask Final mask 3D representation

Figure 4.5: Mask based on anatomy

4.2.3 Clustering

The clustering algorithm used was K-means (Algorithm 1). From each case, K random points, named centroids, are selected and the other points are assigned to the closest centroid. For each cluster, new centroids are defined based on the average of the points position. This process is repeated until the centroids do not change.

Algorithm 1 K-means algorithm

Require: Volume with n elements, V

Number of clusters, K

Ensure: Label Matrix, L

Select K random centroids from $V, c_1, \cdot, c_k$

while $A_i \leq c_i$ **do**

for i from 1 to K **do**

Set $A_i = 0$

for m from 1 to n **do**

Compute $d_{m,i}$, distance between v_m and c_i

Find $d_{m,j}$, the minimum $d_{m,i}$

Set $v_m = c_j$

end for

Compute A_j, the average of $v_n = c_j$

Set $A_j = c_j$, the new centroid

end for

end while

As represented in Figure 4.8, it started by choosing the optimal K. K values between 5 and 195 were tested by steps of 10 in the train data set to segment the bladder in each CT. To each K, it was selected the region with the biggest intersection area with the ground truth. Each of these segmented regions were evaluated with Jaccard Index and the K chosen was the one whose region achieved the best result.

Having set the best value for K, each volume is segmented into K regions. In Figure 4.6, the

division in 115 clusters can be analysed.

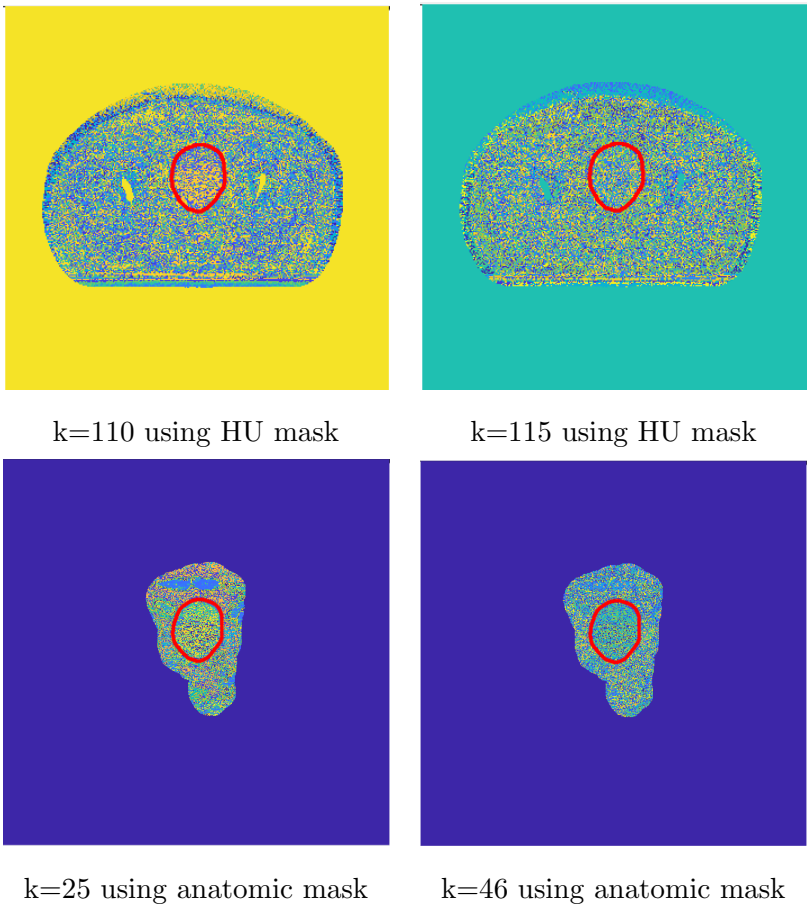


Figure 4.6: Segmentation into K regions. The red line is the GT. The background of each image is empty (classified as zero).

The next step is to choose one of these regions to be the one corresponding to the bladder. With this purpose, 5 features were extracted from each cluster: volume, diameter and the maximum, minimum and mean intensity. The train data set was trained by several classifiers: Classification Trees (ClassT), Discriminant Analysis (DA), k Nearest Neighbors (kNN), Naive Bayes (NB), Support Vector Machine (SVM), Classification Ensembles (CE), Classification Tree Ensembles (CTE) and RUSboost.

Two different post-processing were applied to the selected cluster. For both approaches, the background was removed. In the first one, the regions and holes were filled by implementing flood fill. The second approach, similarly to the detection of the RoI, consists of choosing the biggest connect region. The post-processing result of the previous cases can be visualised in Figure 4.7.

This entire process is described in Figure 4.8.

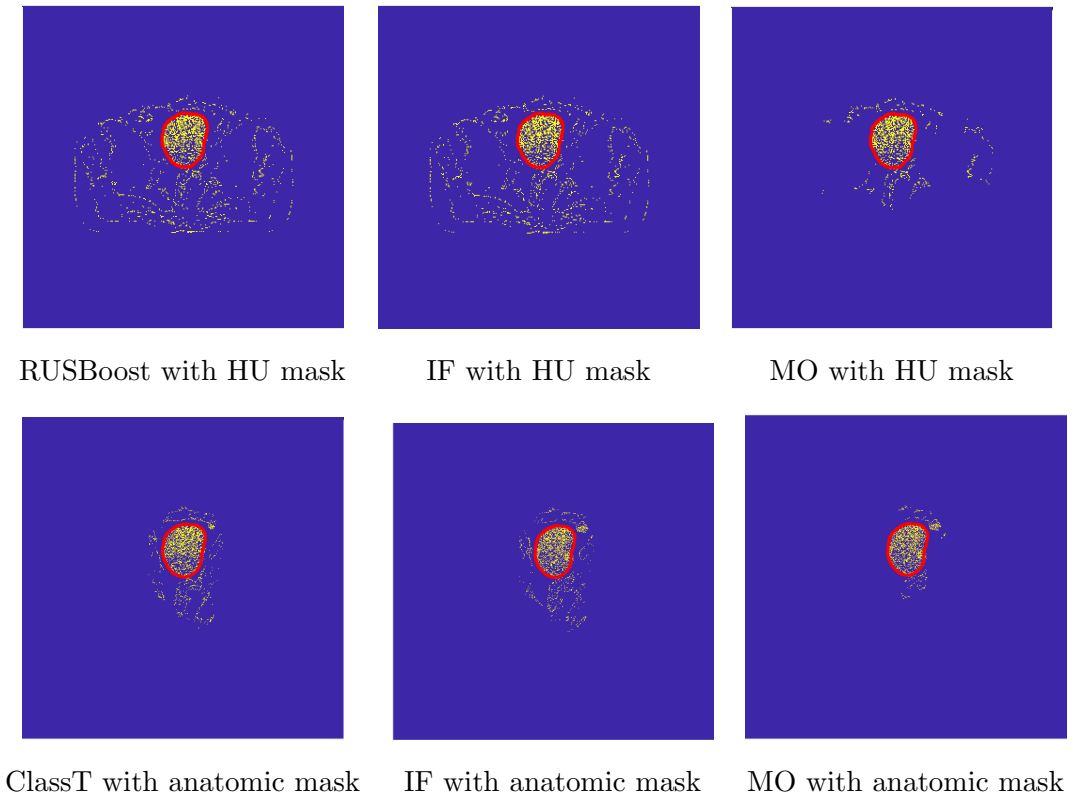


Figure 4.7: Clustering post-processing. "IF" stands for post-processing with flood fill and "MO" stands for post-processing with morphological operations. The red line is the GT

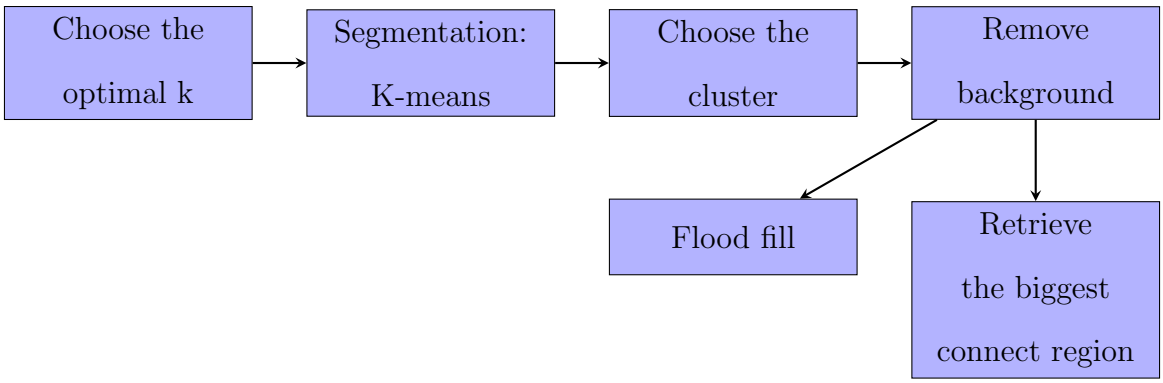


Figure 4.8: Clustering Process

4.2.4 U-Net

This algorithm was based and adapted from [58], a brain tumor segmentation tool for MRI. Training a network on the full input volume is impractical due to the amount of memory needed to store and process 3-D volumes. This problem is solved by training the network on image patches. On this step, several parameters were estimated:

- patches size;

- number of patches per image;
- batch size: size of the subset of train, which is used to evaluate the loss function gradient and update the weights;
- number of output channels for the first convolutional layer;
- number of epochs: number of times that the train set went through the train algorithm;
- validation frequency: number of network iterations;
- patience: number of epochs that the network waits to halve the learning rate, after learning rate does not improve.

An overlap-tile strategy is used to stitch test patches into a complete segmented test volume. The class imbalance in the data hampers training when using conventional cross entropy loss, which is fixed by using a weighted multi class Dice loss function [51], defined in Equation 4.1:

$$L_{dc} = -\frac{2}{|k|} \sum_{k \in K} \frac{\sum_i u_{i,k} v_{i,k}}{\sum_i u_{i,k} + \sum_i v_{i,k}} \tag{4.1}$$

where u is the prediction of the network, v is the true value, i are the train voxels, k are the classes and $u_{i,k}$ and $v_{i,k}$ are the output and GT for the class k in voxel i, respectively.

Weighting the classes helps to counter the influence of larger regions on the Dice score, making it easier for the network to learn how to segment smaller regions.

As a pre-procedure for train and validation data, each RoI was normalized by using its mean and standard deviation. For each image, random patches were extracted from ground truth images and corresponding pixel label data to feed the training and validation data to the network and to validate the training progress. To prevent overfitting due to data limited size, the training and validation data were augmented by randomly rotating and reflecting training data to make the training more robust. The response patches were cropped to the output size of the network.

To set up the 3D U-Net layers, a default 3D U-Net network was used. As represented in

Figure 4.9, the network is formed by an input layer, followed by 21 encoder layers. The first one is a convolutional layer that is responsible for halving the number of feature maps. The input is standardized to a layer for each mini-batch by the batch normalization layer. Then, the activation function, represented in Equation 4.2, is applied by the ReLU layer.

$$R(z) = max(0, z) \tag{4.2}$$

These three layers are repeated. The input halves down size by the max pooling layer. This process is repeated twice. The bridge layers are between the encoder and decoder layers and they are composed by a convolutional, a batch normalization and a ReLU layers. There are 23 decoder layers where, initially, the size of the images are expanded by a transposed convolutional layer. The images are concatenated by a concatenation layer, followed by a convolutional layer, a batch normalization layer and a ReLU layer. Similarly to the encoder layers, these three layers are repeated and this sequence happens twice more. The network ends with 3 more decoder layers: a final convolutional layer; a softmax layer, which applies a softmax activation function, defined in Equation 4.3, that converts the output of the last layer into a probability distribution, and finally the output layer. There are three crop 3D layers that connect the encoder and decoder layers, which is necessary due to the loss of border pixels in every convolution.

$$S(x_i) = \frac{e^{x_i}}{\sum_{j=1}^{n} e(x_j)} \tag{4.3}$$

In order to avoid border artifacts when using the overlap-tile strategy for prediction of the test volumes, valid convolution padding was specified. This model replaces the pixel classification layer with the Dice pixel classification layer, to better segment smaller regions and reduce the influence of larger background regions [58]. Then, the network was trained using the Adam

optimization solver, which is an optimization algorithm that can be used instead of the classical stochastic gradient descent procedure to update network weights iterative based on training

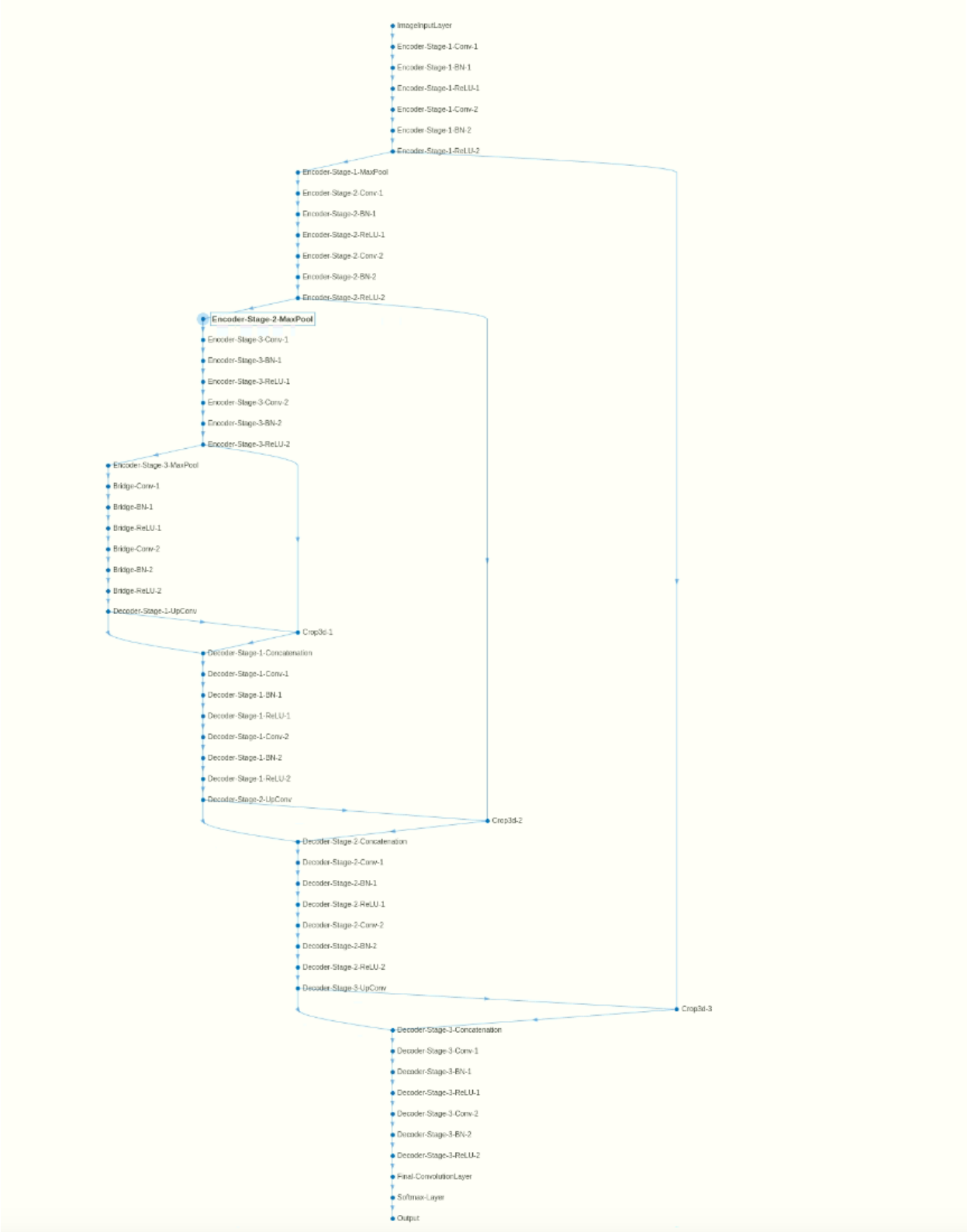


Figure 4.9: U-Net layers

data [59]. Using the test data, the overlap-tile strategy was applied to predict the labels for each test volume. Each test volume was padded to make the input size a multiple of the output size of the network and compensates for the effects of valid convolution [58]. The overlap-tile algorithm selects overlapping patches, predicts the labels for each patch and then recombines the patches.

A post-processing was implemented by removing the background and choosing the biggest connected region. This whole process is represented in Figure 4.10.

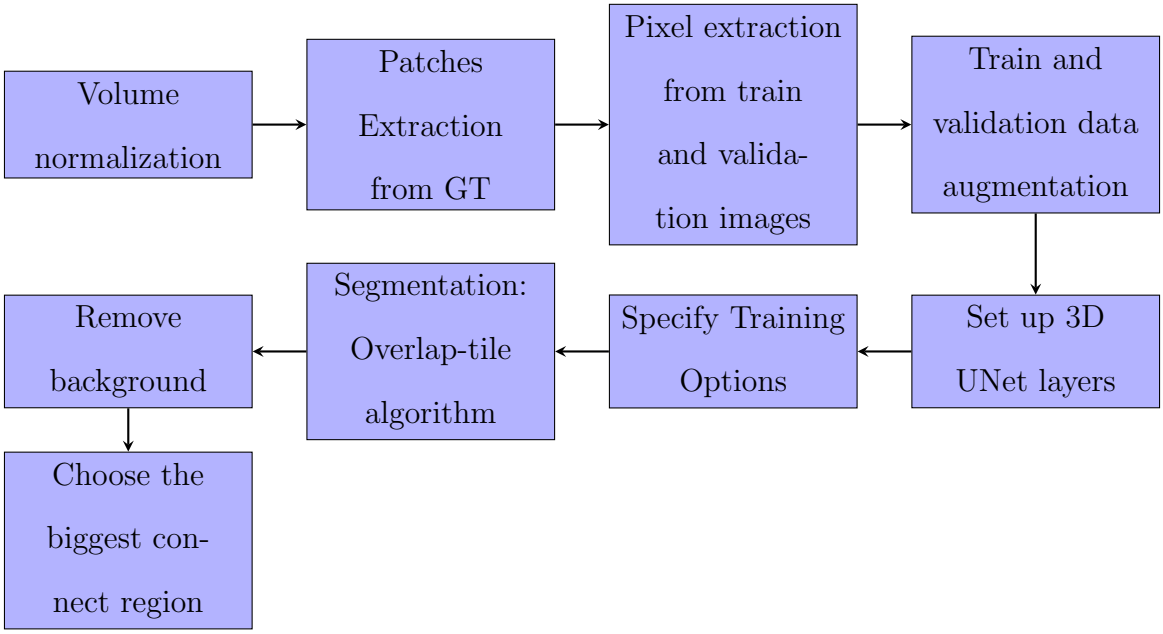


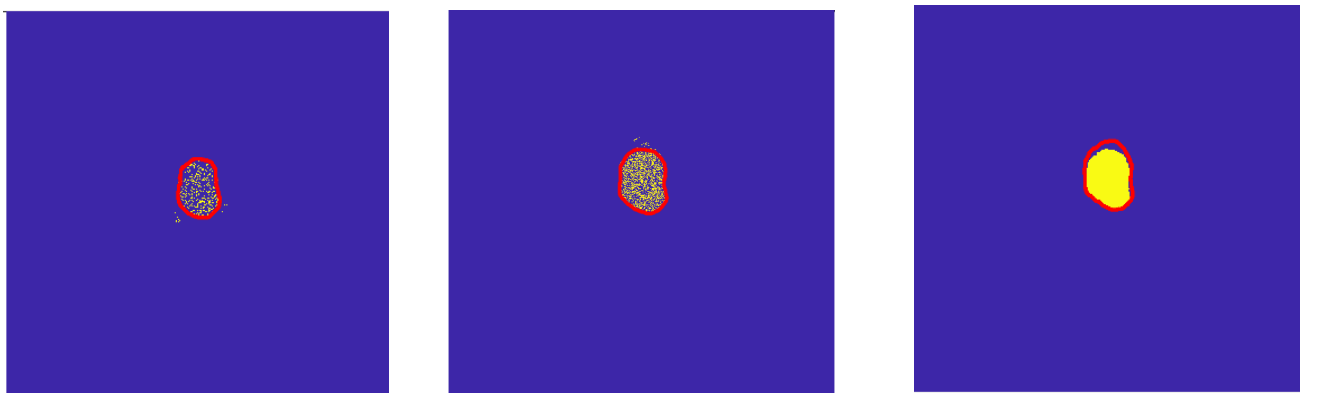
Figure 4.10: U-Net Process

4.2.5 Active Contours

Active Contours algorithm, also known as Snakes, consists of deforming the image domain and capture a desired feature through the constraint and image forces that pull it towards object contours and the internal forces resist deformation [60].

