中图分类号: 学校代码: 10055
UDC: 密级: 公开

南開大學

Nankai University
硕士学位论文
Master's Dissertation

论文题目: 基于深度学习的前列腺磁共振图像的全自动分割
Title: Automatic segmentation of prostate in MRI based on deep learning

论文作者 MUHAYIMANA ODETTE Author MUHAYIMANA ODETTE

指导教师 马玲副教授 Supervisor A/Prof. Ling Ma

申请学位 工学硕士 Degree Master of Engineering

培养单位 软件学院 Institute College of Software

学科专业 软件工程 Discipline Software Engineering

研究方向 医学图像分割 Research Area Medical Image Segmentation

答辩委员会主席 谢茂强 Chairman of Committee Maoqiang Xie

评阅人 宫云战王靖 Reviewers Yunzhan Gong , Jing Wang

南开大学研究生院
Graduate School of Nankai University
二○二一年六月

The Authorization of Dissertation in Nankai University

I fully understand the management regulations of Nankai University on the retention and utilization of graduate student's dissertation and agree to submit my electronic form and corresponding paper version of my dissertation to Nankai University.

I understand that Nankai University has the right to utilize the dissertation within "The Law of Copyright" of the People's Republic of China, and I agree to authorize the university in the following aspects. That is:

1. The university will adopt the dissertation into"The Full-text Database of Nankai University's Dissertations ", and can provide the public dissertations to teachers and students as the reading in the library and free information services including catalog search, abstracts and the first 16-pages browsing on the campus network;

2. The university can adopt photocopying, micro-printing or other means to save the copy of dissertation. According to the provisions of the Ministry of Education, the university should submit public dissertations to the institution designated by the Ministry of Education;

3. This authorization is applicable to confidential dissertations after decryption.

I promise: the dissertation is completed by myself during studying in Nankai University and has been passed the defend; the electronic version of the dissertation is consistent with the contents of this paper; if these contents are different, I undertake my responsibility.

I have signed one copy of this authorization (that is one page of the dissertation), which shall be submitted to the library for retention.

Author's signature:MUHAYIMANA ODETTE_____ Date: 2021.05.31 _____

Title	Automatic segmentation of prostate in MRI based on deep learning				
Name	MUHAYIMANA ODETTE	Student ID	2120196018	The Date of Oral Defense	2021.05.29
The Type of Dissertation:	Master				
Institution/Department	College of Software	Major:	Software Engineering		
Email Address:	muhayiodette06@gmail.com				

Note: The authorization applies to all students' dissertations. The author fills in the duplicate authorization and hands in the library. The non-public dissertation should be with "The Approval Form of non-public dissertation in Nankai University".

The Original Declaration of Dissertation in Nankai University

I solemnly declare: The presented dissertation is my own original work under the guidance of my supervisor. All references have been stated and the dissertation does not contain any others works created, published or not published. Other individuals and the collectives that contribute to the dissertation have been stated in the dissertation. I will be responsible for all the legal of the statement.

Author's signature:_____ Date: _____

非公开学位论文标注说明

（本页表中填写内容须打印）

根据南开大学有关规定，非公开学位论文须经指导教师同意、作者本人申请和相关部门批准方能标注。**未经批准的均为公开学位论文，公开学位论文本说明为空白。**

论文题目			
申请密级	□限制(≤2 年)	□秘密(≤10 年)	□机密(≤20 年)
保密期限	20　年　月　日至 20　年　月　日		
审批表编号		批准日期	20　年　月　日

南开大学学位评定委员会办公室盖章(有效)

注：限制★2 年(可少于 2 年)；秘密★10 年(可少于 10 年)；机密★20 年(可少于 20 年)

摘 要

前列腺癌是世界范围内男性癌症死亡的主要原因之一，也是一个重大的公共卫生问题。如果前列腺癌能够在早期发现则很大程度上会被治愈。由于磁共振成像（MRI）能够提供详细的解剖结构，因此是前列腺癌诊断与治疗的最常用的模态之一。从 MRI 中准确地分割前列腺，对于前列腺癌的诊断和治疗计划至关重要。深度学习已在疾病的早期检测、图像处理和分析、图像分类、图像配准、图像分割和医疗计划等医学领域提供了重要的支持。本文提出了一种基于深度学习的 MRI 前列腺自动分割方法。采用卷积神经网络，3D U-Net，对 MRI 前列腺进行了分割。在包含 50 个 MRI 影像的公共数据集 Promise12 上进行了 10 折交叉验证实验，最终得到平均 84.92 %的 Dice 相似系数和 5.3 mm 的 Hausdorff 值，实验结果表明提出的方法能够得到满意的分割结果。

关键词: 深度学; 前列腺分割; 核磁共振图像

Abstract

Prostate cancer is one of the most causes of cancer death for men worldwide and is still a significant public health problem. However, the prostate cancer can be cured if it is detected at early stage. Due to its ability to produce detailed anatomical structure, Magnetic Resonance Imaging (MRI) is one of the most used modality for prostate cancer diagnosis and treatment. Accurate segmentation of the prostate from MRI is crucial for diagnosis and treatment planning of prostate cancer. Deep learning has provided an important support in early disease detection, image processing and analysis, especially in image classification, image registration, image segmentation and in medical treatment plan. In this thesis, we propose an automatic segmentation of prostate in MRI based on deep learning methods. A Convolution Neural Network special type named 3D U-Net is used to segment the prostate in MRI. We conduct the 10 fold cross-validation experiments on the public Promise12 Data set of 50 prostate images, and achieved a mean Dice similarity coefficient of 84.92 % and a mean Hausdorff distance of 5.3 mm. The experiments proved that the proposed algorithm performed with promising results.