This algorithm was used to segment the image in foreground and background, by defining a mask with the initial contour at which the evolution of the segmentation begins. Two initial contours were used: the result of the segmentation by clustering after applied the post-processing and the result of the U-Net algorithm with anatomic mask. The initial contours are represented in Figure 4.11.

With Chan-Vese method choice, the edges were ignored completely. Instead, it optimally fitted



Cluster result with HU mask | Cluster result with anatomic mask | U-Net result with anatomic mask

Figure 4.11: Initial contour for Active Contours technique. The red line is the GT

a two phase piecewise constant model [61]. Moreover, with this method, the contour is free to either shrink or expand based on the image features.

The optimal number of iterations was found by evaluating some cases with the metrics DSC, Jaccard and BF score. It was estimated values between 150 and 325 with steps of 25. The algorithm stops the evolution of the Active Contour when it reaches the number of iterations.

The segmentation results by defining the U-Net results as initial contours went through the same morphological operations as the ones used to create the mask based on HU values. The results by defining the Clustering results as initial contours did not, since there were no improvements. The process is represented in Figure 4.12.

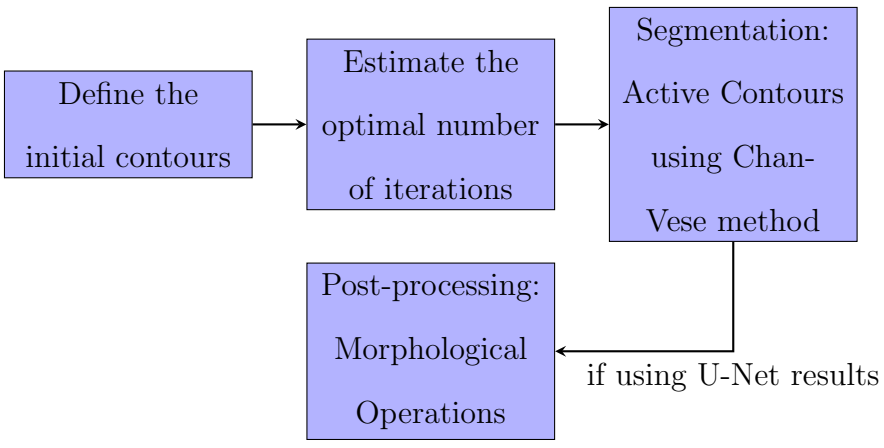


Figure 4.12: Active Contours Process

4.2.6 Graph Based

Graph-based image processing methods typically operate on pixel adjacency graphs. Adjacency graphs are graphs whose vertex set V is the set of image elements, and whose edge set E is given by an adjacency relation on the image elements.

Initially, the number of superpixels for the label matrix was found. Superpixels are the result of perceptual grouping of pixels. It carries more information than pixels. Values between 250 and 500 with steps of 25 were estimated by evaluating some cases with the metrics DSC, Jaccard and BF score. Those superpixels were used to define the label mask, that specifies the sub regions of the volumes. The foremask and the backmask were created in order to designate pixels in the image as foreground and background, respectively. If a region of the label matrix contains pixels belonging to both the foreground mask and background mask, the algorithm segments the region as background. As represented on the Figure 4.13, these masks were created with three approaches: defining them manually by choosing ranges based on ground truth position where the bladder is fully in (for the foremask) and fully out (for the backmask); defining the clustering post-processing results as foremask and defining the backmask by implementing a seed found based on ground truth coordinates; defining the U-Net results as foremask and using again the manual backmask.

The seed used to define the backmask was obtained by applying the formula 4.4 to each CT:

$$min(gt.s) + \frac{max(gt.s) - min(gt.s)}{2} \tag{4.4}$$

where $gt.s$ is the ground truth quota set. The seed is the average of the results.

The entire process is described in Figure 4.14.

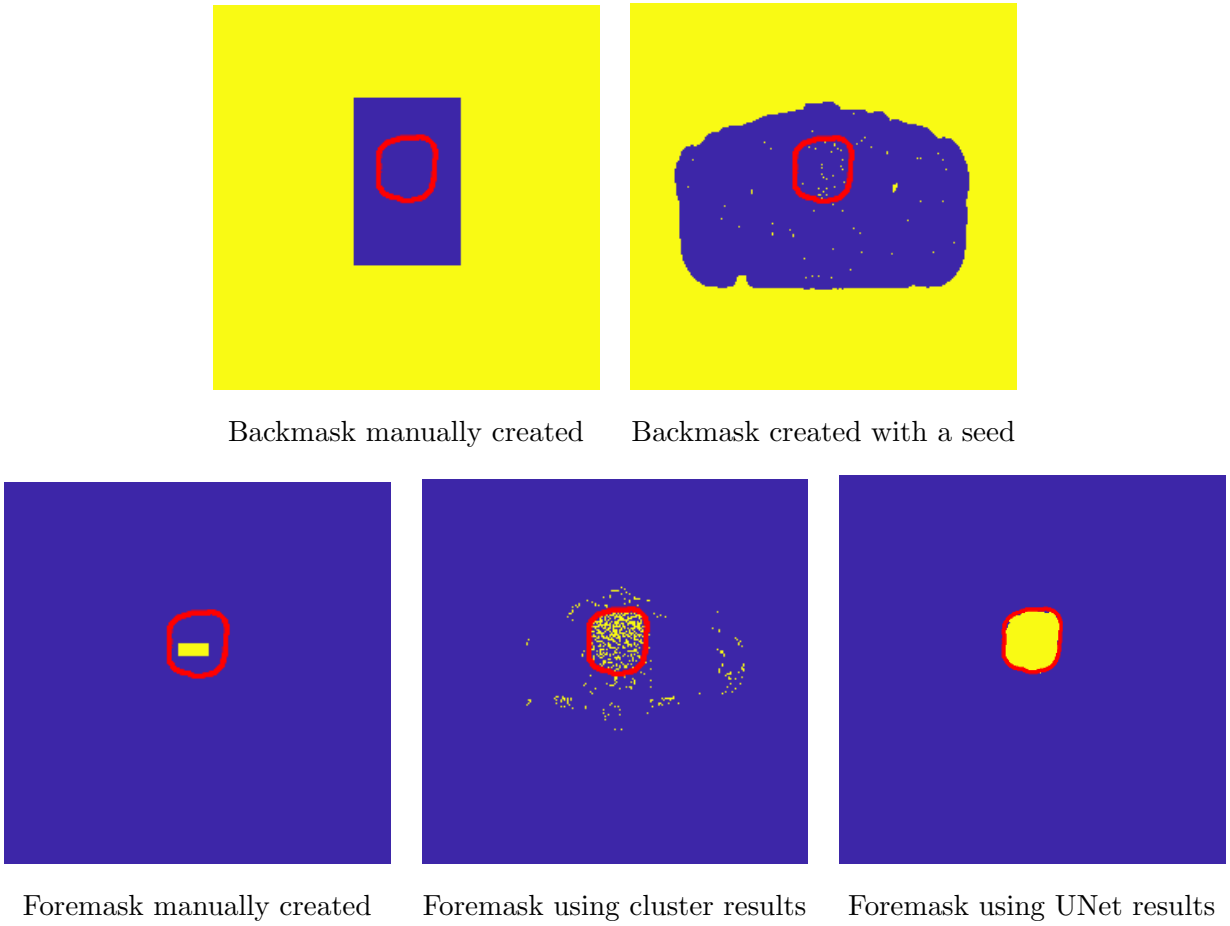


Figure 4.13: Creation of backmask and foremask. For the backmasks, the yellow region represents the pixels that are initialized as background and the blue ones the pixels that are not. For the foremasks, the yellow region represents the pixels that are initialized as foreground and the blue one the pixels that are not. The red line is the GT contour

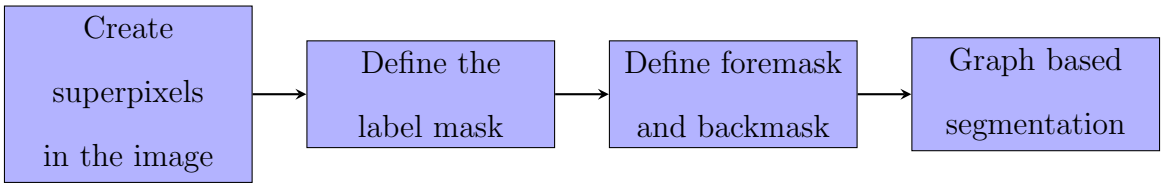


Figure 4.14: Graph Based Process

4.3 Evaluation methodology

To compare the algorithms performance, the evaluation methodology was the same to each one. Each bladder segmentation was evaluated in the train, validation and test set individually by the metrics DSC, Jaccard, BF score, Precision and Recall. For each set, the mean and the standard deviations were computed. Then, the average of the three sets was computed for each metric.

Chapter 5.

Results

The algorithms used were evaluated and the results can be analysed and compared in this chapter. It begins with the evaluation of the two pre-procedures used, in Section 5.1. Then, the methods used to segment were also evaluated: Clustering (Section 5.2), U-Net (Section 5.3), Active Contours (Section 5.4) and Graph Based (Section 5.5). At last, in Section 5.6, the results are compared and commented.

5.1 Detection of the Region of Interest

By thresholding the HU values, a RoI was achieved. It was expected more cases whose thresholds were centered in zero. Since it did not happen, the RoI was bigger than desired. In Figure 5.1, it is possible to visualize the ground truth, the ground truth applied to HU and the thresholded mask (initial mask) with the ground truth, in the first row. In the second one, the initial mask suffered morphological operation, originating the RoI (final mask). In these operations, the biggest connected region of the initial mask was found. At last, the RoI was applied to the HU values.

To evaluate the thresholding, five evaluation metrics were used: DSC, Jaccard index, BF score, Precision and Recall. The method was evaluated on the three sets of train, validation and test. The results can be analyzed in Table 5.1.

By applying the anatomic mask to each case, the RoI became smaller than the one obtained with the thresholding. The results obtained with the same metrics can be consulted in Table 5.2.

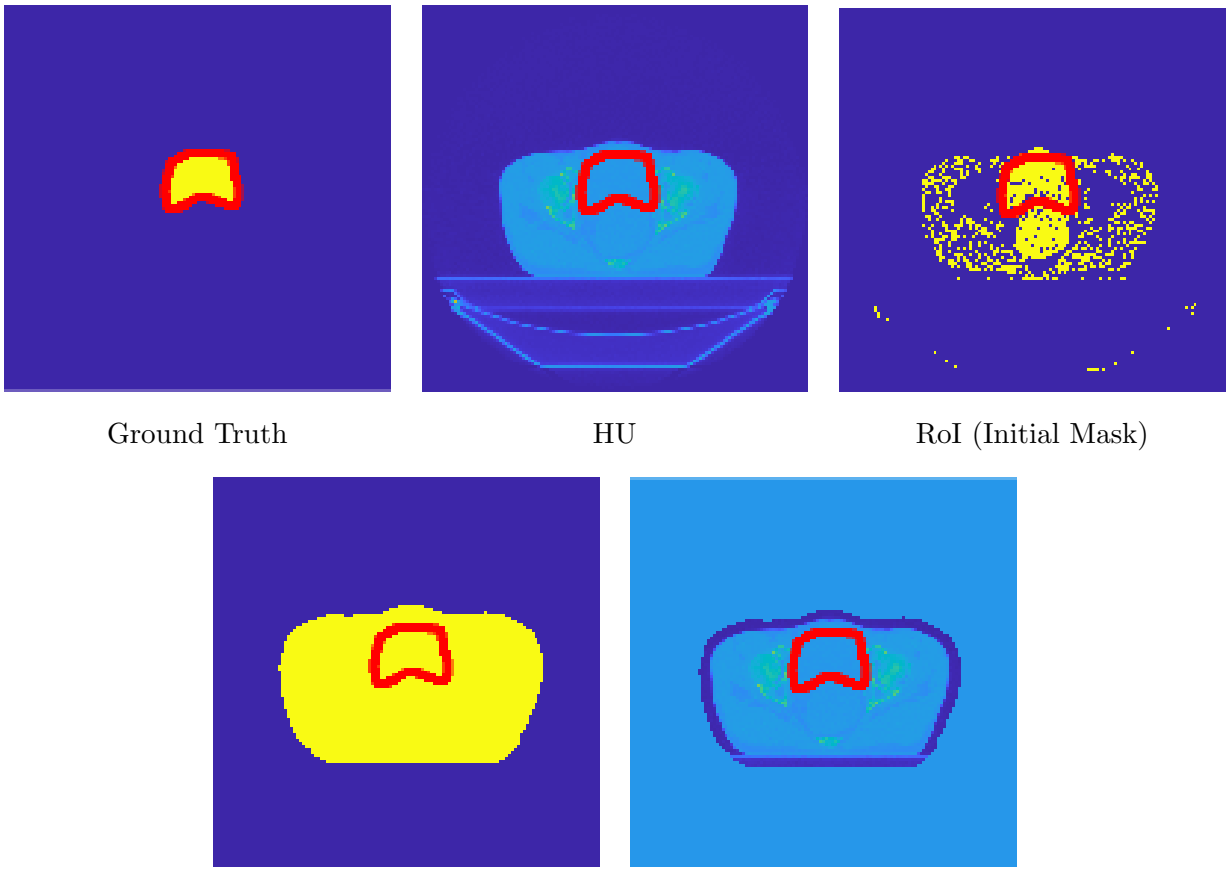


Figure 5.1: Slice with the RoI by applying the HU mask. The yellow regions are the pixels classified as bladder and the red lines are the GT contours.

Table 5.1: Results for the detection of the RoI by thresholding

Set	DSC	Jaccard	BF score	Precision	Recall
Train	0.0255 ± 0.0131	0.0130 ± 0.0067	0.8103 ± 0.0343	0.7991 ± 0.0473	0.8233 ± 0.0337
Validation	0.0153 ± 0.0025	0.0077 ± 0.0013	0.8009 ± 0.0233	0.7912 ± 0.0296	0.8113 ± 0.0259
Test	0.0270 ± 0.0114	0.0137 ± 0.0059	0.8043 ± 0.0252	0.7948 ± 0.0316	0.8147 ± 0.0297
Overall	0.0255 ± 0.0120	0.0129 ± 0.0062	0.8062 ± 0.0286	0.7961 ± 0.0377	0.8177 ± 0.0255

Table 5.2: Results for the detection of the RoI by using a mask based on anatomy

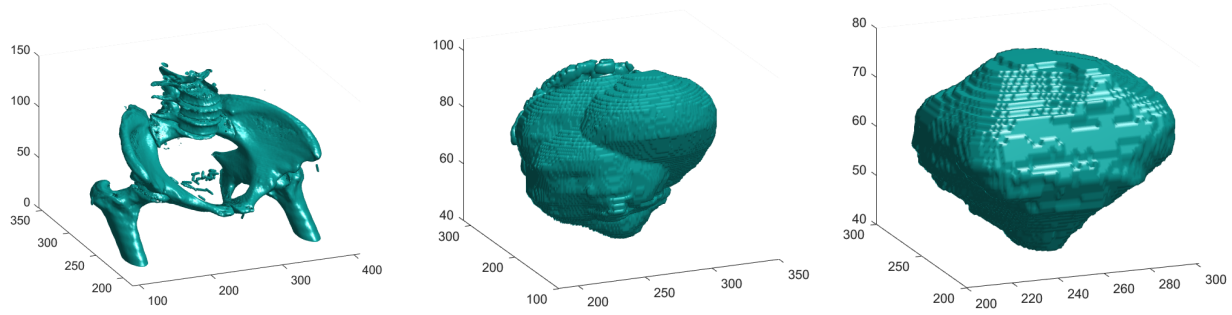
Set	DSC	Jaccard	BF score	Precision	Recall
Train	0.2973 ± 0.1191	0.1801 ± 0.0826	0.9619 ± 0.0027	0.9414 ± 0.0029	0.9832 ± 0.0047
Validation	0.2218 ± 0.0626	0.1260 ± 0.0412	0.9647 ± 0.0047	0.9439 ± 0.0051	0.9866 ± 0.0051
Test	0.3300 ± 0.1277	0.2045 ± 0.0921	0.9635 ± 0.0039	0.9437 ± 0.0051	0.9842 ± 0.0044
Overall	0.3087 ± 0.1230	0.1888 ± 0.0875	0.9630 ± 0.0036	0.9429 ± 0.0045	0.9840 ± 0.3087

In Figure 5.2, the ground truth was drawn once again to be compared with the anatomic mask and with the anatomic mask applied to the HU values.

Ground Truth | Anatomic mask | Anatomic mask applied to HU

Figure 5.2: Slice with the RoI by applying the anatomic mask.

The two approaches can be compared in Figure 5.3, where the isosurfaces of the volumes can be analysed.



Mask based on HU values | Mask based on anatomy | GT

Figure 5.3: RoI detection

5.2 Clustering

To choose the optimal K for K-means, K values between 5 and 195 were tested by steps of 10, as mentioned in Chapter 4, Section 4.2.3. Analysing the histogram of the Figure 5.4, the chosen values using the train set vary between 35 and 185 with the mode 115 and mean 110. To choose between these two values, the segmentation was evaluated using the metrics DSC, Jaccard, BF score and Precision in the validation set. The results were quite similar for the two K values, as one can see in the Figure 5.4.

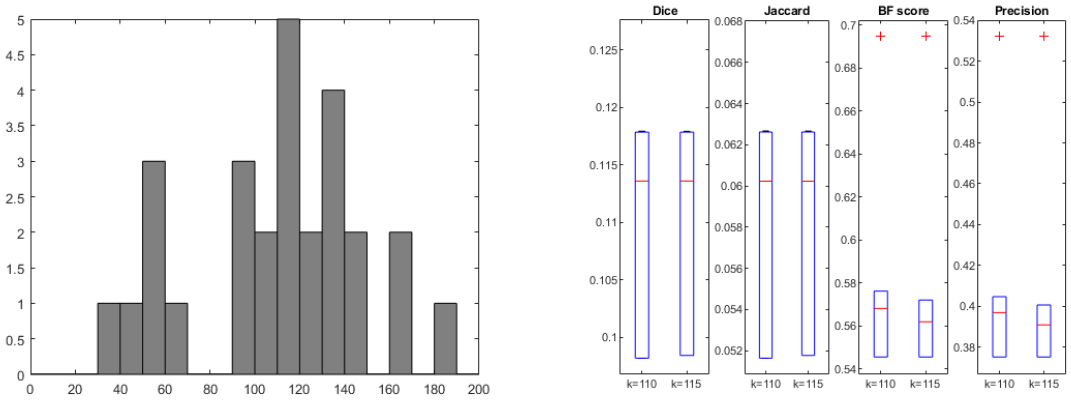


Figure 5.4: Selection of K. Left: histogram of the chosen K values on the train set. Right: Metrics for $K = 110$ and $K = 115$ on the validation set.

Since the results are very alike, the two values of K will be used. The next step is to automatically choose one cluster from the 110 and 115 clusters. After the train data set being trained by the classifiers, its performance was evaluated in the validation data set with and without data normalization. Analysing these results in Table 5.3, the best one is 115 clusters with data normalization with the mean and the standard deviation.

Table 5.3: Clustering choice results with the HU mask (accuracy). "K" stands for the number of clusters used for segmentation.

K	normalization	ClassT	DA	kNN	NB	SVM	CE	CTE	RUSBoost
110	no	0.9909 ± 0.0950	0.9818 ± 0.0950	0.9712 ± 0.1673	0.9818 ± 0.1337	0.0182 ± 0.1337	0.9939 ± 0.0777	0.9879 ± 0.1095	0.0091 ± 0.0950
115	no	0.9841 ± 0.1253	0.9826 ± 0.1253	0.9855 ± 0.1196	0.9826 ± 0.1308	0.1493 ± 0.3566	0.9913 ± 0.0929	0.9913 ± 0.0929	0.0087 ± 0.0929
110	yes	0.9909 ± 0.0950	0.9818 ± 0.0950	0.9848 ± 0.1222	0.9818 ± 0.1337	0.9909 ± 0.0950	0.9939 ± 0.0777	0.9939 ± 0.0777	0.0091 ± 0.0950
115	yes	0.9841 ± 0.1253	0.9826 ± 0.1253	0.9870 ± 0.1135	0.9826 ± 0.1308	0.9913 ± 0.0929	0.9913 ± 0.0929	0.9971 ± 0.0538	0.9870 ± 0.1135

Since the data set is highly imbalanced, it is not enough to study only the accuracy. For that reason, the confusion matrices were computed by training in the train set and evaluating in the validation set. The results can be consulted in Table 5.4.

Since the goal is to correctly localise the bladder, it is assigned more importance to TP. The RUSBoost seems to be the best classifier, since it detects all existing TPs. A comparison of Clustering when using the ground truth and the RUSBoost for region selection is given in Table 5.5.

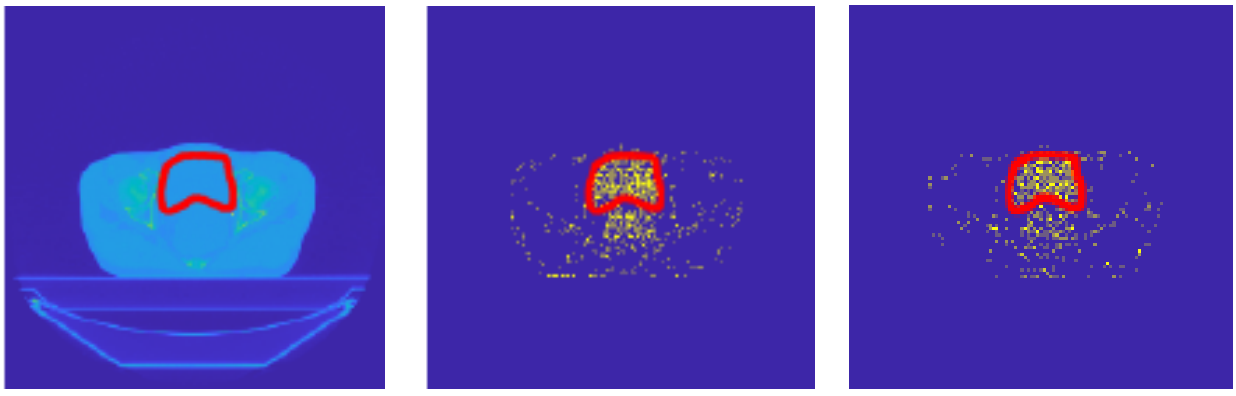
In Figure 5.5, the result of K-means application in one validation set patient with the K value 110 and 115 is represented, respectively.

Table 5.4: Confusion Matrices of each classifier computed by using the data set with HU mask

Model	TP	FN	FP	TN
ClassificationTrees	2	4	7	677
DiscriminantAnalysis	0	6	6	678
kNearestNeighbors	3	3	6	678
NaiveBayes	0	6	6	678
SupportVectorMachines	0	6	0	684
ClassificationEnsembles	4	2	4	680
ClassificationTreeEnsembles	5	1	5	679
RUSBoost	6	0	6	678

Table 5.5: Clustering results summary table. "RS=GT" stands for region selected using the Ground Truth information and "RS=RUSBoost" stands for region selected using the RUSBoost classifier.

Method	Set	DSC	Jaccard	BF score	Precision	Recall
	Train	0.1120 ± 0.1055	0.0626 ± 0.0613	0.4827 ± 0.2158	0.3659 ± 0.1517	0.7424 ± 0.3723
k=115;	Validation	0.1100 ± 0.0095	0.0582 ± 0.0053	0.5804 ± 0.0573	0.4108 ± 0.0606	1.0000 ± 0.0000
RS=GT	Test	0.1303 ± 0.1008	0.0727 ± 0.0582	0.5130 ± 0.1915	0.3844 ± 0.1315	0.8028 ± 0.3402
	Overall	0.1218 ± 0.0980	0.0677 ± 0.0568	0.5073 ± 0.1939	0.3796 ± 0.1346	0.7966 ± 0.1218
	Train	0.1120 ± 0.1055	0.0626 ± 0.0613	0.4827 ± 0.2158	0.3659 ± 0.1517	0.7424 ± 0.3723
k=115;	Validation	0.1100 ± 0.0095	0.0582 ± 0.0053	0.5804 ± 0.0573	0.4108 ± 0.0606	1.0000 ± 0.0000
RS=RUSBoost	Test	0.0981 ± 0.0974	0.0544 ± 0.0561	0.4868 ± 0.2027	0.3711 ± 0.1353	0.7442 ± 0.3672
	Overall	0.1043 ± 0.0959	0.0578 ± 0.0555	0.4930 ± 0.1999	0.3725 ± 0.1365	0.7649 ± 0.1043



Ground Truth | Segmentation with $K = 110$ | Segmentation with $K = 115$

Figure 5.5: Clustering results for k=110 and k=115

At last, the two different post-processing were applied. Analysing and comparing the results on Table 5.6, the best results are achieved with the choice of the biggest connected region.

Table 5.6: Clustering results after post-processing with the HU mask summary table. "PP=IF" stands for post-processing with flood fill and "PP=MO" stands for post-processing with morphological operations.

Method	Set	DSC	Jaccard	BF score	Precision	Recall
	Train	0.1635 ± 0.0827	0.0912 ± 0.0497	0.6188 ± 0.0837	0.4530 ± 0.0868	1.0000 ± 0.0000
k=115;	Validation	0.1117 ± 0.0111	0.0592 ± 0.0062	0.5804 ± 0.0573	0.4109 ± 0.0606	1.0000 ± 0.0000
PP=IF	Test	0.1591 ± 0.0855	0.0887 ± 0.0506	0.6290 ± 0.0720	0.4628 ± 0.0785	1.0000 ± 0.0000
	Overall	0.1568 ± 0.0813	0.0872 ± 0.0485	0.6211 ± 0.0758	0.4548 ± 0.0807	1.0000 ± 0.1568
	Train	0.2589 ± 0.1457	0.1563 ± 0.0943	0.8156 ± 0.1541	0.7152 ± 0.2083	0.9989 ± 0.0042
k=115;	Validation	0.1402 ± 0.0905	0.0775 ± 0.0529	0.7378 ± 0.1374	0.6036 ± 0.1921	0.9975 ± 0.0060
PP=MO	Test	0.2356 ± 0.1551	0.1422 ± 0.1020	0.8399 ± 0.1496	0.7540 ± 0.2196	0.9969 ± 0.0066
	Overall	0.2364 ± 0.1490	0.1421 ± 0.0972	0.8223 ± 0.1510	0.7269 ± 0.2146	0.9977 ± 0.2364

Some cases and the respective isosurfaces of the segmentation can be observed in Figure 5.6.

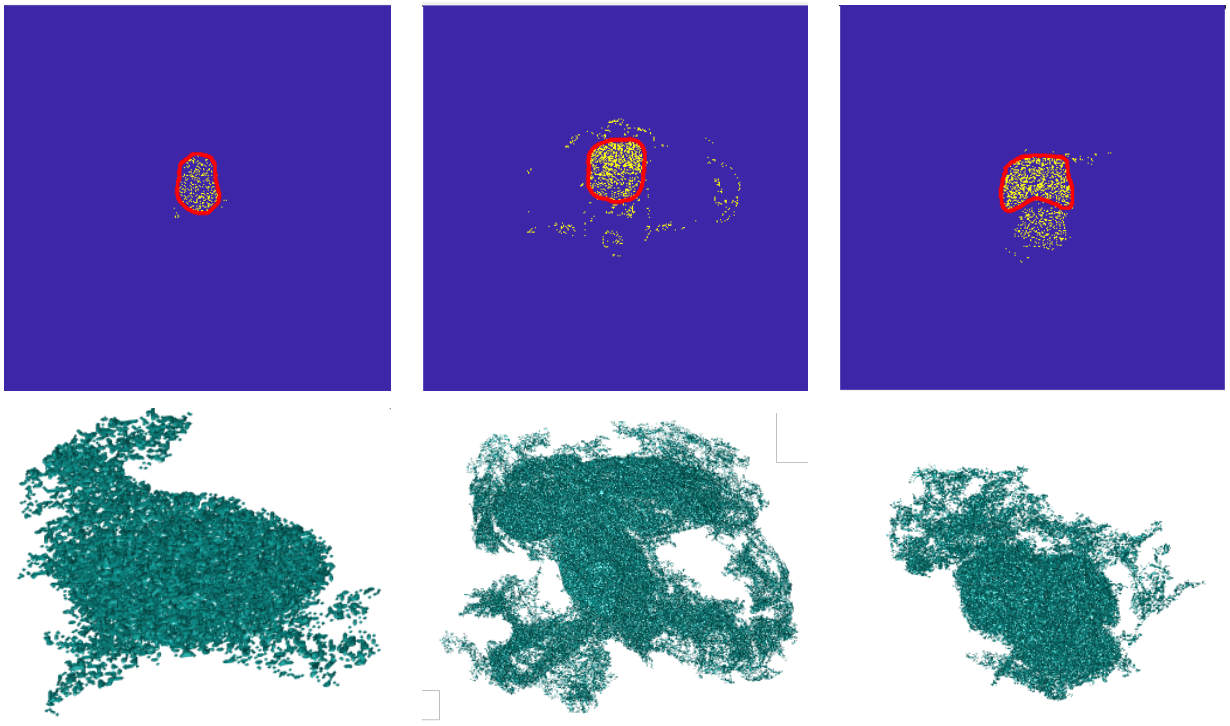


Figure 5.6: Selected example of clustering with morphological operations with the HU mask

The data set with the anatomic mask went through the same process. It started by choosing the optimal K by testing values between 5 and 195 by steps of 10. In this case, the mode was 25 and the mean 46. The segmentation with these two K values was evaluated by the previous process, as the Table 5.7 shows.

Table 5.7: Clustering choice results with the anatomic mask (accuracy). "K" stands for the number of clusters used for segmentation.

K	normalization	ClassT	DA	kNN	NB	SVM	CE	CTE	RUSBoost
46	no	0.9819 ± 0.1336	0.9601 ± 0.1336	0.9601 ± 0.1960	0.9565 ± 0.2043	0.6667 ± 0.4723	0.9710 ± 0.1681	0.9819 ± 0.1336	0.0217 ± 0.1461
25	no	0.9800 ± 0.1405	0.9533 ± 0.2116	0.9467 ± 0.2255	0.8133 ± 0.3909	0.9600 ± 0.1966	0.9467 ± 0.2255	1.0000 ± 0.0000	0.0400 ± 0.1966
46	yes	0.9819 ± 0.1336	0.9601 ± 0.1336	0.9746 ± 0.1575	0.9565 ± 0.2043	0.9746 ± 0.1575	0.9710 ± 0.1681	0.9746 ± 0.1575	0.0217 ± 0.1461
25	yes	0.9800 ± 0.1405	0.9467 ± 0.1405	0.9533 ± 0.2116	0.9467 ± 0.2255	0.9533 ± 0.2116	0.9600 ± 0.1966	0.9667 ± 0.1801	0.9667 ± 0.1801

The results are quite similar. However, the chosen one was 46 clusters with data normalization, since it had the best results with exception of RUSBoost.

Once again, the confusion matrices were computed by training and evaluating in the train and validation data set, respectively. Analysing the results in Table 5.8, Classification Trees classifier was the one that achieved better results.

Table 5.8: Confusion Matrices of each classifier

Model	TP	FN	FP	TN
ClassificationTrees	4	2	1	143
DiscriminantAnalysis	2	4	4	140
kNearestNeighbors	2	4	3	141
NaiveBayes	2	4	4	140
SupportVectorMachines	0	6	1	143
ClassificationEnsembles	1	5	1	143
ClassificationTreeEnsembles	2	4	1	143
RUSBoost	2	4	1	143

Clustering was performed by using the ground truth information and Classification Trees for region selection. The results are noted in Table 5.9.

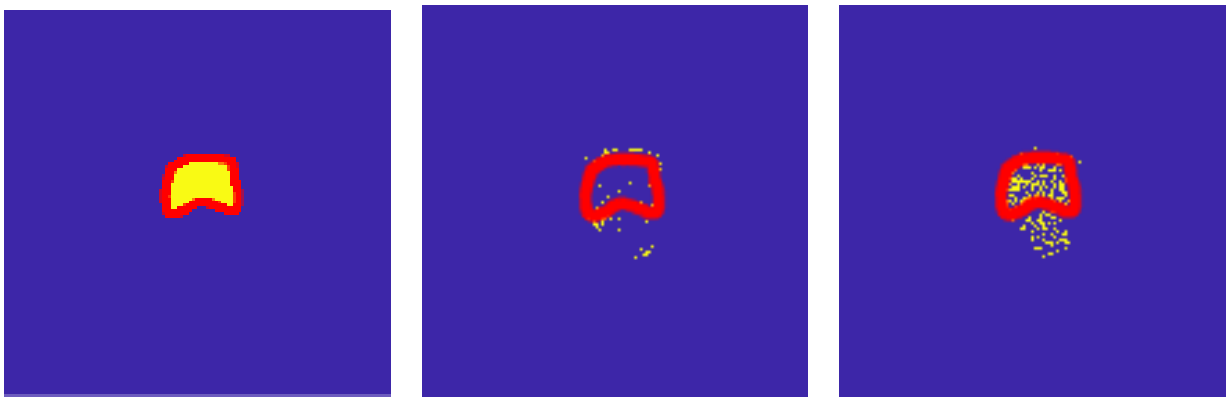
In Figure 5.7, the result of K-means application in one validation set patient with the K value 25 and 46, respectively, can be compared.

The same two post-processing were applied. Comparing the results in Table 5.10, the best results are, once again, achieved with the choice of the biggest connected region.

Some cases and the respective isosurfaces of the result can be analysed in Figure 5.8.

Table 5.9: Clustering applied to data set with anatomic mask results summary table. "RS=GT" stands for region selected using the Ground Truth information and "RS=ClassT" stands for region selected using Classification Tree classifier.

Method	Set	DSC	Jaccard	BF score	Precision	Recall
	Train	0.3403 ± 0.1090	0.2099 ± 0.0770	0.9781 ± 0.0145	0.9476 ± 0.0278	1.0000 ± 0.0000
k=46;	Validation	0.3373 ± 0.1173	0.2080 ± 0.0876	0.9905 ± 0.0034	0.9812 ± 0.0066	1.0000 ± 0.0000
RS=GT	Test	0.3789 ± 0.0948	0.2378 ± 0.0710	0.9741 ± 0.0162	0.9500 ± 0.0308	1.0000 ± 0.0000
	Overall	0.3611 ± 0.1024	0.2249 ± 0.0748	0.9771 ± 0.0155	0.9556 ± 0.0295	1.0000 ± 0.0000
	Train	0.2469 ± 0.1821	0.1522 ± 0.1141	0.6385 ± 0.4304	0.7011 ± 0.2595	0.6742 ± 0.4706
k=46;	Validation	0.3511 ± 0.0403	0.2135 ± 0.0293	0.9345 ± 0.0039	0.8772 ± 0.0068	1.0000 ± 0.0000
RS=ClassT	Test	0.1541 ± 0.1668	0.0925 ± 0.1025	0.5355 ± 0.4511	0.6184 ± 0.2952	0.5646 ± 0.4936
	Overall	0.2058 ± 0.1758	0.1252 ± 0.1092	0.6092 ± 0.4338	0.6720 ± 0.2767	0.6441 ± 0.2058



Ground Truth | Segmentation with $K = 25$ | Segmentation with $K = 46$

Figure 5.7: Clustering results for k=25 and k=46

Table 5.10: Clustering results after post-processing with the anatomic mask summary table. "PP=IF" stands for post-processing with flood fill and "PP=MO" stands for post-processing with morphological operations.

Method	Set	DSC	Jaccard	BF score	Precision	Recall
	Train	0.2147 ± 0.1798	0.1310 ± 0.1119	0.5628 ± 0.4614	0.9302 ± 0.0600	0.5957 ± 0.4976
k=46;	Validation	0.3443 ± 0.0403	0.2085 ± 0.0292	0.9333 ± 0.0026	0.8750 ± 0.0045	1.0000 ± 0.0000
PP=IF	Test	0.1433 ± 0.1640	0.0859 ± 0.1014	0.4944 ± 0.4604	0.9360 ± 0.0639	0.5220 ± 0.5003
	Overall	0.1863 ± 0.1722	0.1127 ± 0.1068	0.5561 ± 0.4527	0.9288 ± 0.0614	0.5889 ± 0.1863
	Train	0.2684 ± 0.2069	0.1702 ± 0.1324	0.6340 ± 0.4578	0.7196 ± 0.3120	0.6469 ± 0.4805
k=46;	Validation	0.3910 ± 0.0397	0.2437 ± 0.0306	0.9664 ± 0.0072	0.9351 ± 0.0134	1.0000 ± 0.0000
PP=MO	Test	0.1623 ± 0.1812	0.0993 ± 0.1142	0.5631 ± 0.4683	0.6531 ± 0.3391	0.5746 ± 0.4925
	Overall	0.2215 ± 0.1960	0.1381 ± 0.1248	0.6247 ± 0.4533	0.7024 ± 0.3209	0.6387 ± 0.2215

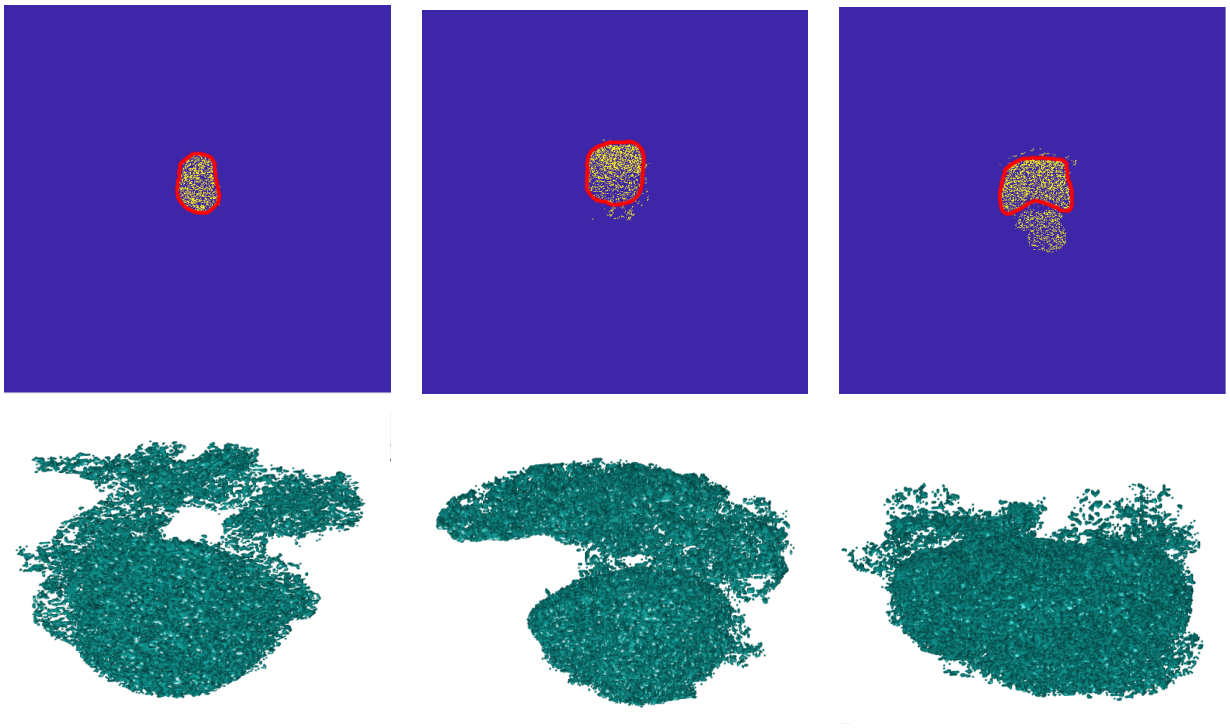


Figure 5.8: Selected examples of clustering with morphological operations using anatomic mask

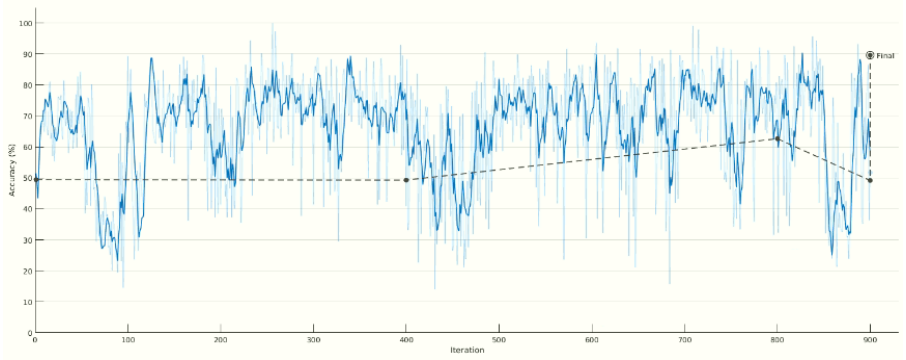
5.3 U-Net

The U-Net algorithm was applied to the data set with the mask based on HU values by training the network on train data set. The parameters estimated were the ones represented in Table 5.11. The bold parameters are the optimal ones. In Figure 5.9 it is possible to analyse the accuracy and loss during training phase. It ended with 89.63% validation accuracy. The train took 22 min 44 sec by using a GPU. The GPU used was a Nvidia with Core GPU temperature $28C$, GPU performance $P8$, power capability $260W$ and memory $24220MiB$.