Key Words: deep learning; prostate segmentation; magnetic resonance image

Table of Contents

TABLE OF CONTENTS

Chapter 1: Introduction

Healthcare sector is completely different from other sectors in the industry due to its sensitivity. It is attributed a high priority since people's life matters than anything else. Therefore, people expect a loft from the healthcare regardless of the price it may take.However, this sector is not yet at the top in achieving social needs although it consumes a huge Deep learning has recognized huge success in various real world applications such as early detection of diseases, natural image processing and analysis, image registration; and it is also providing a great support in medical practices, especially in medical image segmentation for the treatment planning decisions.

Prostate cancer is one of the top most diseases causing deaths among men that it was ranked the second to lung cancer as the most frequent cancer death leading disease in men worldwide[1]. In 2018, the prostate cancer in men counted 1,276,106 new cases and 358, 989 deaths i.e. 3.8% of all deaths caused by cancer in men all over the world. The GLOBCAN 2020 reported an estimate of 1,414,259 or 14.1% of new cases with 375,000 deaths of prostate cancer worldwide; which ranks this disease as the second most frequent cancer and the fifth leading cause of cancer death among men in 2020. The late detection of prostate cancer is at the base of the observed high mortality rate. As published by cancer.net, the studies have shown that 87% of men treated when their cancer is diagnosed at early, can have a life expectancy of five years.

Prostate cancer starts when the prostate gland healthy cells start growing out of control, forming a tumor in the prostate gland, which can be identified as benign or malignant tumor. The majority of prostate cancers grow slowly and hide to the prostate gland where they may not cause serious harm and may require minimal or even no treatment whereas other

types are aggressive and can attack nearby parts of the body.The prostate cancer or malignant tumor should not be confused with the Benign Prostatic Hypertrophy (BPH) which is a natural increase of the prostate gland volume that mainly occurs in men over 65 years old and rarely in men as young as 40 years old. The most common types of prostate cancers are adenocarcinoma which attack men to the rate of 95% and other types rarely attack them to the rate of 5%. Due to the remarkable difference in clinical behaviors of the prostate biological structure, it is necessary to clinically differentiate significant prostate cancer (cs PCa) tumors from the indolent cancer.

Generally, the PCa is diagnosed using trans-rectal ultrasound (TRUS) guided biopsy in the patients whose medical exam show the higher prostate serum antigen or abnormal digital rectal examination (DRE). However, TRUS commonly considered as a typical imaging modality for the prostate diagnosis samples, only captures a small portion of the gland. Therefore, there is a high possibility for the biopsy results to not catch the most aggressive tumor in the prostate gland as accurate as needed. This gives a good reason to count on using the more advanced prostate cancer imaging modalities, including MRI, CT, that play an important role in the accurate prostate cancer diagnosis [2].

Once the prostate cancer is diagnosed, the next step is to evaluate the status of the detected cancer by performing the test. The test aims at determining whether the cancer is confined to the prostate gland, or if it has spread outside to the other parts of the body. The image can be examined by an expert physician to evaluate and propose the most suitable treatment solutions.

The volume of the prostate is the key indicator of the prostate' s health because it reveals crucial information about the stage of the prostate cancer, the probable prognosis and gives a direction to urologists about possible treatments to assign to the prostate cancer holder. Moreover, the volume of the prostate provides experts with the useful information to avoid over-diagnosis and over-treatment of a slow tumor which could be dangerous to

the man' s life. The segmentation operation on the T2-weighted (T2W) prostate magnetic resonance images (MRIs) achieves good results for the accurate treatment planning and provides higher support to the automated prostate cancer diagnosis algorithms. It also provides crucial guidance in various existing and developing clinical applications. For instance, radiotherapy planning for prostate cancer treatment relies on proper delineation of the prostate in imaging modality which is mostly MRI[3].

Some developed countries in their recent practices and currently in developing countries, the prostate delineation operation is achieved by manual contouring of the prostate in slice by slice manner with the help of either sagittal, axial, coronal view or using the combination of various views. However, the manual contouring requires high expertise, it is time consuming and it is prone to error due to in-homogeneous structure of the prostate and its surrounding parts[4].

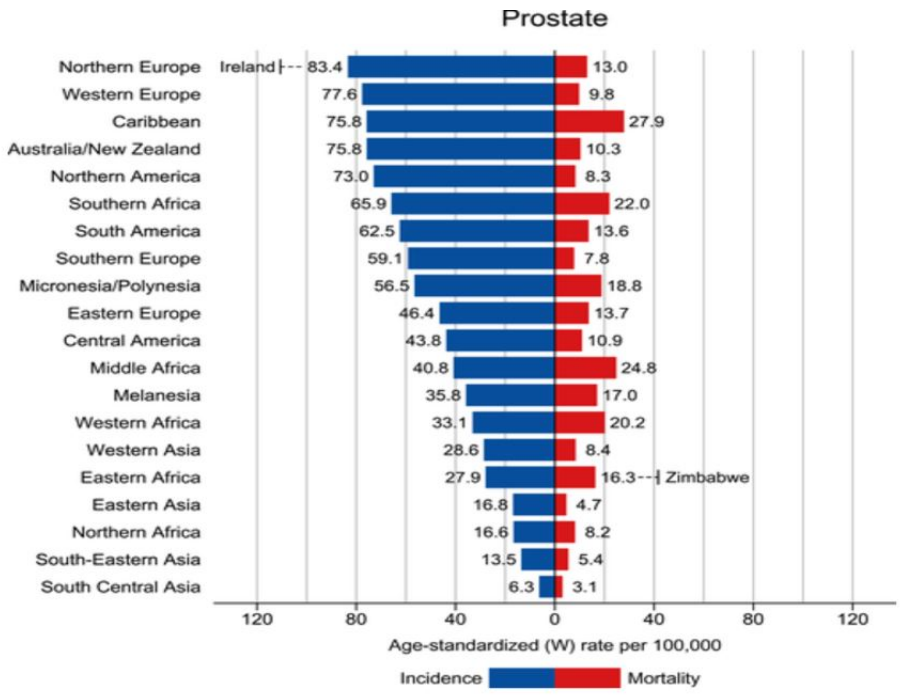
Developing an accurate and automatic segmentation system based on three dimensional MRI would be a good solution for identifying the malignant prostate tumor from the BPH and provide a better guidance for cancer diagnosis, localization, surgery operation planning, treatment plan, follow-up of the disease progress and eliminate inconsistencies or biases observed in manual segmentation. Note that making a choice of an automatic prostate segmentation method poses challenges depending on the imaging modality used and on the information used to guide the segmentation operation[4].

With the rapid advancement in technology, researchers are developing automatic and semi-automatic medical systems to assist the radiologists in prostate cancer diagnosis, in the disease prognosis and treatment plan. For instance, Zhu et al.[5] proposed a deeply-supervised CNN for accurate prostate segmentation from MRI based on a convolution information. A novel prostate cancer segmentation graphical user interface (GUI) tool has been introduced to support an easy implementation of several segmentation algorithms[6]. The strengths of this proposed GUI tool reside in that it can hold several segmentation

algorithms, such as a modified version of the snake-based active contour segmentation (ACS) algorithm referred to as Basic ACS. Then, the basic ACS model was extended in order to integrate a Gradient Vector Flow (GVF) component to improve the segmentation process[6]. And these segmentation algorithms outperformed the manual segmentation operations performed by medical experts. Compared with the semi-automatic segmentation methods which had the disadvantages of vast cost, inefficiency and subjective deviation, the automatic prostate segmentation algorithms may be robust and fast. Hence, in this thesis work, a special type of CNN architecture, a 3DU-Net architecture is proposed for the automatic segmentation of the prostate in MRI.

1.1 Research motivation

Cancer is a challenging public health problem in Africa and all over the world in the 21rst Century[1]. Over the past decades, most cases were mostly observed in countries with low- and middle-income, living with varying lifestyle, geographic and environmental factors. The GLOBOCAN 2020 has reported that all over the world, the prostate cancer ranked as the second most frequent cancer and the fifth leading cause of cancer death in men in the year 2020. The report estimated around 1.4 million new cases and 375,000 deaths caused prostate cancer worldwide. The incidence rates are estimated to 37.5 and 11.3 per 100,000 in transitioned and transitioning countries respectively, while the mortality rates does not vary a lot: 8.1 and 5.6 per 100,000 in transitioned and transitioning countries respectively. The prostate cancer has been reported as the most frequently diagnosed cancer among men in more than one-half, i.e. 112 of 185 of the countries worldwide. Moreover, prostate cancer became the leading cause of cancer death among men in 48 countries, especially in Sub-Saharan Africa, Caribbean, Central and South America and in Sweden. Back in 2018, the prostate cancer was the most diagnosed among men in sub-Saharan Africa, with an estimate of 69,000 cases (23% of all cases). Considering the incidence and mortality rates, the prostate cancer also ranks the first common diagnosed cancer in the same region. Rwanda, officially The Republic of Rwanda, is a landlocked sub-Saharan country located



Incidence and mortality rates of the most common cancers in sub-Saharan Africa,2018. Source: canceratlas.cancer.org] 图

Figure 1.1: Region-specific Incidence and Mortality Age-standardized Rates for Prostate Cancer 2020. Source: GLOBOCAN 2020

in East Africa, precisely in the Great Lift Valley. It is populated with over 12.6 million on the 26.338 km2 of land. Prostate cancer is a leading cause of cancer deaths in Rwanda whereby based on the GLOBOCAN 2020,the International Agency for Research on Cancer (IARC) has reported the estimates of 1054 or 28.6% new cases of prostate cancer among the 3,683 overall cancer new cases. Prostate cancer was also reported by GLOBOCAN 2018 as one of the leading cause of morbidity and mortality among men in Rwanda, where there was an estimate of 31.9% per 100,000 populations and 24.4% per 100,000 populations as incidence and mortality rates respectively[7]. In 2014, the World Health Organization (WHO) reported 16.3% of prostate cancer death among 2700 cancer deaths that were observed as all cancer deaths in Rwanda. According to the Rwanda Ministry of Health' s report, the primary risk factors to the high rising of prostate cancer are multi-factorial including lack of early screening practices and poor prostate cancer awareness among men.