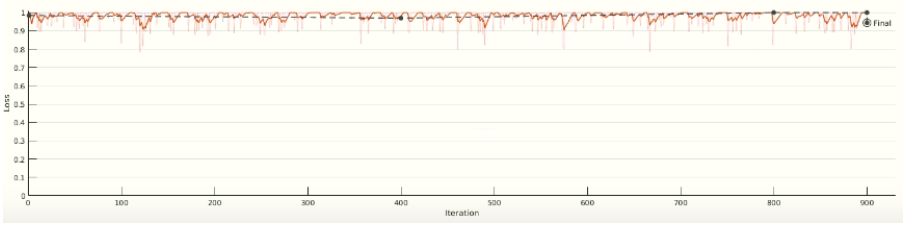
With the Overlap-tile algorithm, segmentation was performed on test data. After removing the background and choosing the biggest connected region, the data was evaluated and the results are shown in Table 5.12.

Table 5.11: Parameters estimated in U-Net algorithm (HU mask)

Parameter	Value
Patch Size	**100^3**
	92^3
Patch per Image	**6**
	16
	26
Batch Size	6
	8
	10
Output Channels	**15**
	20
	32
Epochs	25
	50
	75
Validation Frequency	200
	400
	600
Patience	**50**
	40



Training Accuracy



Training Loss

Figure 5.9: Accuracy and loss on the network training the data set with the HU mask. The blue line represents train accuracy, the red line represents train loss and the black lines are validation accuracy (first image) and validation loss (second image)

In Figure 5.10, some cases and the respective isosurfaces can be analysed, respectively.

The same process was implemented for the mask based on anatomy. The parameters estimated are represented in Table 5.13. Those that are highlighted are the ones with which the best results were obtained. The accuracy and loss during train can be analysed in Figure 5.11, respectively. The train ended with 10.52% validation accuracy and took 64 min 30 sec by using the same GPU.

Once again, segmentation was performed on test data with the Overlap-tile algorithm and the post-processing was applied. The results can be consulted in Table 5.14.

In Figure 5.12, some cases and the respective isosurfaces can be observed, respectively.

Table 5.12: U-Net results using HU mask

Set	DSC	Jaccard	BF score	Precision	Recall
Train	0.0571 ± 0.0288	0.0296 ± 0.0153	0.6928 ± 0.0618	0.5446 ± 0.0751	0.9624 ± 0.0069
Validation	0.0354 ± 0.0044	0.0180 ± 0.0023	0.6697 ± 0.0513	0.5161 ± 0.0573	0.9596 ± 0.0045
Test	0.0690 ± 0.0338	0.0360 ± 0.0182	0.7154 ± 0.0542	0.5699 ± 0.0645	0.9672 ± 0.0101
Overall	0.0617 ± 0.0318	0.0321 ± 0.0171	0.7030 ± 0.0580	0.5559 ± 0.0693	0.9648 ± 0.0617

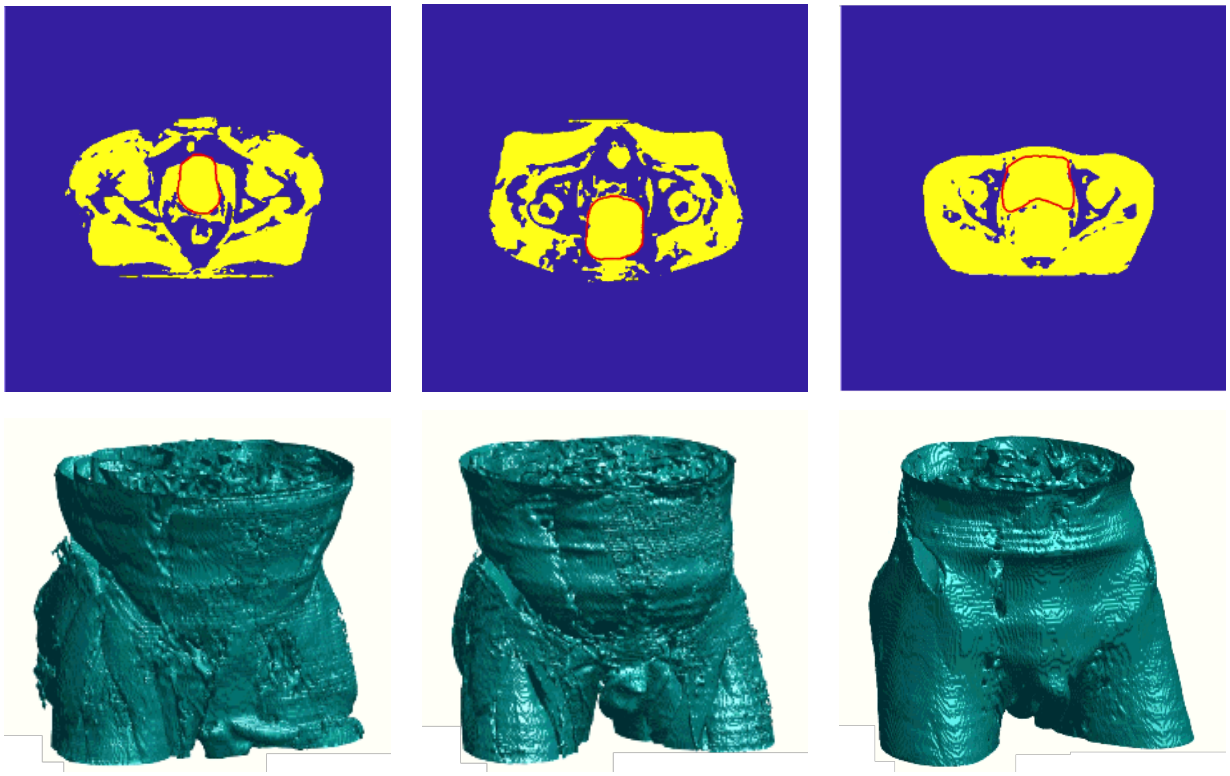
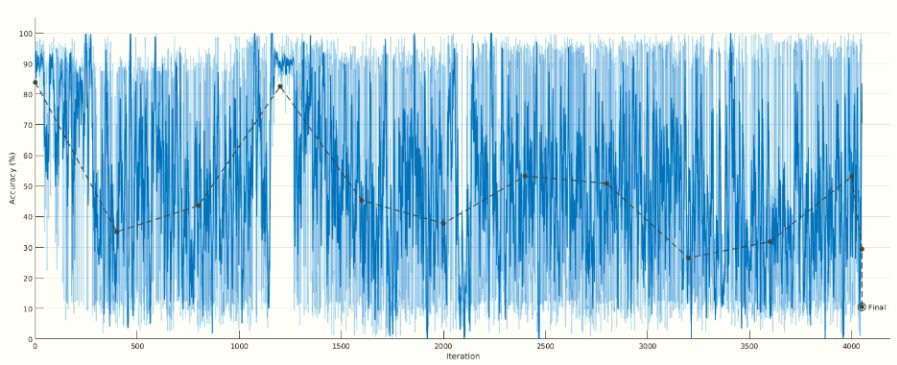
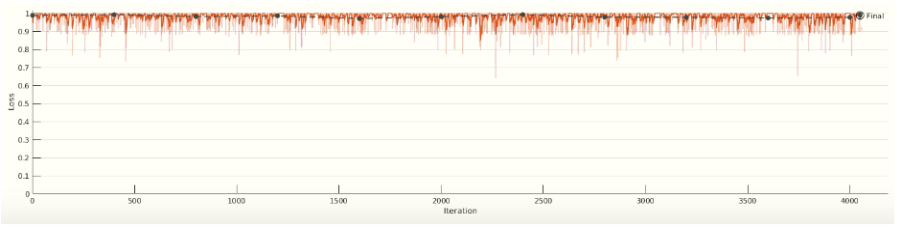


Figure 5.10: Selected results of U-Net using the HU mask



Training Accuracy



Training Loss

Figure 5.11: Accuracy and loss on the network training the data set with the anatomic mask. The blue line represents train accuracy, the red line represents train loss and the black lines are validation accuracy (first image) and validation loss (second image)

Table 5.13: Parameters estimated in U-Net algorithm (anatomic mask)

Parameter	Value
Patch Size	**100^3**
	92^3
Patch per Image	16
	26
	36
Batch Size	6
	8
	10
Output Channels	**20**
	32
	50
Epochs	25
	50
	75
Validation Frequency	200
	400
	600
Patience	**50**
	40

Table 5.14: U-Net results using the anatomic mask

Set	DSC	Jaccard	BF score	Precision	Recall
Train	0.7929 ± 0.2013	0.6958 ± 0.2448	0.9980 ± 0.0033	0.9976 ± 0.0054	0.9984 ± 0.0032
Validation	0.7249 ± 0.1669	0.5909 ± 0.2053	0.9966 ± 0.0030	0.9956 ± 0.0039	0.9977 ± 0.0030
Test	0.8252 ± 0.2478	0.7538 ± 0.2537	0.9984 ± 0.0038	0.9984 ± 0.0046	0.9984 ± 0.0045
Overall	0.8048 ± 0.2249	0.7186 ± 0.2482	0.9981 ± 0.0036	0.9979 ± 0.0049	0.9984 ± 0.8048

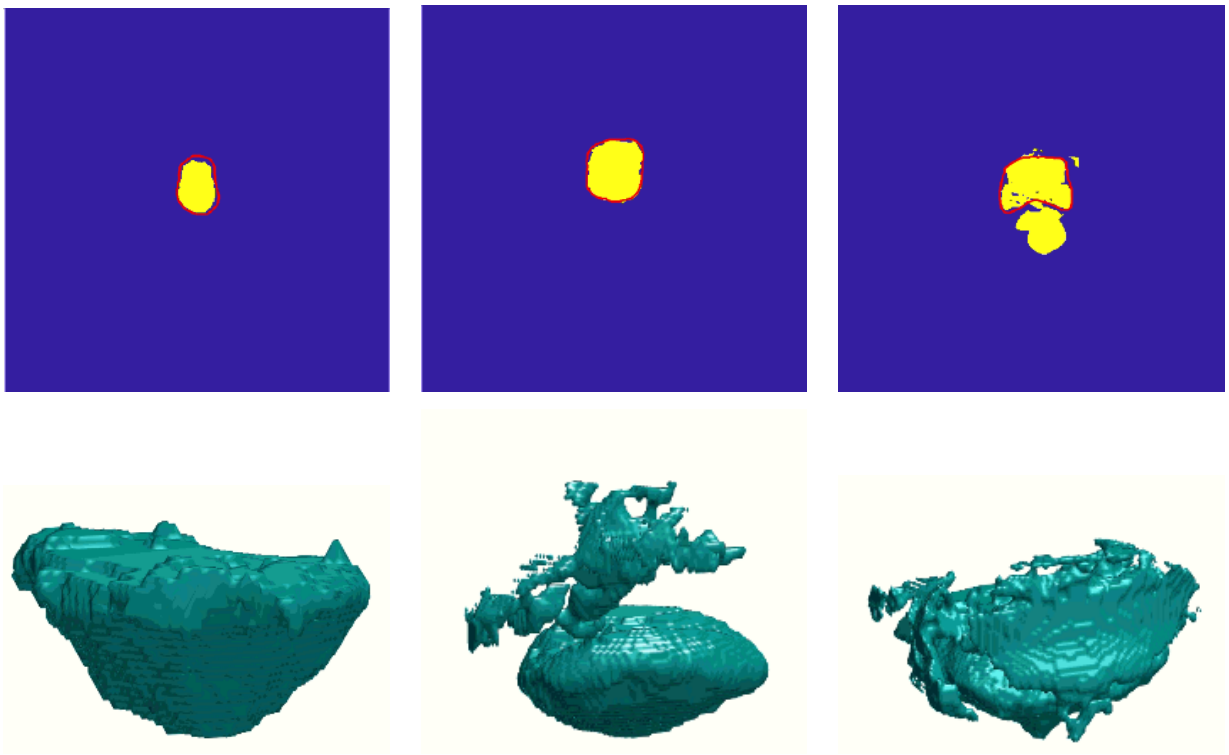


Figure 5.12: Selected results of U-Net using the anatomic mask

5.4 Active Contours

To implement the Active Contours algorithm, the first step is to define the initial contours from which the evolution starts. For this purpose, two masks were used: the Clustering results and the U-Net results with anatomic mask. As mentioned in the Subsection 4.2.5, the number of iterations was estimated using the data set with the mask based on HU and the optimal value was 225. Using Chan Vese method, the active contours algorithm were applied to each image.

The results of Active Contours by using the clustering results as initial contour are represented in Table 5.15.

Table 5.15: Active Contours results using the Clustering results with the HU mask

Set	DSC	Jaccard	BF score	Precision	Recall
Train	0.0031 ± 0.0048	0.0016 ± 0.0024	0.6514 ± 0.4696	0.7058 ± 0.4001	0.6475 ± 0.4745
Validation	0.0000 ± 0.0000	0.0000 ± 0.0000	0.9887 ± 0.0039	0.9909 ± 0.0055	0.9865 ± 0.0030
Test	0.0589 ± 0.1459	0.0374 ± 0.0962	0.6013 ± 0.4711	0.6795 ± 0.3734	0.5988 ± 0.4815
Overall	0.0334 ± 0.1109	0.0210 ± 0.0730	0.6556 ± 0.4572	0.7182 ± 0.3730	0.6526 ± 0.0334

Some segmentation's results and the respective isosurfaces were drawn in Figure 5.13.

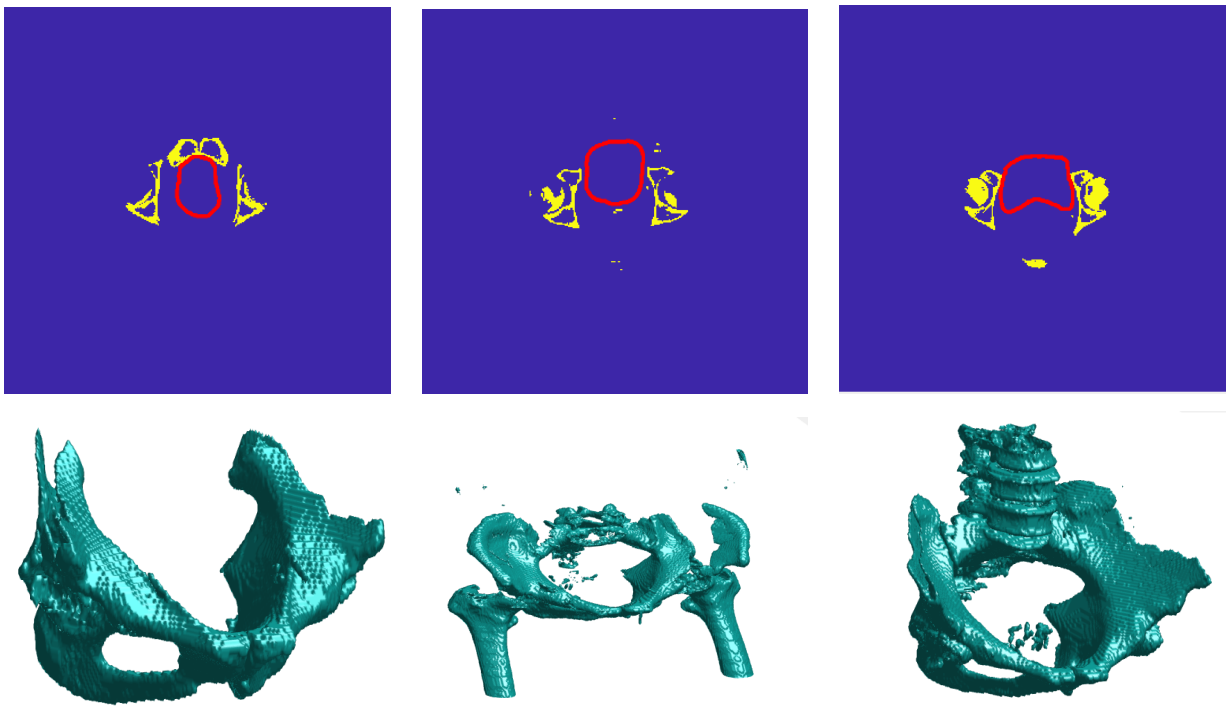


Figure 5.13: Selected examples of Active Contours results using Clustering results with the HU mask

Then, the initial mask was defined by U-Net results. Some results can be observed in Figure 5.14.

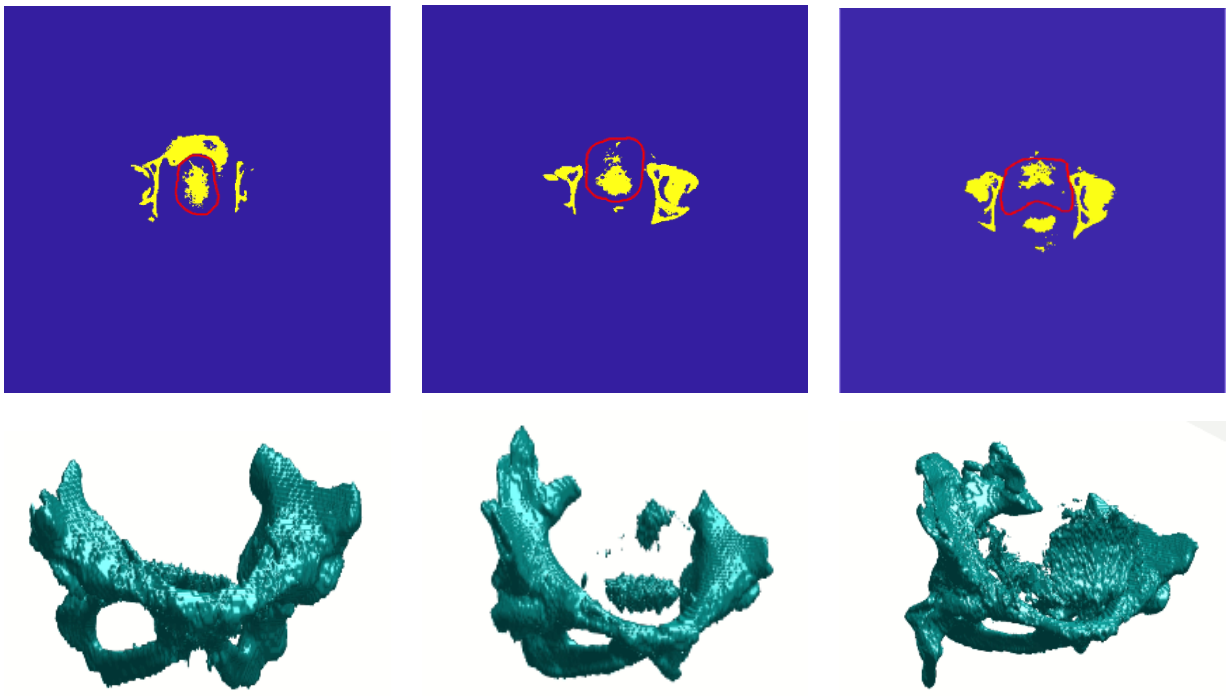


Figure 5.14: Selected examples of Active Contours using U-Net results with the HU mask

The morphological operations used to create the HU mask were implemented to the results obtained by defining the U-Net results as initial contour. These results can be analysed in Table 5.16.