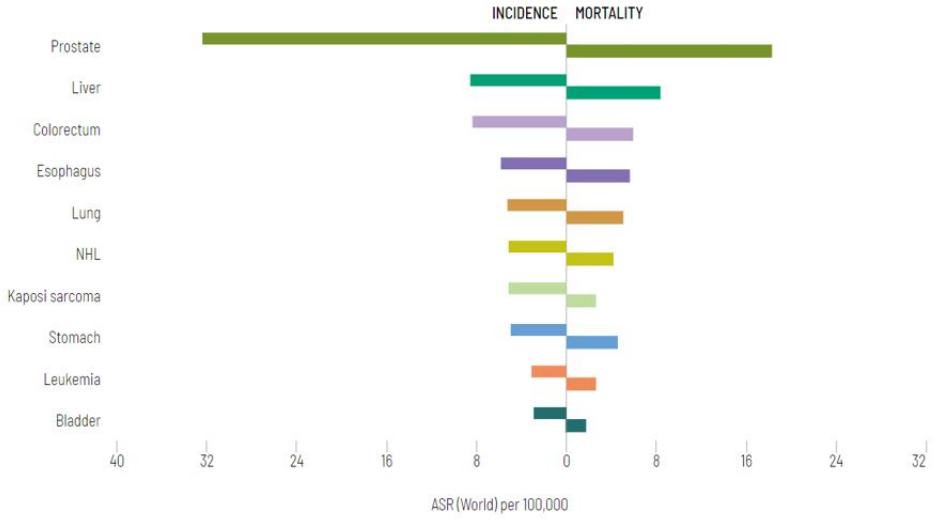


Figure 1.2: Incidence and mortality rates of the most common cancers in sub-Saharan Africa, 2018. Source: canceratlas.cancer.org

However, the country is always putting much efforts in improving the long term preventive measures. For instance, not long ago, the cancer unit at Rwanda Biomedical Center (RBC) which is under the Ministry of Health inaugurated the country' s 5-year National Cancer Control Plan. In the past decades, there was only one public hospital, Butaro Cancer Center that could offer chemotherapy and surgery services to patients with cancer but recently, Rwanda Cancer Center has launched two linear accelerator radiotherapy machines; which reduced the number of transfer for treatment abroad. In other words, the Rwanda Cancer registry system has improved. The new Information and communication Technology infrastructures are also improving rapidly that the country can accommodate advanced medical technology[8]. It is worth noting that although Rwanda has done a lot to improve the health care sector, there is still a lot of work to be done among others being, building capacity in prostate cancer awareness among citizens in order to shift towards earlier diagnosis to facilitate early treatment plan and reduce cancer death rates[8]. Hence an automatic segmentation system would provide an efficient way to diagnose cancer diseases at early stage and facilitate the early treatment planning such as giving a clear direction on surgery operations on patients. This would save the life of citizens at early stage of the disease and

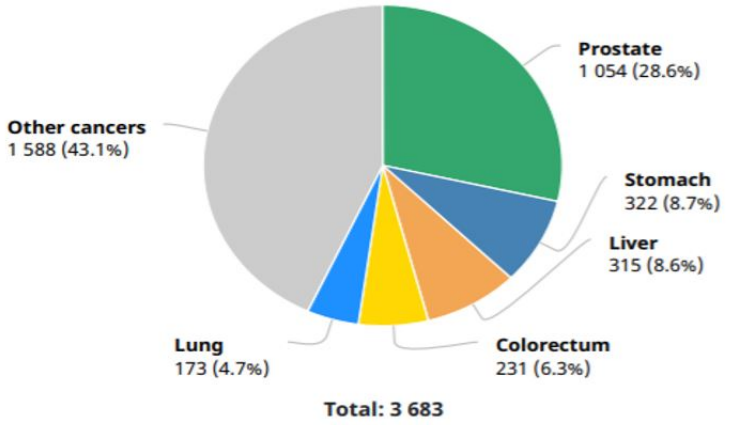


Figure 1.3: Rwanda: Number of new cancer cases in 2020, males, all ages. Source: GLOBCAN 2020

increase the life expectancy.

1.2 Problem definition

According to Prostate Cancer Foundation, prostate cancer is a type of cancer found in male reproductive system. The prostate cancer starts when normal prostate cells become abnormal and start multiplying out of control. Some types of prostate cancer grow slowly and do not spread out of the prostate gland while other category behaves aggressively and metastasize to other parts surrounding the prostate gland and to other organs of the body. The majority of prostate cancers are the slow-growing ones and take years and years to be detectable. For this category, monitoring is recommended for the purpose to verify if it is a malignant type that requires treatment or if it is just a normal growing of prostate commonly named Benign Prostatic Hyperplasia (BPH) that goes with aging. The aggressive category requires serious treatment as it rapidly attacks neighboring organs and can cause severe pain, erectile dysfunction and urinary incontinence to patients. After diagnosing the patient with prostate cancer, the custom-design treatment path can start from a wait-and-watch approach to higher medical and surgical plan for a better treatment.

The prostate anatomical structure is complicated due to its unclear boundary specifically at the apex which is located at the bottom of the prostate gland and at the base which is situated just at its top. Moreover, the shape and texture of the prostate vary from one patient to another and this heterogeneous structure makes the prostate surgery operation a challenge when it is the only treatment to assign to the patients[5]. However, carrying a manual segmentation operation in order to decide on prostate parts that undergo the surgery operation requires medical experts and physicians. This is time consuming and is prone to inter and intra-operative errors[9]. An automatic segmentation method would provide health centers with a suitable way to save time and use internal expertise instead of recruiting from outside, as it can be trained to internal medical staff without requiring external interventions.

In this research study, an automatic prostate segmentation from MR images is developed in order to provide medical institutions with an efficient way to conduct early prostate cancer prognosis, staging and make an early treatment plan, and once the decision about treatment is taken, complete information about the treatment should be provided as well[6]. This work makes sure to evaluate if the proposed method is going to outperform the manual existing segmentation method. The development of the automatic prostate segmentation method from MRI is going to be beneficial to both medical staff and patients as it will ease the work for medical doctors by using the system instead of doctors spending days and days on patients and also provides the patients' reports in a short time. This will facilitate the early prostate prognosis and treatment plan according to the output results from the system. Also, patients will no longer queue at the hospital and will be able to access their test results in a short period instead of waiting for months and months, which could even lead to their death before treatment. To develop the automatic prostate segmentation method, we build a special convolution network architecture called U-Net. This U-Net architecture is built upon the rearrangement of deep learning tools such as convolution layers and max pooling layers.

In fact, the main purpose of this study is to show the effectiveness of proposed automatic

segmentation of prostate in MRI compared with the manual segmentation and the other automatic segmentation methods. We also aim to create a U-Net architecture that is capable to extract features from the input image that help to capture its context then provide a segmented image which is the same as the input image but with a correct localization of the information that can guide medical doctors monitoring the prostate cancer prognosis and staging, treatment plan and surgery operation. The advantage of this proposed segmentation method is that it can avoid subjective judgments and time-consuming process.

1.3 Research objectives

1.3.1 General objectives

The general objective of this research is to perform an automatic prostate segmentation task on a relatively small data set using deep learning tools.

1.3.2 Specific objectives

Below specific objectives will help us achieve the general objective of the study:

- To read and review the literature to identify gaps, suitable methods and approaches used to perform prostate segmentation tasks
- To point out various techniques and methods for designing automatic prostate segmentation system from MRI
- Building the appropriate deep learning method for automatic prostate segmentation that achieves promising results with outperforming accuracy
- Evaluating the performance of the proposed method on the Promise12 data set
- Comparing the developed segmentation method with other segmentation methods in order to get future recommendations

1.3.3 **Significance of the study**

Convolution Neural Network is a class of Artificial Neural Networks that is attracting the attention of researchers in various domains including radiology and now it is dominating in computer vision tasks among others being disease diagnosis, tumor recognition, organ segmentation and so on [10]. U-Net is a special CNN architecture that has proved to outperform in biomedical image segmentation. Moreover, when the segmentation task is carried out on relatively small data set, U-Net architecture makes use of data augmentation approach on available annotated samples to provide outperforming results [11]. This work will create a U-Net architecture by varying parameter and choose the best parameters for the prostate segmentation on the small data sets. In addition, the successful implementation of the proposed segmentation methods will help radiologists to save time and patients' life at the same time by providing an efficient and rapid way to examine the patients' cases and provide results within a minimum period of time.

1.4 **Thesis outline**

The remaining part of the thesis work is outlined as follows: Chapter 2 covers the background of the topics that relate to the proposed method, including segmentation methods on the general images, segmentation methods on medical images, and prostate segmentation methods. Chapter 3 explains the MRI, CNN and U-Net, which are used to develop the proposed automatic prostate segmentation algorithm. Chapter 4 will present experimental work details comprising the description of the data set used in the method development, the experimental design, experimental parameters, evaluation metrics that are used to assess the proposed method performance, experimental results and results discussion. Chapter 5 gives a conclusion by summarizing the whole thesis work with the methodology, findings, and evaluation results, then gives recommendations about the future work plan for the later improvement of the proposed algorithm.

Chapter 2: Literature Review

If the prostate cancer is diagnosed at early stage, then it can easily and successfully be treated with adequate measures. Therefore, the accurate detection of the prostate cancer is very important for the sake of saving people' s life and reducing cancer deaths that are increasing daily all over the world. Nowadays, MRI is considered as the most common imaging method for PCa diagnosis since it has the ability to produce high soft tissue resolution that facilitate the internal and external view of the prostate tissues. As the MRI has shown promising PCa diagnosis results, the segmentation of prostate from MRI can provide radiologists with accurate information about the cancer prognosis, treatment plan and surgery guidance. In the past decades, the prostate MR images were segmented manually, which was a tedious and time consuming task, prone to intra-operators' errors and requiring higher expertise and attentive observation. To alleviate this problem, many researchers invested a lot of efforts in creating ML and deep learning methods to perform semi-automatic and automatic prostate segmentation tasks. Traditional ML methods including SVM and RF achieved good performance in image segmentation, but later on the emerging of huge amounts of data with a lot of parameters, ML methods could not handle the feature engineering and the model generalization ability in an appropriate manner. Thus, deep learning emerged to support higher dimensional data and alleviate the model generalization ability problems that were observed in ML models [12].