Table 5.16: Active Contours using UNet results with the HU mask

Set	DSC	Jaccard	BF score	Precision	Recall
Train	0.2255 ± 0.2429	0.1486 ± 0.1641	0.9678 ± 0.0102	0.9442 ± 0.0189	0.9927 ± 0.0067
Validation	0.0529 ± 0.1287	0.0314 ± 0.0764	0.9701 ± 0.0023	0.9512 ± 0.0088	0.9899 ± 0.0048
Test	0.3174 ± 0.2093	0.2063 ± 0.1467	0.9688 ± 0.0105	0.9446 ± 0.0205	0.9947 ± 0.0066
Overall	0.2603 ± 0.2280	0.1696 ± 0.1558	0.9686 ± 0.0099	0.9450 ± 0.0191	0.9935 ± 0.2603

The same cases of Figure 5.14 are represented in Figure 5.15 with the morphological operations applied and the respective isosurfaces.

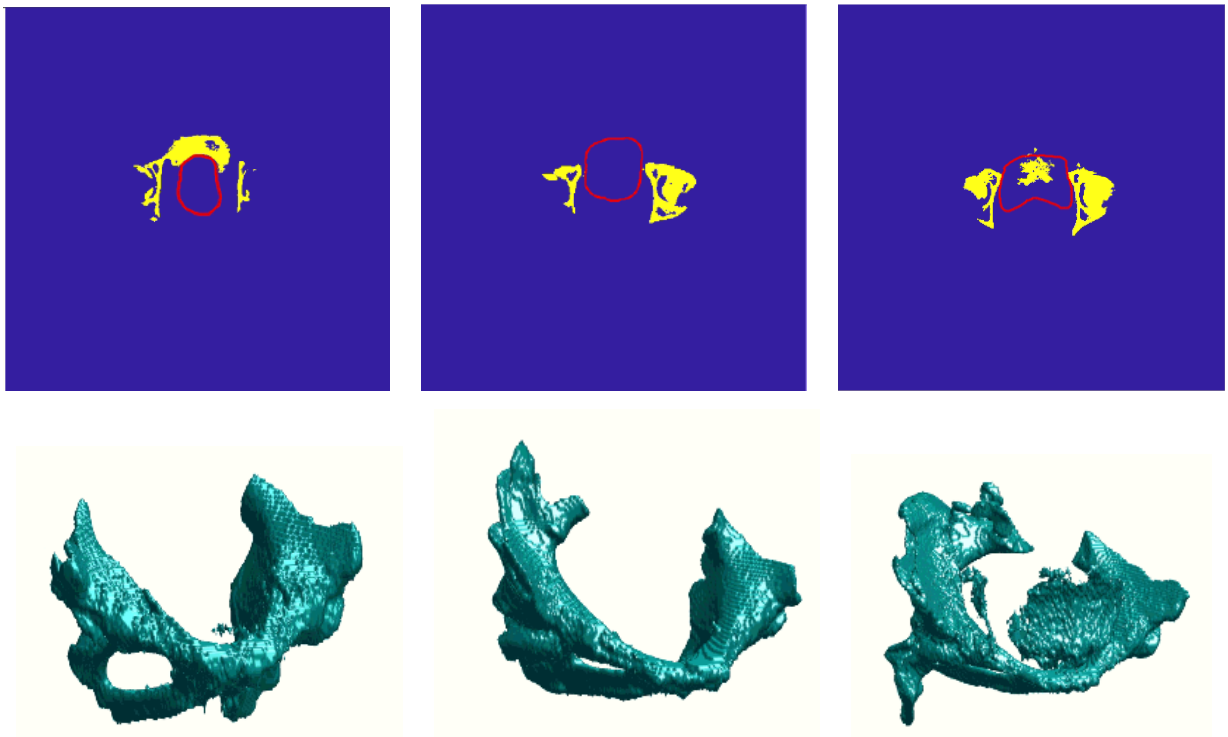


Figure 5.15: Selected examples of Active Contours using UNet results with post-processing with the HU mask

The process was repeated for the data set with the anatomic mask. Once again, the process started by defining the Clustering algorithms as the initial contours. For the anatomic mask, the optimal number of iteration obtained by using the same methodology was 275. The results of active contours implementation to each case using the Chan Vese method are represented in Table 5.17.

Table 5.17: Active Contours using Clustering results with the anatomic mask

Set	DSC	Jaccard	BF score	Precision	Recall
Train	0.0029 ± 0.0099	0.0015 ± 0.0050	0.9016 ± 0.0797	0.8392 ± 0.1169	0.9861 ± 0.0055
Validation	0.0000 ± 0.0000	0.0000 ± 0.0000	0.8988 ± 0.0223	0.8260 ± 0.0388	0.9867 ± 0.0031
Test	0.0000 ± 0.0001	0.0000 ± 0.0000	0.9394 ± 0.0243	0.9004 ± 0.0432	0.9831 ± 0.0043
Overall	0.0011 ± 0.0060	0.0005 ± 0.0031	0.9222 ± 0.0545	0.8717 ± 0.0836	0.9845 ± 0.0011

The results of some cases and the respective isosurfaces can be observed in Figure 5.16.

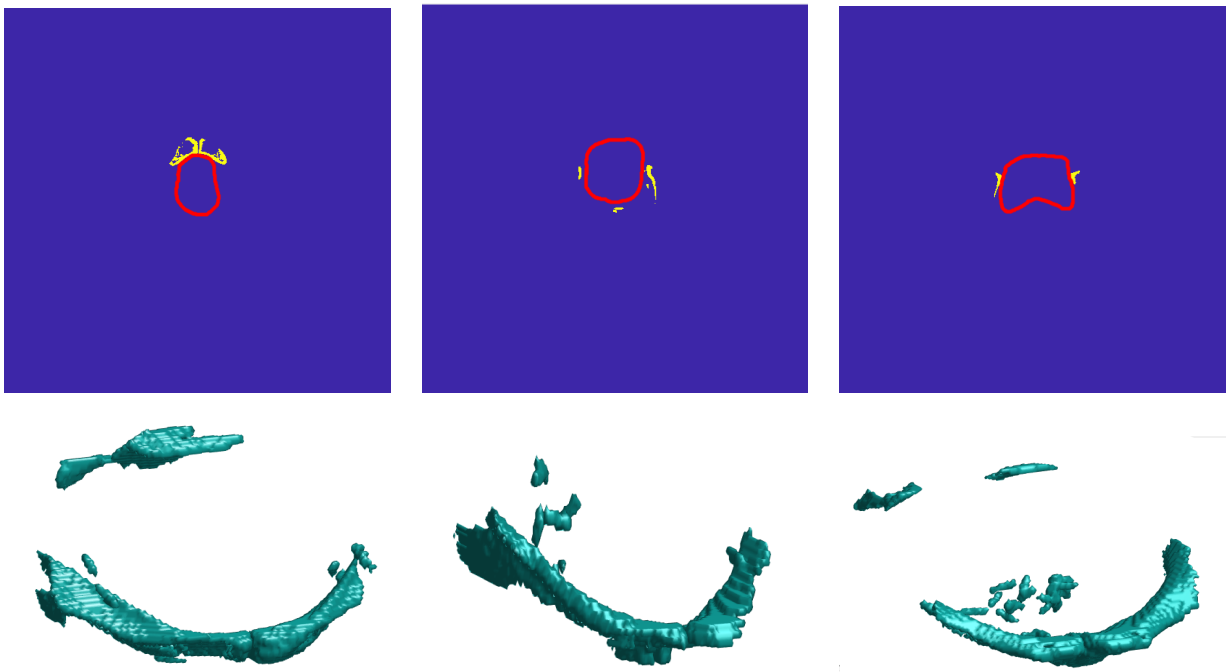


Figure 5.16: Selected examples of Active Contours using clustering results with anatomic mask

Similarly with the HU mask, the U-Net results were also used to define the initial contour. Some results and the respective isosurfaces can be observed in Figure 5.17.

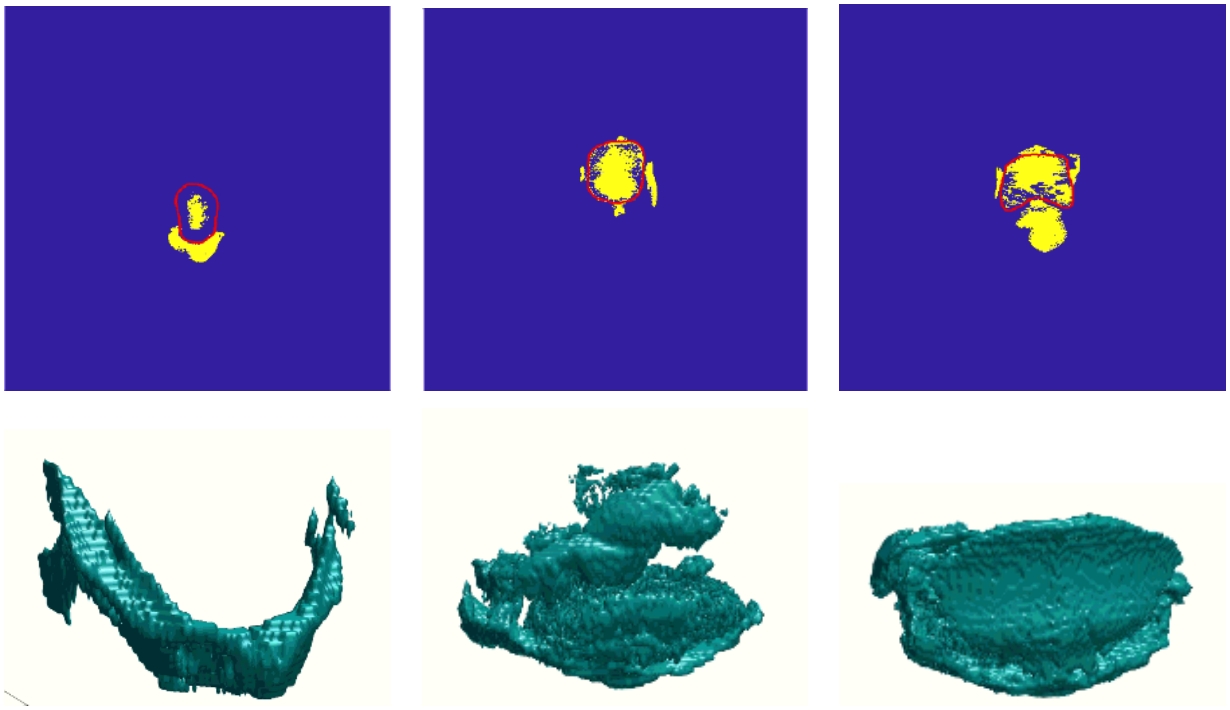


Figure 5.17: Selected examples of Active Contours using the U-Net results with anatomic mask

The results after applying the morphological operations can be consulted in Table 5.18.

Table 5.18: Active Contours results mask using U-Net results with the anatomic mask

Set	DSC	Jaccard	BF score	Precision	Recall
Train	0.2552 ± 0.2677	0.1744 ± 0.1925	0.9637 ± 0.0081	0.9747 ± 0.0184	0.9932 ± 0.0060
Validation	0.2086 ± 0.3231	0.1518 ± 0.2352	0.9911 ± 0.0048	0.9927 ± 0.0014	0.9896 ± 0.0086
Test	0.3579 ± 0.3343	0.2716 ± 0.2743	0.9856 ± 0.0082	0.9775 ± 0.0166	0.9941 ± 0.0069
Overall	0.3069 ± 0.3112	0.2253 ± 0.2459	0.9654 ± 0.0081	0.9778 ± 0.0171	0.9933 ± 0.0059

The cases represented in Figure 5.18 are the same ones that were drawn in Figure 5.17 after the post-processing. The respective isosurfaces are also represented.

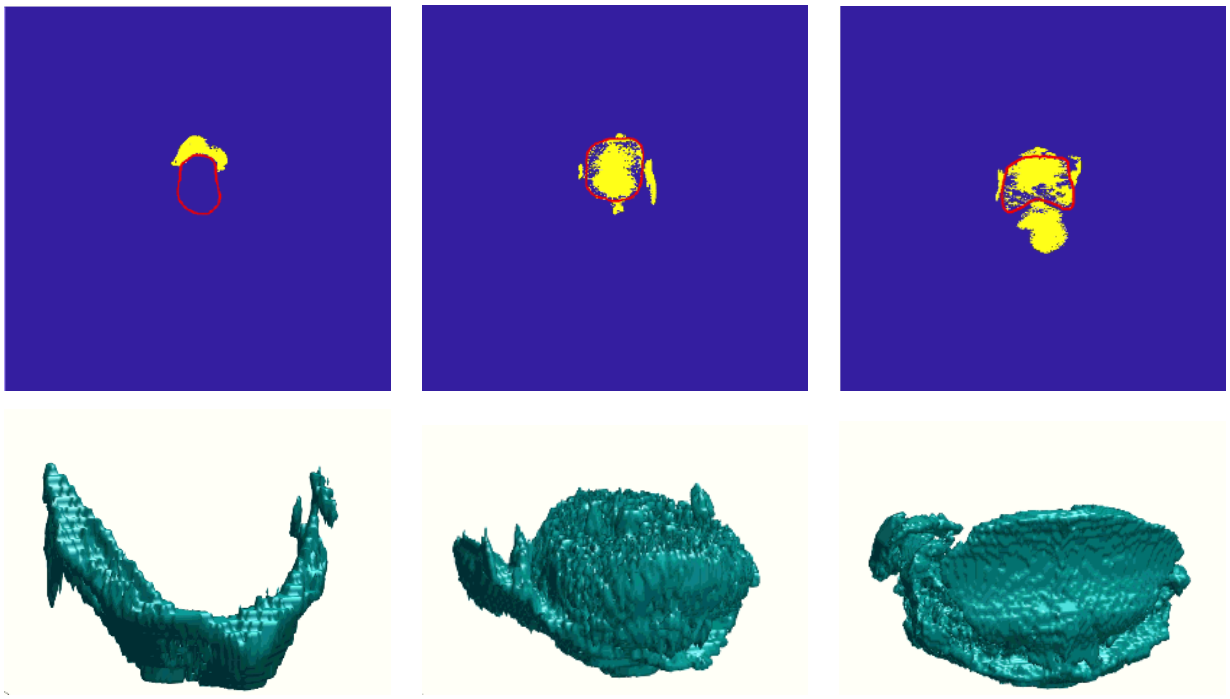


Figure 5.18: Selected examples of Active Contours mask using U-Net results with post-processing with the anatomic mask

5.5 Graph Based

From the estimated superpixels' values, the data set with the mask based on HU values achieved better results by defining the label mask with 400 superpixels. The foremask and backmask were defined with three approaches. In the first one, they were chosen manually. The segmentation results can be analysed in Table 5.19.

Some cases and the respective masks and isosurfaces are represented in Figure 5.19.

Table 5.19: Graph Based results using manual foremask and backmask with the HU mask

Set	DSC	Jaccard	BF score	Precision	Recall
Train	0.2637 ± 0.1332	0.1584 ± 0.0886	0.9805 ± 0.0068	0.9757 ± 0.0134	0.9854 ± 0.0036
Validation	0.2458 ± 0.0854	0.1424 ± 0.0559	0.9831 ± 0.0047	0.9795 ± 0.0089	0.9869 ± 0.0042
Test	0.2681 ± 0.1231	0.1605 ± 0.0829	0.9792 ± 0.0038	0.9746 ± 0.0077	0.9839 ± 0.0051
Overall	0.2646 ± 0.1230	0.1582 ± 0.0823	0.9800 ± 0.0052	0.9754 ± 0.0102	0.9847 ± 0.2646

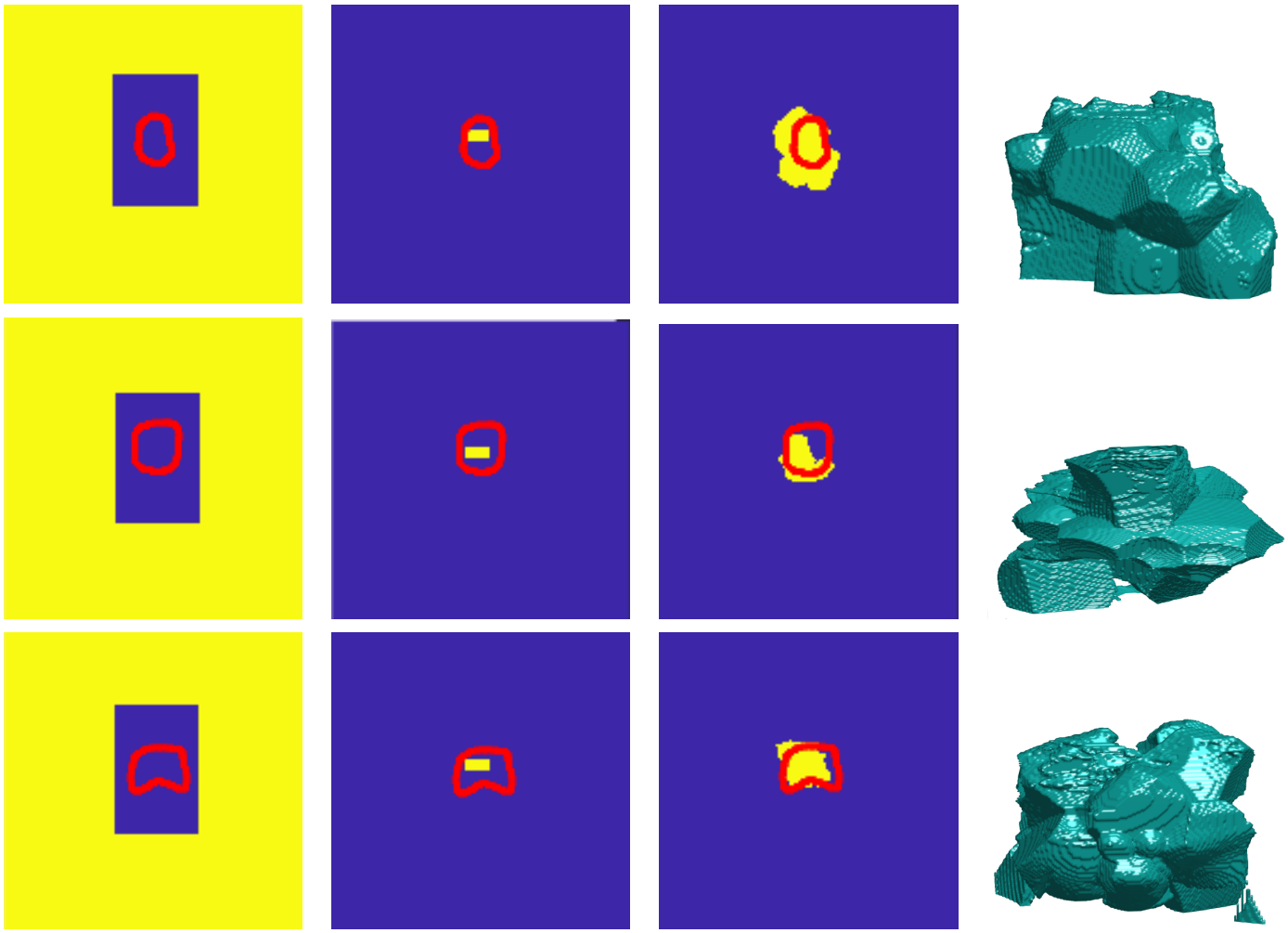


Figure 5.19: Selected results of Graph Based algorithm using HU mask (third column) and respective manual backmasks (first column), foremasks (second column) and isosurfaces (fourth column)

Then, the bladder was segmented by defining the clustering post-processing results as foremask and creating the backmask by implementing a seed found based on ground truth coordinates. The results can be analysed in Table 5.20.