2.1 General image segmentation methods

Accurate segmentation is an important step for successful digital image processing operations. Segmentation consists of classifying or clustering an image into a set of visually distinct sub-regions to create a homogeneous region based on a certain criterion. This cri-

terion can be brightness, contrast, color, texture or gray level. The resulting homogeneous region provides a clear way to analyze and extract the knowledge from the image data. This knowledge gives an important impact on the overall success of that image understanding. In other words, the key motive of image segmentation process is to get a lot of information from the region of interest in an image. This provides a detailed explanation about the object scene.

Generally, segmentation methods are classified as either supervised or unsupervised paradigm. The unsupervised segmentation prioritizes the image structure whereby, various techniques depend on the pixel intensity and gradient analysis to delineate the region of interest in an image. These techniques may be region based, edge detection based, graph cut based or deformation based. Applying these techniques produces promising results on well-defined boundaries. However, the performance degrades vis a vis complicated shapes. For instance, the gradient based segmentation methods are sensitive to noise and artifacts which may result in missing tissue boundaries. In addition, the graph-based methods that solves the latter issues are often computationally expensive thanks to the iterative scheme used to improve the segmentation results.

The supervised segmentation methods include prior knowledge about the image processing task via the training samples. For instance, atlas-based segmentation methods, notably statistical shape models and probabilistic atlases can perform well by capturing the organ’ s shape with promising results compared to unsupervised models [13]. Another class of supervised segmentation methods includes Support Vector Machine (SVM) and Random Forest (RF) that also perform well in image segmentation tasks. Nevertheless, in regard to fuzzy radio-logical image delineation, their performance is limited. Supervised machine learning frameworks also have shown a huge advancement in segmentation tasks, especially Convolution Neural Network (CNN) performed well in semantic segmentation of natural images with promising results. The research done by Litjens in [1] revealed that deep learning achieved the state-of-art performance in solving various complex tasks

including image classification, object detection, disease prognosis and image segmentation.

To carry out the image segmentation, the segmentation methods choice depends on the selected parameter for segmenting an image including pixel intensity, homogeneity, discontinuity, clustering or topology [4]. It may also depend on the modality together with the acquisition protocols used for image data collection or on the problem to be solved. However, to date, there is no universally recognized image segmentation method that can perform perfectly since image segmentation results are most of the time affected by the variability of intensity homogeneity in an image, partial volume effects, artifacts, etc. and every imaging system has its own limitations.

2.1.1 Traditional ML methods for image segmentation

- **Region-based Segmentation**

  Intensity based segmentation commonly called threshold based approach, it consists of segmenting an image based on intensity values. To segment an image, this approach chooses a threshold value and classifies the image into two classes referring on a hypothesis that pixels belonging to a certain range of intensity values are part of one class and the rest of the pixels represent another class. The threshold based segmentation can be either global or local.

  The global threshold separates the object and the background in the image by making a comparison of pixel values with the chosen threshold value then use the binary partition to segment an image whereby, the pixels whose values are greater than the threshold value are assigned a binary value 1 while pixels with values lesser than the threshold value are assigned a value 0 as shown below.

$$f(a,b) = \begin{cases} 1, & \text{for } i(a,b) \geq t \\ 0, & \text{for } i(a,b) < t \end{cases} \tag{2.1}$$

  Where $f(a,b)$ is the output image and $i(a,b)$ is the input image.

Note that the global thresholding uses only one threshold value for the entire image segmentation. Local thresholding also known as adaptive thresholding divides an image into sub-regions and the chosen thrshold value varies over the entire image depending on the considered properties of local pixels in each sub-region until all sub-regions are processed. Otsu method is the most common threshold segmentation algorithm used in the selection of an optimal threshold by maximizing the variance between classes [14].The intensity based segmentation methods are advantageous that they are computationally cheap, fast, simpler to implement and are applicable to the real-time applications. However, they ignore the spatial information of the image which make them highly sensitive to noise and can miss some edges [15]. The correct threshold value selection is difficult and often results in over or under-segmentation [14].

Region growing segmentation aims at extracting a region from the image by comparing the candidate pixel to neighboring pixels then check the homogeneity criteria for classifying pixels into the appropriate region of an image. Note that the regional growing method is classified as seeded or unseeded growing method depending on whether it initiates the seed manually or automatically before segmenting an image.

The seeded region growing method or semi-automatic region growing method is the fastest and robustest of its counterparts. It starts by initiating a seed pixel in the image to be segmented then examines the neighboring pixels to the seed pixel that are added to increase the region accordingly.

The unseeded region growing is a fully automatic and flexible segmentation approach that bases on the pixel similarities within regions of the image to be segmented. This approach is applicable to multi-spectral and 3D images and does not count on neither parameters tuning nor manual inputs for the segmentation process. The inclusion of adaptive filters in unseeded region growing segmentation made it outperforming.

Region merging is another region based method that starts by initializing some seed points and the seed value should be chosen correctly since the segmentation results

will depend on its value. Then the regions grow iteratively by merging neighboring pixels according to the merging criterion. This process continues repeatedly until all pixels belong to their corresponding regions as per merging criterion.

Region splitting operate in an opposite way to the region merging method whereby, the split criterion is defined and the image is split repeatedly into homogeneous regions based on a split criterion until there is no possibility to split.