In Figure 5.20, some cases and the respective isosurfaces can be observed.

The masks were defined in a third way. The U-Net results with the anatomic mask were used

Table 5.20: Graph Based results using the Clustering post-processing results with the HU mask

Set	DSC	Jaccard	BF score	Precision	Recall
Train	0.0321 ± 0.0175	0.0164 ± 0.0091	0.7957 ± 0.0716	0.7405 ± 0.0816	0.8617 ± 0.0617
Validation	0.0239 ± 0.0130	0.0121 ± 0.0067	0.8174 ± 0.0299	0.7513 ± 0.0265	0.8963 ± 0.0351
Test	0.0415 ± 0.0319	0.0215 ± 0.0170	0.7941 ± 0.0698	0.7332 ± 0.0772	0.8677 ± 0.0634
Overall	0.0366 ± 0.0267	0.0188 ± 0.0141	0.7967 ± 0.0676	0.7375 ± 0.0753	0.8680 ± 0.0366

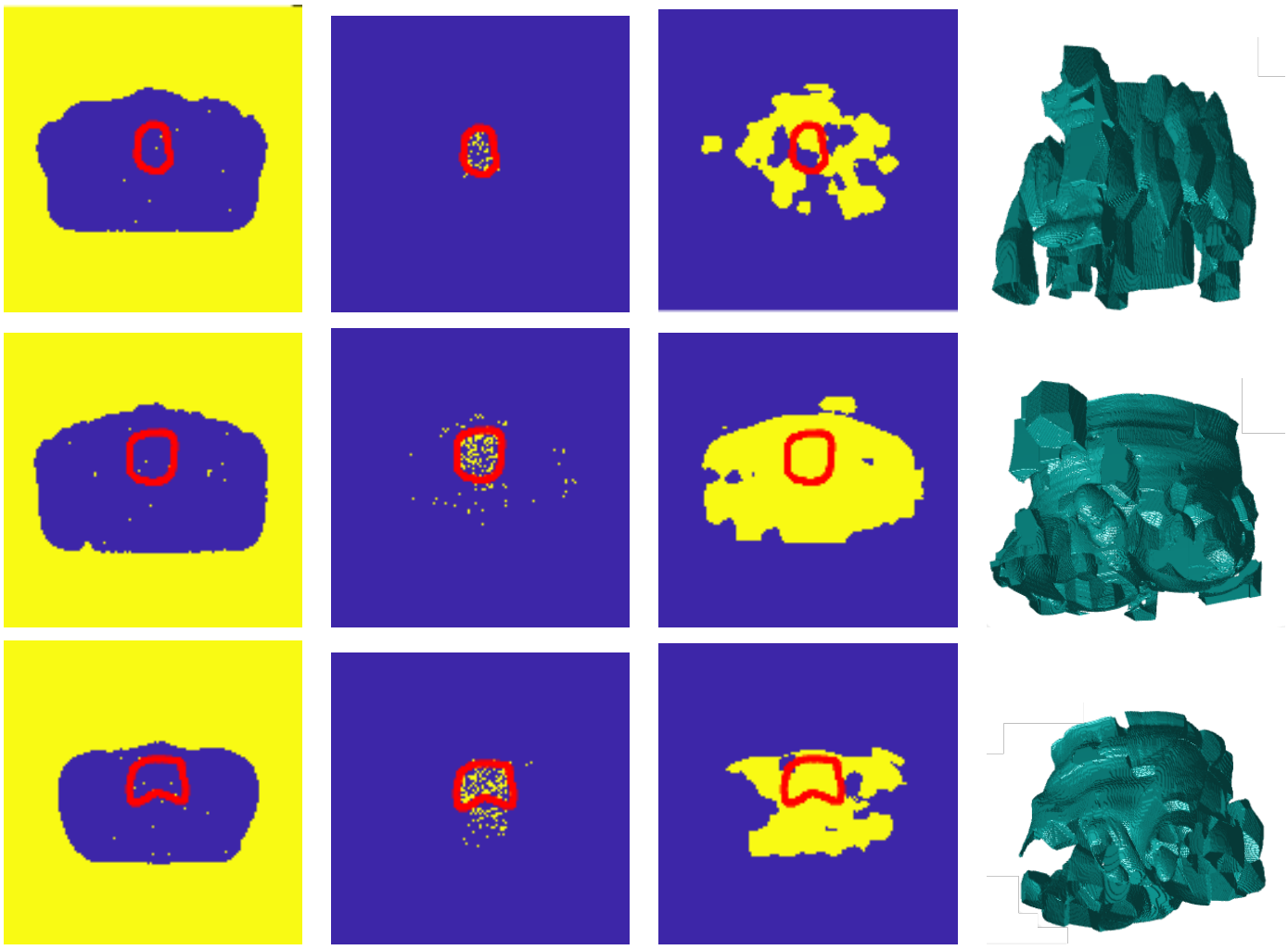


Figure 5.20: Selected results of Graph Based algorithm using the Clustering results (third column) using the clustering results and respective backmasks (first column), foremasks (second column) and isosurfaces (fourth column) with the HU mask

to define the foremask and the manual backmask was used again. The results are represented in Table 5.21. Some results and the respective isosurfaces can be consulted in Figure 5.21.

The process was repeated to the data set with the mask based on anatomy. This time, the optimal number of superpixels was 350. Once again, the foremask and backmask were defined manually. The results are represented in Table 5.22

In Figure 5.22, one can observe some cases and the respective isosurfaces.

As it was done with the mask based on HU values, the clustering post-processing results were used to define the foremask and the backmask was created by implementing a seed found based on ground truth coordinates. The results can be consulted in Table 5.23.

Some results and the respective isosurfaces can be observed in Figure 5.23.

Table 5.21: Graph based results using the U-Net results with the HU mask

Set	DSC	Jaccard	BF score	Precision	Recall
Train	0.2802 ± 0.0855	0.1657 ± 0.0573	0.9681 ± 0.0083	0.9526 ± 0.0130	0.9841 ± 0.0047
Validation	0.2447 ± 0.0728	0.1411 ± 0.0473	0.9703 ± 0.0059	0.9555 ± 0.0101	0.9856 ± 0.0041
Test	0.3045 ± 0.1071	0.1844 ± 0.0776	0.9666 ± 0.0088	0.9516 ± 0.0143	0.9823 ± 0.0053
Overall	0.2903 ± 0.0975	0.1737 ± 0.0690	0.9675 ± 0.0084	0.9523 ± 0.0134	0.9832 ± 0.2903

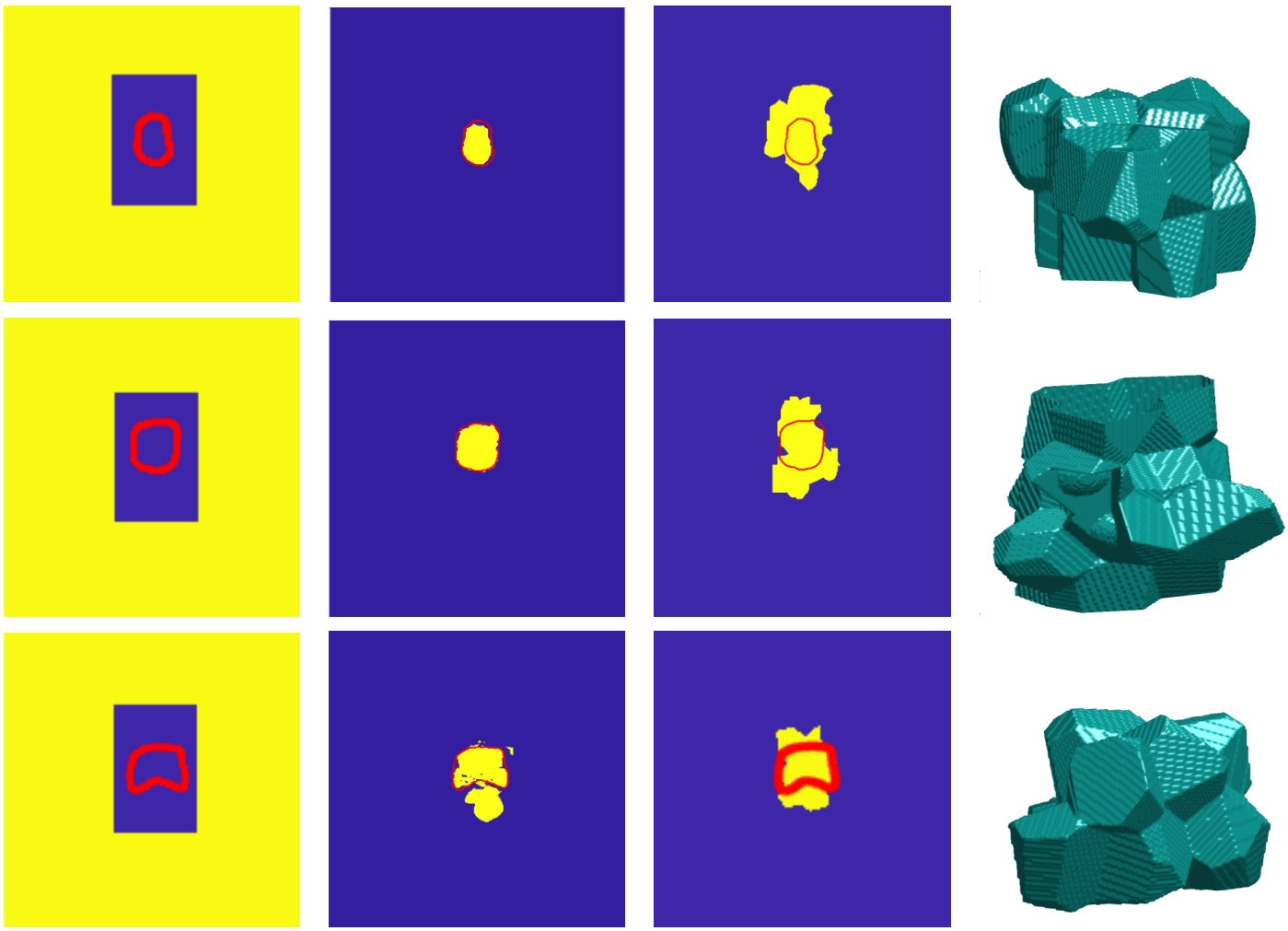


Figure 5.21: Selected results of Graph Based algorithm (third column) using U-Net results and respective backmasks (first column), foremasks (second column) and isosurfaces (fourth column) with the HU mask

Table 5.22: Graph Based results by creating foremask and backmask manually with the anatomic mask

Set	DSC	Jaccard	BF score	Precision	Recall
Train	0.1970 ± 0.0882	0.1119 ± 0.0559	0.9593 ± 0.0096	0.9340 ± 0.0173	0.9861 ± 0.0046
Validation	0.1844 ± 0.0558	0.1025 ± 0.0341	0.9613 ± 0.0115	0.9370 ± 0.0208	0.9870 ± 0.0021
Test	0.2026 ± 0.0905	0.1156 ± 0.0576	0.9675 ± 0.0112	0.9509 ± 0.0217	0.9850 ± 0.0054
Overall	0.1990 ± 0.0863	0.1131 ± 0.0548	0.9640 ± 0.0112	0.9435 ± 0.0215	0.9856 ± 0.1990

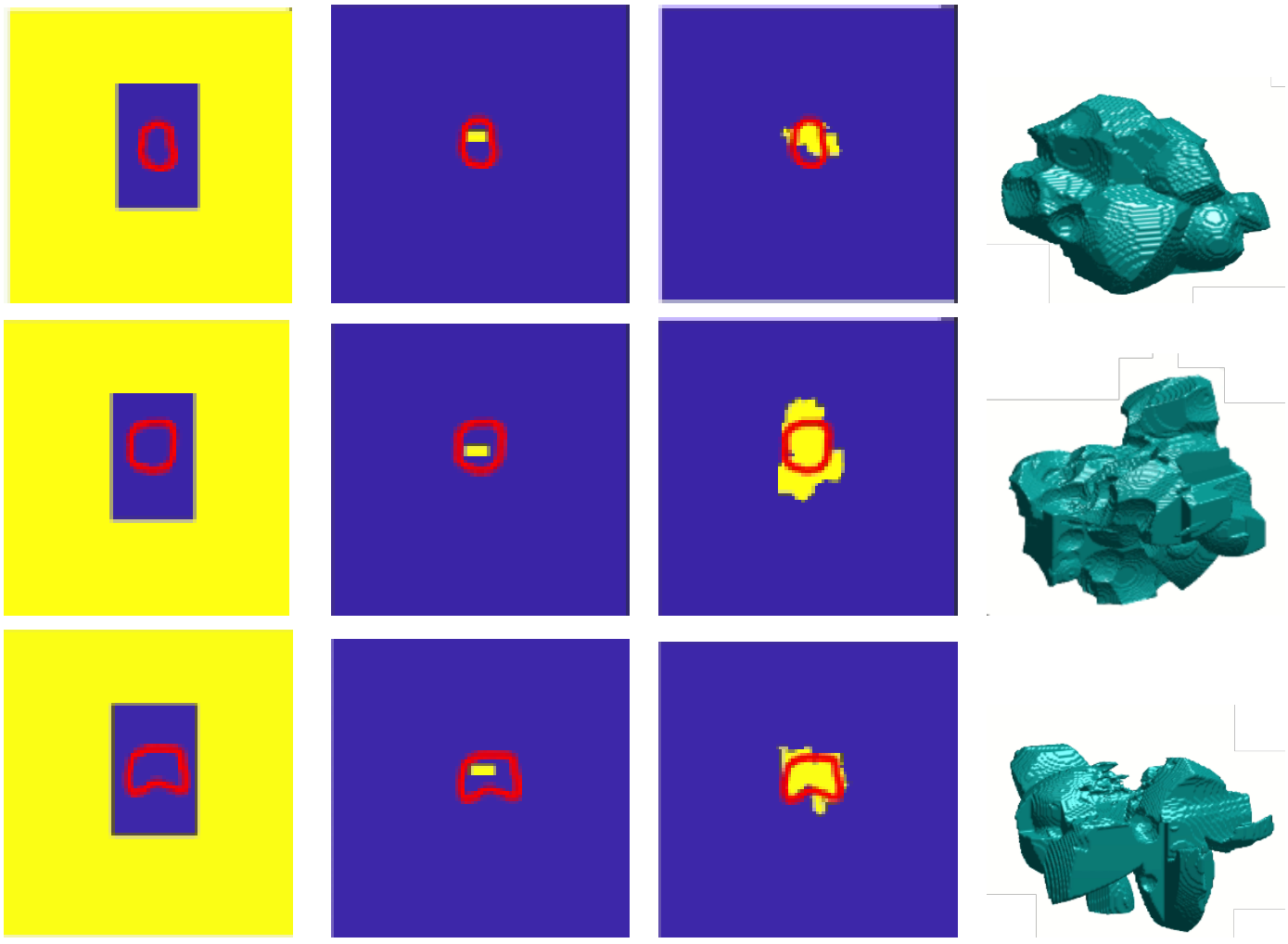


Figure 5.22: Selected results of Graph Based (third column) and respective manual backmasks (first column), foremasks (second column) and isosurfaces (fourth column) with the anatomic mask

At last, the U-Net results with anatomic mask were used as foremask and the backmask was defined manually. The results can be consulted in Table 5.24.

Table 5.23: Graph Based results using the Clustering post-processing results with the anatomic mask

Set	DSC	Jaccard	BF score	Precision	Recall
Train	0.0954 ± 0.0615	0.0512 ± 0.0339	0.9386 ± 0.0170	0.9051 ± 0.0252	0.9748 ± 0.0091
Validation	0.0496 ± 0.0194	0.0255 ± 0.0102	0.9334 ± 0.0092	0.8953 ± 0.0150	0.9750 ± 0.0085
Test	0.0863 ± 0.0719	0.0466 ± 0.0402	0.9373 ± 0.0218	0.9067 ± 0.0350	0.9705 ± 0.0098
Overall	0.0863 ± 0.0655	0.0463 ± 0.0365	0.9374 ± 0.0191	0.9051 ± 0.0301	0.9725 ± 0.0863

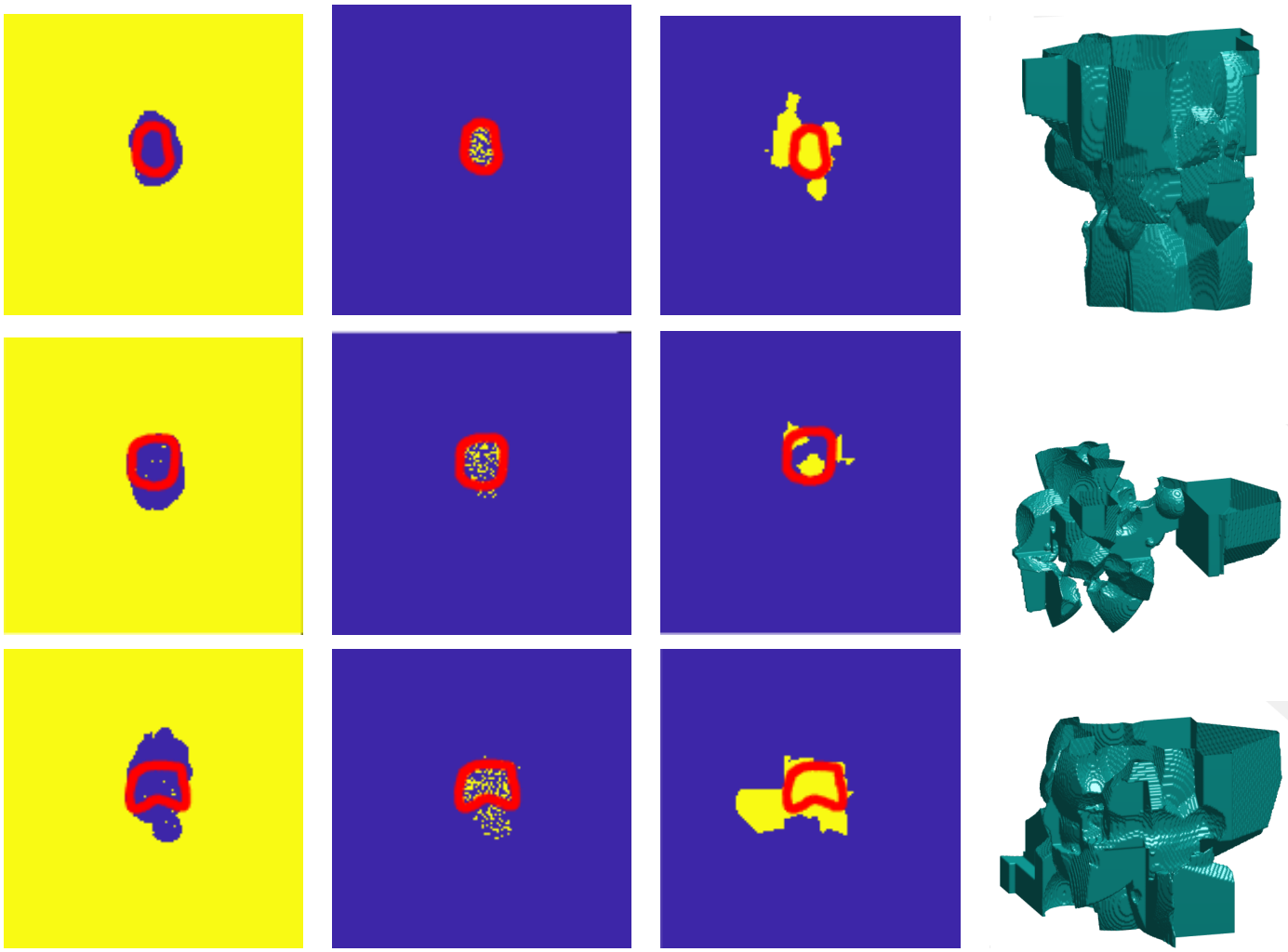


Figure 5.23: Selected results of Graph Based (third column) using the Clustering results and respective backmasks (first column), foremasks (second column) and isosurfaces (fourth column) with the anatomic mask

Table 5.24: Graph based results using U-Net results with the anatomic mask

Set	DSC	Jaccard	BF score	Precision	Recall
Train	0.1892 ± 0.0789	0.1066 ± 0.0494	0.9566 ± 0.0110	0.9311 ± 0.0174	0.9837 ± 0.0060
Validation	0.1666 ± 0.0331	0.0912 ± 0.0198	0.9608 ± 0.0076	0.9389 ± 0.0142	0.9840 ± 0.0040
Test	0.2105 ± 0.0743	0.1195 ± 0.0472	0.9601 ± 0.0152	0.9395 ± 0.0250	0.9819 ± 0.0079
Overall	0.1988 ± 0.0740	0.1123 ± 0.0467	0.9589 ± 0.0132	0.9364 ± 0.0218	0.9827 ± 0.1988

In Figure 5.24 some cases and the respective isosurfaces can be analysed.

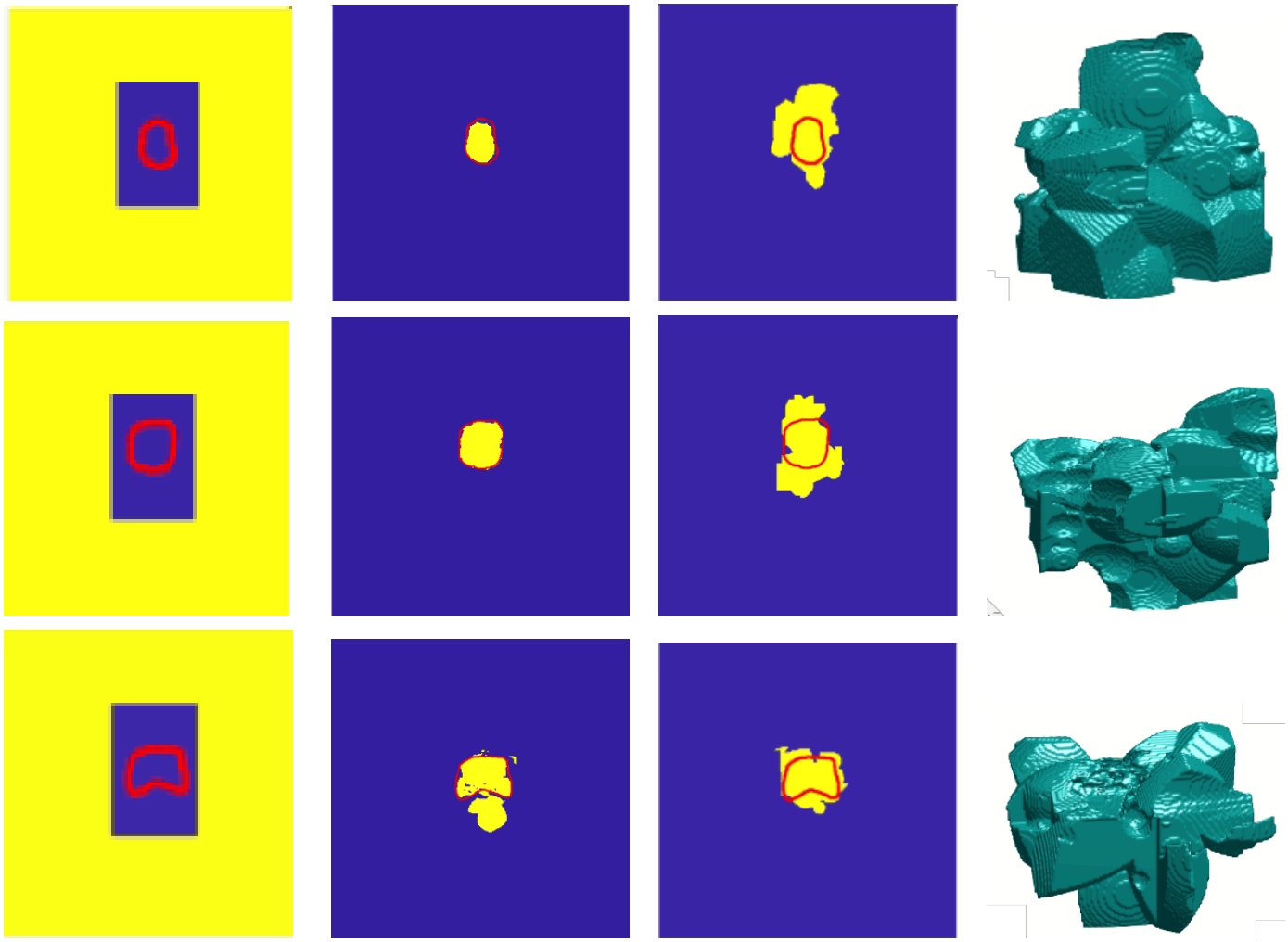


Figure 5.24: Selected results of Graph Based (third column) using U-Net results and respective backmasks (first column), foremasks (second column) and isosurfaces (fourth column) with the anatomic mask

5.6 Conclusion

The results vary greatly depending on the mask used for each algorithm and the algorithm itself.

The RoI detected by thresholding was intended to be smaller, more similar to the one based on anatomy.

Initially, the clustering algorithm achieved better results with the mask based on anatomy than with the mask based on HU values. Two post-processing were applied to both results. The results of the volumes with the mask based on HU values improved with the two approaches,

reaching the best results with the one that uses morphological operations. The improvements were not so noticeable for the results of the volumes with the mask based on anatomy. Actually, the results got worse with the post-processing based on imfill. At the end, the best results were the ones achieved with the mask based in HU values after suffering morphological operations. This approach has the advantage of not requiring previous knowledge but it has the disadvantage of the need to estimating the K values manually.

In U-Net algorithm, the difference between the two RoIs was more accentuated. The results with the mask based on HU values had a lot of background poorly classified. However, the results with the anatomic mask are quite good and are the best of all methods. This method is quite slow, since the network runs separately for each patch but it increases the training data when compared with the number of training images.

Active contours algorithm was the one who computed worst results by defining the clustering results as initial contours, since most of the bladder pixels are classified as background. With the anatomic mask, the results improved slightly, since there are fewer background pixels classified as bladder and those misclassified as background did not decrease much. The results improved in both masks when the initial contours were defined by the U-Net results. In spite of being simple to implement, this algorithm needs an initial contour.

For the graph based algorithm, three approaches were used to achieve the foremask and backmask. Getting the backmask by defining a seed and the foremask as the clustering results after morphological operations led to not so good results. The backmask created manually is bigger and contains more information on the background, giving better results. With the HU mask, the best results were achieved by using the U-Net results as foremask. For the anatomic mask, the results were better by defining the foremask manually. This method is also simple to implement, but has the disadvantage of needing the manual initialization of the foreground and background.

The results can be compared in Table 5.25, where the best results of each algorithm are represented.

Table 5.25: Algorithms comparison. "PP=MO" stands for morphological operations applied as post-processing

Mask	Algorithm	DSC	Jaccard	BF score	Precision	Recall
HU	Clustering (k=115 PP=MO)	0.2364 ± 0.1490	0.1421 ± 0.0972	0.8223 ± 0.1510	0.7269 ± 0.2146	0.9977 ± 0.2364
	U-Net	0.0617 ± 0.0318	0.0321 ± 0.0171	0.7030 ± 0.0580	0.5559 ± 0.0693	0.9648 ± 0.0617
	Active Contours (U-Net)	0.2603 ± 0.2280	0.1696 ± 0.1558	0.9686 ± 0.0099	0.9450 ± 0.0191	0.9935 ± 0.2603
	Graph Based (U-Net)	0.2903 ± 0.0975	0.1737 ± 0.0690	0.9675 ± 0.0084	0.9523 ± 0.0134	0.9832 ± 0.2903
Anatomic	Clustering (k=46 PP=MO)	0.2215 ± 0.1960	0.1381 ± 0.1248	0.6247 ± 0.4533	0.7024 ± 0.3209	0.6387 ± 0.2215
	U-Net	$\mathbf{0.8048 \pm 0.2249}$	$\mathbf{0.7186 \pm 0.2482}$	$\mathbf{0.9981 \pm 0.0036}$	$\mathbf{0.9979 \pm 0.0049}$	$\mathbf{0.9984 \pm 0.8048}$
	Active Contours (U-Net)	0.3069 ± 0.3112	0.2253 ± 0.2459	0.9654 ± 0.0081	0.9778 ± 0.0171	0.9933 ± 0.0059
	Graph Based (manual masks)	0.1990 ± 0.0863	0.1131 ± 0.0548	0.9640 ± 0.0112	0.9435 ± 0.0215	0.9856 ± 0.1990

Chapter 6.

Conclusions and directions for future work

Different approaches were studied with the aim of helping specialists to detect the bladder more quickly and effectively. The approach whose results stand out is using the anatomic mask to detect the RoI, implementing the U-Net algorithm to predict the label of each volume and applying morphological operations to select the biggest connect region.

Two pre-procedures were implemented to detect the RoI. Using GT information, it was possible to create a smaller mask than the one based on the HU values of the bladder. Consequently, the results were better with the smaller one, which DSC, Jaccard, BF score, Precision and Recall were 31%, 19%, 96%, 94% and 98%, respectively. The mask based on HU values were evaluated on the same metrics and the results were 3%, 1%, 81%, 80% and 82%, respectively. Those two masks were applied to each volume, creating two RoIs for each one. Four methods were implemented to each RoI of each volume.

By applying K-means algorithm, the images were divided in clusters. From these clusters, one was chosen to be the one corresponding to the bladder. The classifiers RUSBoost and Classification Trees were used to make this choice. To improve the segmentations, two approaches were used. With morphological operations, the segmentations in the data set with the mask based on HU values achieved 24%, 14%, 82%, 73% and 100% of DSC, Jaccard, BF score, Precision and Recall, respectively. With the mask based on anatomy, the results evaluated with the same metrics were 22%, 14%, 62%, 70% and 64%, respectively. U-Net was the algorithm whose results vary the most for each mask. With morphological operations, this method obtained 6% of DSC, 3% of Jaccard, 70% of BF score, 56% of Precision and 96% by using the data set with the mask based on HU values. With the one based on anatomy, the DSC was 80%, Jaccard was

72% and BF score, Precision and Recall were 100%. Clustering and U-Net results were used to define the initial contours in active contours algorithm. This method achieved the worst results by using the clustering results but it improved by using UNet's: 26% of DSC, 17% of Jaccard, 97% of BF score, 95% of Precision and 99% of Recall for the data set with mask based on HU values applied and 31% of DSC, 23% of Jaccard, 97% of BF score, 98% of Precision and 99% of Recall for the mask based on anatomy. A foremask and a backmask were created to design the pixels in the foreground and background for Graph Based algorithm. For the HU mask, the results were better when the backmask was manually computed and the foremask was defined by U-Net results: DSC was 29%, Jaccard was 17%, BF score was 97%, Precision was 95% and Recall was 98%. For the data set with the mask based on anatomy, the best results were achieved by defining the backmask and foremask manually: 20%, 11%, 96%, 94% and 99% for DSC, Jaccard, BF score, Precision and Recall, respectively.

In the future, it is intend to use hyperparameter optimization techniques to choose the optimal parameters in each algorithm, to obtain better results. Since the rectum is an OR in context of prostate cancer, it is planned to implement the previous methods, with the aim of studying whether their performance is similar by segmenting the rectum.

Bibliography

[1] Jesús Rogel-Salazar. *Data science and analytics with python*. Springer Singapore, 2018, pp. 1–376. ISBN: 9781498742115. DOI: 10.1201/9781315151670. URL: http://dx.doi.org/10.1007/978-981-10-8527-7.

[2] National Institute of Biomedical Imaging and Bioengineering. *Computed Tomography (CT)*. URL: https://www.nibib.nih.gov/science-education/science-topics/computed-tomography-ct (visited on 02/22/2021).

[3] Associação Portuguesa de Urologia. *Cancro da Próstata*. Apr. 2010. URL: https://www.apurologia.pt/publico/frameset.htm?https://www.apurologia.pt/publico/cancro_da_prostata.htm (visited on 01/05/2021).

[4] eurostat. *Deaths from prostate cancer in EU regions*. Nov. 2019. URL: https://ec.europa.eu/eurostat/web/products-eurostat-news/-/EDN-20191119-2 (visited on 01/05/2021).

[5] Neil G. Burnet et al. "Defining the tumour and target volumes for radiotherapy". In: *Cancer Imaging* 4.2 (2004), pp. 153–161. ISSN: 14707330. DOI: 10.1102/1470-7330.2004.0054.

[6] J. Grimaud et al. "Quantification of MRI lesion load in multiple sclerosis: A comparison of three computer-assisted techniques". In: *Magnetic Resonance Imaging* 14.5 (1996), pp. 495–505. ISSN: 0730725X. DOI: 10.1016/0730-725X(96)00018-5.

[7] Leo Joskowicz et al. "Inter-observer variability of manual contour delineation of structures in CT". In: *European Radiology* 29.3 (2019), pp. 1391–1399. ISSN: 14321084. DOI: 10.1007/s00330-018-5695-5.

[8] MathWorks. *strel*. URL: https://www.mathworks.com/help/images/ref/strel.html (visited on 04/05/2021).

[9] Image Processing. *7 . Morphological operations on binary images*. URL: https://www.cs.auckland.ac.nz/courses/compsci773s1c/lectures/ImageProcessing-html/topic4.htm (visited on 04/05/2021).

[10] David Shih Chun Jin, Chien Hao Chu, and Jyh Cheng Chen. "Trabecular bone morphological analysis for preclinical osteoporosis application using micro computed tomography scanner". In: *Journal of Medical and Biological Engineering* 36.1 (2016), pp. 96–104. ISSN: 21994757. DOI: 10.1007/s40846-016-0109-0.

[11] MathWorks. *Image Processing Toolbox User's Guide*. URL: http://matlab.izmiran.ru/help/toolbox/images/morph12.html (visited on 04/05/2021).

[12] Dzung L Pham, Chenyang Xu, and Jerry L Prince. "Current Methods in Medical Image Segmentation". In: (2000).

[13] Tim Mcinerney and Demetri Terzopoulos. "Deformable models in medical image analysis: a survey". In: 1.2 (1996), pp. 91–108.

[14] Hongwei Yue et al. "A markov random field model for image segmentation of rice planthopper in rice fields". In: *Journal of Engineering Science and Technology Review* 9.2 (2016), pp. 31–38. ISSN: 17912377. DOI: 10.25103/jestr.092.06.

[15] "We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists TOP 1 %". In: *Intech* tourism (2016), p. 13. URL: https://www.intechopen.com/books/advanced-biometric-technologies/liveness-detection-in-biometrics.

[16] Yi Gao and Allen Tannenbaum. "Combining atlas and active contour for automatic 3D medical image segmentation". In: *Proceedings - International Symposium on Biomedical Imaging* May 2014 (2011), pp. 1401–1404. ISSN: 19457928. DOI: 10.1109/ISBI.2011.5872662.

[17] K. Santle Camilus and V. K. Govindan. "A Review on Graph Based Segmentation". In: *International Journal of Image, Graphics and Signal Processing* 4.5 (2012), pp. 1–13. ISSN: 20749074. DOI: 10.5815/ijigsp.2012.05.01.

[18] Hydin John and J Anitha. "A Study of Image Segmentation Approaches". In: *International Journal of Advanced Research in Electronics and Communication Engineering* 1.4 (2012), pp. 62–68.

[19] Keh Shih Chuang et al. "Fuzzy c-means clustering with spatial information for image segmentation". In: *Computerized Medical Imaging and Graphics* 30.1 (2006), pp. 9–15. ISSN: 08956111. DOI: 10.1016/j.compmedimag.2005.10.001.

[20] Machine Learning Mastery. *A Gentle Introduction to Expectation-Maximization (EM Algorithm)*. URL: https://machinelearningmastery.com/expectation-maximization-em-algorithm/ (visited on 08/28/2020).

[21] Programmer Sought. *Python-based decision tree classifier and pruning*. URL: https://www.programmersought.com/article/29374957715/ (visited on 08/17/2021).

[22] Mohammed J. Islam et al. "Investigating the performance of naïve- bayes classifiers and K- nearest neighbor classifiers". In: *Journal of Convergence Information Technology* 5.2 (2010), pp. 133–137. ISSN: 19759320. DOI: 10.4156/jcit.vol5.issue2.15.

[23] Yash Rawat. *Introduction to SVM*. Apr. 2020. URL: https://medium.datadriveninvestor.com/introduction-to-svm-5ea47373a03b (visited on 05/28/2021).

[24] Toward data science. *The differences between Artificial and Biological Neural Networks*. URL: https://towardsdatascience.com/the-differences-between-artificial-and-biological-neural-networks-a8b46db828b7 (visited on 04/28/2021).

[25] Yann Lecun, Yoshua Bengio, and Geoffrey Hinton. "Deep learning". In: *Nature* 521.7553 (2015), pp. 436–444. ISSN: 14764687. DOI: 10.1038/nature14539.

[26] Davide Bacciu et al. "Bioinformatics and medicine in the era of deep learning". In: *European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN)* April (2018), pp. 345–354. arXiv: 1802.09791.

[27] Kwang Gi Kim. "Deep learning book review". In: *Nature* 29.7553 (2019), pp. 1–73.

[28] Md Zahangir Alom et al. "A state-of-the-art survey on deep learning theory and architectures". In: *Electronics (Switzerland)* 8.3 (2019), pp. 1–67. ISSN: 20799292. DOI: 10.3390/electronics8030292.

[29] Geert Litjens et al. "A survey on deep learning in medical image analysis". In: *Medical Image Analysis* 42.December 2017 (2017), pp. 60–88. ISSN: 13618423. DOI: 10.1016/j.media.2017.07.005. arXiv: 1702.05747.

[30] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. "U-Net: Convolutional Networks for Biomedical Image Segmentation". In: *International Conference on Medical image computing and computer-assisted intervention*. 2015, pp. 234–241. ISBN: 978-3-319-24573-7. DOI: 10.1007/978-3-319-24574-4_28. arXiv: 1505.04597.

[31] Pattern Recognition and Image Processing. *U-Net: Convolutional Networks for Biomedical Image Segmentation*. URL: https://lmb.informatik.uni-freiburg.de/people/ronneber/u-net/ (visited on 03/30/2021).

[32] Abdel Aziz Taha and Allan Hanbury. "Metrics for evaluating 3D medical image segmentation: Analysis, selection, and tool". In: *BMC Medical Imaging* 15.1 (2015). ISSN: 14712342. DOI: 10.1186/s12880-015-0068-x. URL: http://dx.doi.org/10.1186/s12880-015-0068-x.

[33] Xiaoli Tang et al. "Geometric-model-based segmentation of the prostate and surrounding structures for image-guided radiotherapy". In: ed. by Sethuraman Panchanathan and Bhaskaran Vasudev. Spie Digital Library, 2004, p. 168. DOI: 10.1117/12.526016.

[34] Mikael Rousson et al. *LNCS 3765 - Constrained Surface Evolutions for Prostate and Bladder Segmentation in CT Images*. Tech. rep. 2005.

[35] C. A. McBain et al. "Early clinical evaluation of a novel three-dimensional structure delineation software tool (SCULPTER) for radiotherapy treatment planning". In: *The British Journal of Radiology* 81.968 (2008), pp. 643–652. ISSN: 0007-1285. DOI: 10.1259/bjr/81762224.

[36] Derek Merck et al. "Training models of anatomic shape variability". In: *Medical Physics* 35.8 (2008), pp. 3584–3596. ISSN: 00942405. DOI: 10.1118/1.2940188.

[37] Dominique P Huyskens et al. "A qualitative and a quantitative analysis of an auto-segmentation module for prostate cancer". In: *Radiotherapy and Oncology* 90.3 (2009), pp. 337–345. ISSN: 01678140. DOI: 10.1016/j.radonc.2008.08.007.

[38] Oscar Acosta et al. "Atlas Based Segmentation and Mapping of Organs at Risk from Planning CT for the Development of Voxel-Wise Predictive Models of Toxicity in Prostate Radiotherapy". In: *Springer*, 2010, pp. 42–51. DOI: 10.1007/978-3-642-15989-3_6.