Region split and merge method is a segmentation approach that follows a quadratree data structure to segment an image. It combines the advantage of both region merging and region splitting methods to perform of segmentation. This approach considers the entire image as one region then splits it into four quadrants based on the predefined homogeneity criterion. The quadrants are evaluated on the same criterion and are merged depending on the uniformity of the segment. The process of splitting and merging regions continue until the criterion is satisfied and there is no possibility of splitting or merging again. The below algorithm is followed during the split and merge method.

Algorithm

- Set the homogeneity criterion. Split the image into four quadrants
- Check the resultant squares homogeneity
- If any square is heterogeneous, split it into four quadrants
- Merge two or more neighboring regions that meet the homogeneity criterion
- Repeat the step that precedes the last step until there is no further split and merge possibility

- **Edge detection based segmentation**

  This segmentation approach bases on the principle of intensity variations among the pixels to detect the boundaries or edges of the objects within an image. This approach aims at segmenting an image into regions based on marking the discontinuities be-

tween objects in an image. These discontinuities may be related to gray level, color, etc. The approach creates borders in the image to be segmented by combining detected edges into edge chain, fake edges and weak edges. The application of a variety of edge detecting operators base on the gradient function to facilitate the segmentation. We can mention the first derivative based detecting operators such as Prewitt, Sobel, and Robert' s and the Laplacian which bases on the second derivative.

Algorithm

- Apply the edge detecting operator to detect the edges in the image
- Estimate the strength of edges by measuring the amplitude of the gradient
- Keep all edges having a magnitude greater than the threshold value and drop weak edges
- Find the position of the crack edge. Keep or abandon the crack edge based on the confidence it receives from its predecessor or successor edge
- Repeat the two steps before the last one, using different values of threshold to find out the the relating boundaries that achieve the segmentation process.

The edge detection based segmentation' s weakness lies in that, fake and weak edges impact segmentation results negatively and it is sensitive to noise. In addition, edge detection based segmentation method can not achieve the segmentation process on its standalone. Therefore, it needs to be combined with region-based techniques for better segmentation results.

- **Atlas-based segmentation** An atlas is defined as a description of the relationships between anatomical structures of an image. Mathematically, atlas A is defined as a mapping function from the n-dimensional to the labels. Atlas-based segmentation is another type of image segmentation method other than machine learning method and has also achieved good performance. The segmentation techniques base on different types of image information to segment an image. The Atlas-based segmentation considers the locations and shapes of image anatomical structures, and the spatial

relationships between them for the segmentation process. Having an atlas, for the segmentation success, the registration process must be performed on the new image beforehand. It consists of mapping the new image coordinates space to the atlas' s coordinate in the correct anatomical way [16], i.e. every coordinate in the new image has its corresponding coordinate in another image. This correspondence is designated as a coordinate transformation T that maps the new image coordinates to those of the mapping. Atlas-based segmentation offers an advantage of segmenting images whose relation between regions and pixels' intensities is not clearly known. In addition, it is suitable in clinical services especially for computer-aided diagnosis where it is used to estimate the shape of an object or examine the morphological differences between patient groups. However, atlas-based segmentation tasks become time consuming when iterative procedures are included in the atlas construction. Another challenge of atlas-based segmentation is the objective validation: given that this segmentation approach is used to segment images whose gray level intensities information is not sufficient; it becomes hard for it to validate the objective [17].

- **Deformable models based segmentation** Deformable models are another category of image segmentation method that have recognized a lot of applications in medical vision, especially in motion tracking. This segmentation approach can be used as either of the two classes of which (1) the parametric deformable models also called active contours and (2) the geometric or implicit models. Deformable models are characterized as curves or surfaces used for segmentation of the images with smooth region boundaries, for the segmentation of higher dimensional images, among others being stacks of images which experience the deformation under the influence of internal and external forces to delineate object boundary. The internal forces preserve the shape smoothness whereas the external forces rely on the image features to guide the model towards achieving the target region boundaries.

  The parametric models can be used for two-dimensional or multi-dimensional image segmentation. When used for two-dimensional image segmentation, they are referred

to as curves and their deformations are determined by the displacement of discrete number of control points along the curve while they are referred to as hyper-surface in higher-dimensional image segmentation. These parametric models segmentation methods offer the advantageous of converging very fast depending on the number of control points. However, their downsides lie in that they depend on the topology that they can only capture a single region of interest (ROI). Consequently, if images present multiple ROI, there should be multiple model initialization equal to the number of ROIs.Geometric models in contrast use the distance transformations to determine the shape on the image and proved to perform better than parametric models because they have the ability to capture more than one ROIs using a single model.

- **Clustering based segmentation** This is another approach that consists of segmenting an image by grouping sub-regions with similar pixels into a specific group that is different from other groups. The K-Means is the most frequently used clustering algorithm that aims at clustering the points near the centroid (the average of all points in that cluster) and has the coordinate as the mean divided by all points in the cluster. These algorithms minimize the distance between the data points and the centroids when grouping the data points. However, their drawback lies in that the number of clusters need to be defined before running the algorithm, which can affect the results. The K-mean clustering follows the below algorithm:
  - Choose the number of clusters k. Randomly put the k cluster centers at different initial locations in the image.
  - Assign each data point to the closest centroid
  - Recompute the cluster centroids, and the centroid should be at the average coordinate of the data points that make the cluster
  - Go back to step to until the distance between the centroids and data points does not change.