[39] Oscar Acosta et al. "Evaluation of multi-atlas-based segmentation of CT scans in prostate cancer radiotherapy". In: *IEEE International Symposium on Biomedical Imaging: From Nano to Macro*. IEEE, 2011, pp. 1966–1969. ISBN: 978-1-4244-4127-3. DOI: 10.1109/ISBI.2011.5872795.

[40] W Li et al. "SU-E-J-104: Evaluation of Atlas-Based Auto-Segmentation on Daily In-Room CT for Prostate Cancer". In: *Medical Physics* 39.6Part7 (2012), pp. 3676–3976. ISSN: 00942405. DOI: 10.1118/1.4734940.

[41] N Lay et al. "A Context Integration Framework for Rapid Multiple Organ Parsing". In: *Medical Image Recognition, Segmentation and Parsing*. Ed. by S Kevin Zhou. Elsevier, 2016, pp. 209–230. ISBN: 978-0-12-802581-9. DOI: 10.1016/B978-0-12-802581-9.00010-X.

[42] S. Kevin Zhou and D. Xu. "A Probabilistic Framework for Multiple Organ Segmentation Using Learning Methods and Level Sets". In: *Medical Image Recognition, Segmentation and Parsing*. Ed. by S Kevin Zhou. Elsevier, 2016, pp. 157–178. ISBN: 978-0-12-802581-9. DOI: 10.1016/B978-0-12-802581-9.00008-1.

[43] Marshall Gordon et al. "Segmentation of inner and outer bladder wall using deep-learning convolutional neural network in CT urography". In: *Medical Imaging: Computer-Aided Diagnosis*. Ed. by Samuel G. Armato and Nicholas A. Petrick. Vol. 10134. SPIE, 2017, p. 1013402. DOI: 10.1117/12.2255528.

[44] Xue Dong et al. "Synthetic MRI-aided multi-organ segmentation on male pelvic CT using cycle consistent deep attention network". In: *Radiotherapy and Oncology* 141 (2019), pp. 192–199. ISSN: 01678140. DOI: 10.1016/j.radonc.2019.09.028.

[45] Jan Schreier et al. "Clinical evaluation of a full-image deep segmentation algorithm for the male pelvis on cone-beam CT and CT". In: *Radiotherapy and Oncology* 145 (2020), pp. 1–6. ISSN: 01678140. DOI: 10.1016/j.radonc.2019.11.021.

[46] Hanna Sartor et al. "Auto-segmentations by convolutional neural network in cervical and anorectal cancer with clinical structure sets as the ground truth". In: *Clinical and Translational Radiation Oncology* 25 (2020), pp. 37–45. ISSN: 24056308. DOI: 10.1016/j.ctro.2020.09.004.

[47] Dong Joo Rhee et al. "Automatic contouring system for cervical cancer using convolutional neural networks". In: *Medical Physics* 47.11 (2020), pp. 5648–5658. ISSN: 0094-2405. DOI: 10.1002/mp.14467.

[48] Daguang Zhang et al. "Automatic segmentation and applicator reconstruction for CT-based brachytherapy of cervical cancer using 3D convolutional neural networks". In: *Journal of Applied Clinical Medical Physics* 21.10 (2020), pp. 158–169. ISSN: 1526-9914. DOI: 10.1002/acm2.13024.

[49] Shuai Wang et al. "Boundary Coding Representation for Organ Segmentation in Prostate Cancer Radiotherapy". In: *IEEE Transactions on Medical Imaging* PP (2020), pp. 1–1. ISSN: 0278-0062. DOI: 10.1109/TMI.2020.3025517.

[50] Mark H F Savenije et al. "Clinical implementation of MRI-based organs-at-risk auto-segmentation with convolutional networks for prostate radiotherapy". In: *Radiation Oncology* 15.1 (2020), p. 104. ISSN: 1748-717X. DOI: 10.1186/s13014-020-01528-0.

[51] Yuki Aoki et al. "The utility of a deep learning-based algorithm for bone scintigraphy in patient with prostate cancer". In: *Annals of Nuclear Medicine* 34.12 (2020), pp. 926–931. ISSN: 0914-7187. DOI: 10.1007/s12149-020-01524-0.

[52] Feng Shi, Jie Yang, and Yue Min Zhu. "Automatic segmentation of bladder in CT images". In: *Journal of Zhejiang University: Science A* 10.2 (2009), pp. 239–246. ISSN: 1673565X. DOI: 10.1631/jzus.A0820157.

[53] Xiangfei Chai et al. "Automatic bladder segmentation on CBCT for multiple plan ART of bladder cancer using a patient-specific bladder model". In: *Physics in Medicine and*

Biology 57.12 (2012), pp. 3945–3962. ISSN: 00319155. DOI: 10.1088/0031-9155/57/12/3945.

[54] A. J.A.J. Van De Schoot et al. "Generic method for automatic bladder segmentation on cone beam CT using a patient-specific bladder shape model". In: *Medical Physics* 41.3 (2014). ISSN: 00942405. DOI: 10.1118/1.4865762.

[55] Kenny H. Cha et al. "Urinary bladder segmentation in CT urography using deep-learning convolutional neural network and level sets". In: *Medical Physics* 43.4 (2016), pp. 1882–1886. ISSN: 00942405. DOI: 10.1118/1.4944498.

[56] Xuanang Xu, Fugen Zhou, and Bo Liu. "Automatic bladder segmentation from CT images using deep CNN and 3D fully connected CRF-RNN". In: *International Journal of Computer Assisted Radiology and Surgery* 13.7 (2018), pp. 967–975. ISSN: 18616429. DOI: 10.1007/s11548-018-1733-7. URL: https://doi.org/10.1007/s11548-018-1733-7.

[57] Xiangyuan Ma et al. "U-Net based deep learning bladder segmentation in CT urography". In: *Medical Physics* 46.4 (2019), pp. 1752–1765. ISSN: 00942405. DOI: 10.1002/mp.13438.

[58] MathWorks. *3-D Brain Segmentation Using Deep Learning*. URL: https://www.mathworks.com/help/deeplearning/ug/segment-3d-brain-tumor-using-deep-learning.html?searchHighlight=3-D%20Brain%20Tumor%20Segmentation%20Using%20Deep%20Learning&s_tid=srchtitle# (visited on 03/06/2021).

[59] Machine Learning Mastery. *Gentle Introduction to the Adam Optimization Algorithm for Deep Learning*. July 2017. URL: https://machinelearningmastery.com/adam-optimization-algorithm-for-deep-learning/ (visited on 01/13/2021).

[60] Bing Li and Scott T. Acton. "Active contour external force using vector field convolution for image segmentation". In: *IEEE Transactions on Image Processing* 16.8 (2007), pp. 2096–2106. ISSN: 10577149. DOI: 10.1109/TIP.2007.899601.

[61] Pascal Getreuer. "Chan-Vese Segmentation". In: *Image Processing On Line* 2 (2012), pp. 214–224. ISSN: 2105-1232. DOI: 10.5201/ipol.2012.g-cv.

Appendices

Chapter A.

Scientific Paper

Comparison of bladder segmentation techniques in CT scans

Ana Couto
AnaSofia1995@hotmail.com
Inês Domingues
inesdomingues@gmail.com
João Santos
joao.santos@ipoporto.min-saude.pt

Faculdade de Ciências da Universidade do Porto
Portugal
Medical Physics, Radiobiology and Radiation Protection Group, IPO Porto Research Centre (CI-IPOP), Portugal
Instituto de Ciência Biomédicas Abel Salazar
Porto, Portugal

Abstract

Radiotherapy takes a very important role in cancer treatment. One of its necessary steps is to segment the Organs at Risk. This process is currently done manually, which is time consuming and subject to human error. With the goal of helping the specialists and improving segmentation accuracy, some algorithms were tested to segment the bladder in 47 Computerized Tomography scans from patients with prostate cancer provided by the Institute of Oncology of Porto. The four algorithms were evaluated and the Dice obtained for applying Clustering, U-Net, Active Contours and Graph Based were 24%, 6%, 26% and 29%, respectively, in the data set with the mask based on HU values. For anatomic mask, the same metric for the same algorithms were 22%, 80%, 31% and 20%, respectively.

1 Introduction

Surgery and radiotherapy are the most efficient and used treatments for cancer. 60% of the patients submitted to radiotherapy are treated with curative intent but it also has an important role in the reduction of symptoms [1]. In order to find a balance between eradicating the tumour and sparing the surrounding tissues, it is important to observe in detail where the tumour ends and the surrounding organs begin. This procedure is made manually, which is time consuming and subject to human error, namely variability between different contours made by different specialists (inter-variability) and variability between different contours made by the same specialist (intra-variability).

The main goal of this work is to study different segmentation algorithms to apply to prostate cancer patients' CT scans provided by the Institute of Oncology of Porto (IPO). The organ of interest is the bladder. By doing so, it is intended to reduce the time consuming manual segmentation and improve its accuracy.

The remaining of this paper is organised as follows. Section 2 describes the segmentation techniques evaluated and compared in the present work. The dataset is presented in Section 3. Section 4 gives the results while in Section 5 some conclusions and directions for future work are given.

2 Segmentation Methods

Two pre-procedures were developed in order to detect the Region of Interest (RoI): one based on thresholding each CT using Hounsfield Unit (HU) value of the bladder and another one based on the anatomy. These pre-procedures were applied to each of the 71 bladders. The segmentation algorithms are described next.

2.1 Clustering

The clustering algorithm used was K-means. k values between 5 and 195 were tested by steps of 10. The next step is to automatically choose one cluster. With this purpose, 5 features were extracted from each cluster: volume, diameter and the maximum, minimum and mean intensity. Several classifiers were trained: Classification Trees (CT), Discriminant Analysis (DA), k Nearest Neighbors (kNN), Naive Bayes (NB), Support Vector Machine (SVM), Classification Ensembles (CE), Classification Tree Ensembles (CTE) and RUSboost. Two different post processing were applied to the selected cluster. In the first one, the regions and holes were filled by implementing flood fill. The second approach consists of choosing the biggest connect region. The full pipeline of the Clustering segmentation is shown in Figure 1.

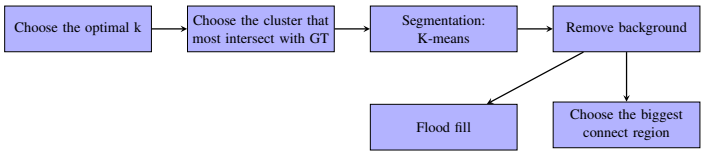


Figure 1: Clustering Process

2.2 U-Net

This algorithm was based and adapted from [2]. Training a network on the full input volume is impractical due to the amount of memory needed to store and process 3-D volumes. This problem is solved by training the network on image patches extracted from the ground truth images. To prevent overfitting due to data limited size, the training and validation data were augmented by randomly rotating and reflecting training data to make the training more robust. In order to avoid border artefacts when using the overlap-tile strategy for prediction of the test volumes, valid convolution padding was specified. The overlap-tile strategy was used to predict the labels for each test volume. The full pipeline of the U-Net segmentation is shown in Figure 2.

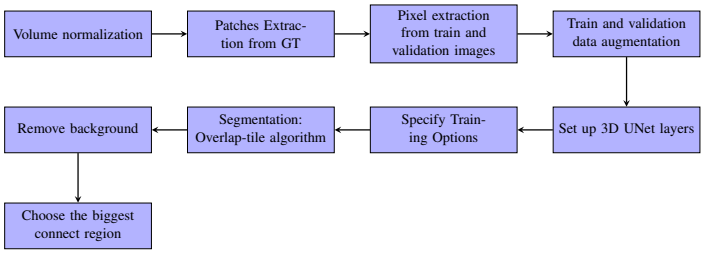


Figure 2: U-Net Process

2.3 Active Contours

Active contours algorithm, also known as snakes, consists of deforming the image domain and capture a desired feature through the constraint and image forces that pull it towards object contours and the internal forces resist deformation [3]. Two approaches to define the initial contour were used: the clustering results and the U-Net results. The optimal number of iteration was found by evaluating some cases with the metrics Dice, Jaccard and the BF (Boundary F1) contour matching score. Values between 150 and 325 with steps of 25 were evaluated. The segmentation results by defining the U-Net results as initial contours, went through the same morphological operations as the ones used to create the mask based on HU values. The results by defining the clustering results as initial contours did not, once there were no improvements. The full pipeline of the Active Contours segmentation is shown in Figure 3.

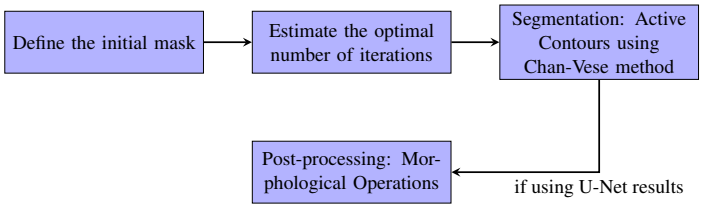


Figure 3: Active Contours Process

2.4 Graph Based

Graph-based image processing methods typically operate on pixel adjacency graphs. Adjacency graphs are graphs whose vertex set V is the set of image elements, and whose edge set E is given by an adjacency relation on the image elements. Initially, the optimal number of superpixels were estimated. For the HU mask, a label mask was computed by creating 400 superpixels on the image to segment and 350 for the anatomic mask. Then, a foreground mask and a background mask were created.These masks were created with three approaches: defining them manually by choosing ranges where the bladder is fully in (for the foreground mask) and where the bladder is fully out (for the background mask); using the clustering results (see Section 2.1) as the foreground mask and defining the background mask as a seed found based on ground truth coordinates; using the U-Net results (see Section 2.2) as the foreground mask and defining the background mask manually, once again. The full pipeline of the Graph Based segmentation is shown in Figure 4.

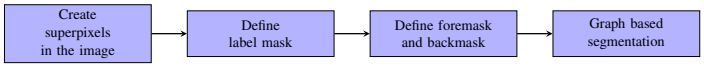


Figure 4: Graph Based Process

3 Dataset

The database was composed by 48 CT scans from 48 patients with prostate cancer, collected at the Institute of Oncology of Porto (IPO). Each CT has the structures manually segmented by specialists. Once most of CTs have more than one manual segmentation of the bladder, there are 71 bladders manually segmented. For some methods, it was needed to split the dataset into train, validation and test datasets. The train dataset consists of 19 patients, the validation has 4 and the remaining 24 went to test dataset. This leads to 37% of the structures to segment in the train dataset, 8% in validation and 55% in the test set.

4 Results

Segmentation illustrations are given in Figure 5, while the best results for each algorithm are given in Table 1.

Table 1: Algorithms comparison. "PP=MO" stands for morphological operations applied as pos processing

Mask	Algorithm	Dice	Jaccard	BF score	Precision	Recall
HU	Clustering (k=115 PP=MO)	0.2364 ± 0.1490	0.1421 ± 0.0972	0.8223 ± 0.1510	0.7269 ± 0.2146	0.9977 ± 0.2364
	U-Net	0.0617 ± 0.0318	0.0321 ± 0.0171	0.7030 ± 0.0580	0.5559 ± 0.0693	0.9648 ± 0.0617
	Active Contours (U-Net)	0.2603 ± 0.2280	0.1696 ± 0.1558	0.9686 ± 0.0099	0.9450 ± 0.0191	0.9935 ± 0.2603
	Graph Based (U-Net)	0.2903 ± 0.0975	0.1737 ± 0.0690	0.9675 ± 0.0084	0.9523 ± 0.0134	0.9832 ± 0.2903
Anatomic	Clustering (k=46 PP=MO)	0.2215 ± 0.1960	0.1381 ± 0.1248	0.6247 ± 0.4533	0.7024 ± 0.3209	0.6387 ± 0.2215
	U-Net	**0.8048 ± 0.2249**	0.7186 ± 0.2482	0.9981 ± 0.0036	0.9979 ± 0.0049	0.9984 ± 0.8048
	Active Contours (U-Net)	0.3069 ± 0.3112	0.2253 ± 0.2459	0.9654 ± 0.0081	0.9778 ± 0.0171	0.9933 ± 0.0059
	Graph Based (manual masks)	0.1990 ± 0.0863	0.1131 ± 0.0548	0.9640 ± 0.0112	0.9435 ± 0.0215	0.9856 ± 0.1990

5 Conclusion

Different approaches were developed with the aim of helping specialists to detect the bladder more effectively and quicker. Two pre procedures were implemented to detect the RoI. Using GT information, it was possible to create a mask smaller than the one based on the HU values of the bladder. Consequently, the results were better with this mask. Next, four segmentation methods were tested. The approach whose results stands out is using the U-Net algorithm to predict the label of each volume, followed by the application of morphological operations and the selection of the biggest connect region.

In the future, it is intended to use hyperparameter optimisation techniques to choose the optimal parameters in each algorithm, to obtain better results. Since the rectum is a region of interest in the context of prostate cancer, it is planned to implement the previous methods, with the aim of studying whether their performance is similar when segmenting the rectum.

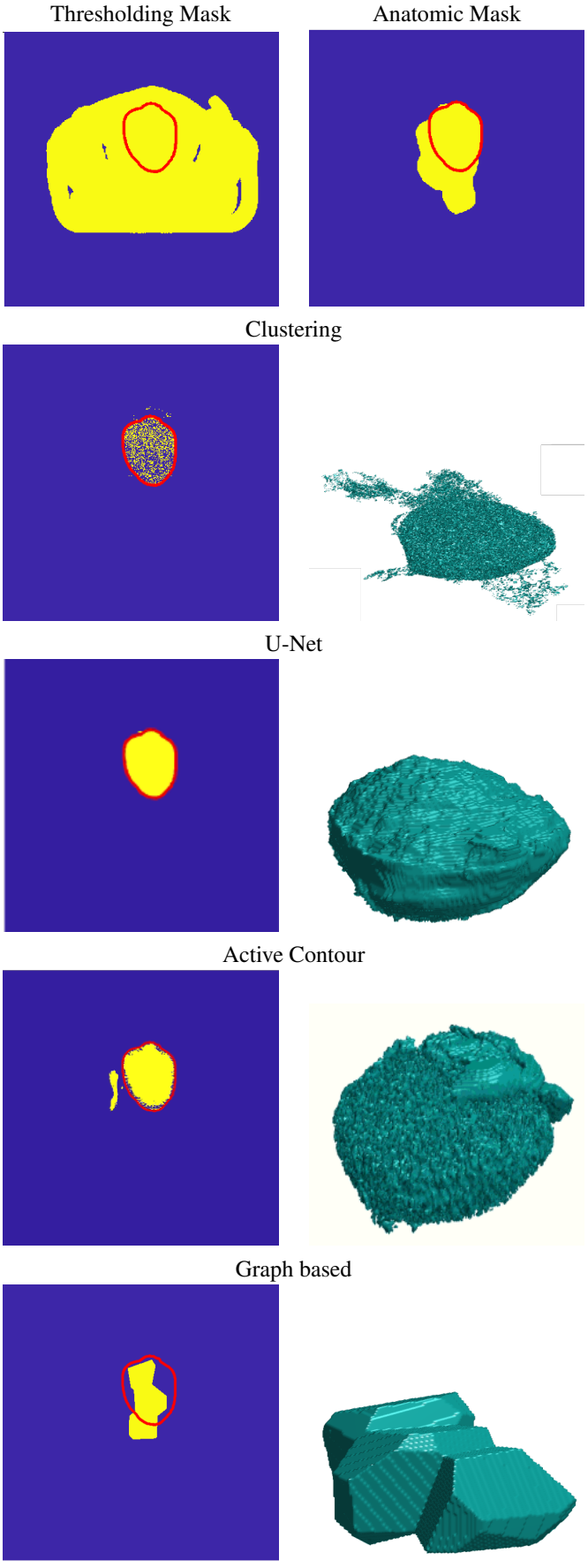


Figure 5: Selected segmentation results

References

[1] Neil G Burnet. Defining the tumour and target volumes for radiotherapy. *Cancer Imaging*, 4(2):153–161, 2004.

[2] Özgün Çiçek, Ahmed Abdulkadir, Soeren S. Lienkamp, Thomas Brox, and Olaf Ronneberger. 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation. In *Medical Image Computing and Computer-Assisted Intervention*, pages 424–432, 2016.

[3] Bing Li and Scott T Acton. Active Contour External Force Using Vector Field Convolution for Image Segmentation. *IEEE Transactions on Image Processing*, 16(8):2096–2106, 2007.