- **Support Vector Machine (SVM)**

  SVM also called Kernel Support Vector Machine, is a great supervised and nonlinear machine learning method that plays an important in classification tasks whereby it seeks to find the optimal hyper-plane that separates new data correctly into any number of classes and maximize the margin between data points of both classes. The kernel support vector machine is a nonlinear classifier that builds upon the pre-established filter. This makes the SVM suitable for medical imaging applications with small training data set. Moreover, this kernel SVM involves the hyper-parameter tuning of the SVM classifier only, which makes the SVM method computationally less expensive and efficient.

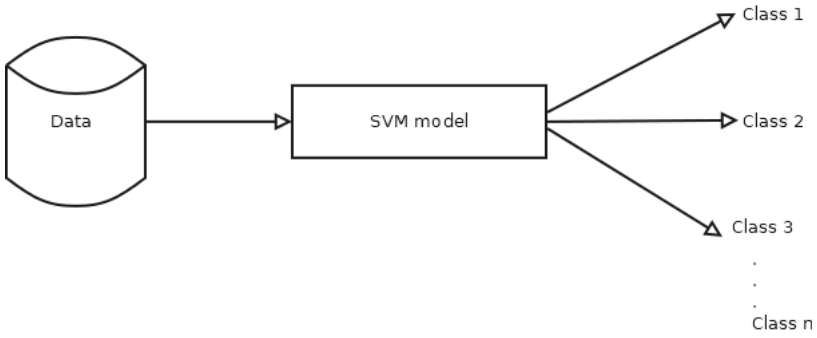


Figure 2.1: SVM classifier

- **Random Forest (RF)**

  Random forests are popular supervised machine learning classifiers, trained to delineate the boundaries between points in a feature space belonging to different classes with a purpose to maximize the margin between classes as done in SVM. The random forest model is built from an ensemble of decision trees, trained to take input values or feature vector and attribute the membership value to pixels (voxels in 3D cases) to belong to either the foreground or background class. The random decision tree training bases on the determined decision rules in each node of the tree.

  Practically, a threshold value is set then the input feature value is tested against it. If the input feature value passes the threshold value, it is passed to the right node

and passed to the left side otherwise. Note that the threshold value is selected from a bunch of tests by choosing the one which maximizes the information gain. The overall decision results from the information flow from the tree root node to its children until reaching the leaf node. Then the possible output in all leafs are averaged to make the final decision of the random forest model. Another way of getting the final result from the model is to use the majority vote method. In other words, for the purpose to prevent the model from over-fitting, the random forest creates a bunch of random uncorrelated decision trees to get the optimal result. These models are referred to as Bootstrap Aggregation or bagging and play a crucial role in overcoming bias-variance trade-off problems. Generally, the model generalization ability can be evaluated based on the bias and variance in a context that the higher bias implies inaccurate test results and the higher variance indicates an over-fitting model, i.e. a model that performs well on the training data and gives poor results on new data.
Typical example of averaging the predictions from all individual trees: Let $X = x_1, x_2, , x_n$ be a training data-set Let $Y = y_1, y_2, , y_n$ be its corresponding labeling set The bagging approach performs a repeated and random sampling from the training data set in m times. The approach replaces the original training data set by fitting the binary trees to these sampled data. Let X_m and Y_m be the sampled data set, where $m = 1, , M$ Let Tb represents the binary tree trained with respect to X_m and Y_m. After the whole training process, the predicted value from the test data set x_{test} can be given by the formula:

$$\hat{y} = \frac{1}{M} \sum T_b(x_{\_test}) \tag{2.2}$$

The random Forest algorithms offer great advantages of reducing the correlation among the trees in the bagging sample through selecting a random subset of features from the candidate splitting. In addition to that, the random forest requires few hyper-parameter tuning whereby only three parameters i.e. the number of trees, the number of features in a tree and the sampling rate for the bagging, which makes it simple to implement and with promising results. However, it is hard to understand

the process that is carried between the input until reaching the output of the model.

- **Markov Random Field (MRF)**

  MRF is a conditional probability model widely used in medical image segmentation, especially in MR images because it has the properties that make it suitable to model inhomogeneities that often occur in MR images. It is characterized as a stochastic process that makes use of local features of the image whereby, the probability of a pixel is affected by its neighboring pixels [18]. Given that MRF represents undirected graph, it can be defined as G (V, E) where G represents the graph, V represents the vertex, and E represents the edge. With the same analogy to image, G represents the image, V represents the pixel and the edge E represents the distance between two neighboring pixels. By applying MRF for classification of pixels in the image, for each object B in image, each pixel represented by v is given a binary umber 1 if it belongs to the object and it is assigned the binary 0 otherwise. Then the voxel is labeled based on its similarity to the object B and the similarity of neighboring pixels to the same object in the image [18]. It is worth noting that although MRF achieves the image segmentation tasks with good accuracy, it has two major downsides that it is computationally complex and it results are sensitive to the model parameters.

2.1.2 Deep learning methods for image segmentation

Classical machine learning based methods have achieved good results but resulting systems were complex that their deployment was limited to available hardware. Since 2000s with the advancement of hardware and the increase of data, deep learning approaches started achieving the state-of-the-art performance in the real world application such as natural image processing, image classification, disease detection, and currently, it is also providing outstanding support in medical fields for instance, in image segmentation tasks, etc. [19].

Convolution Neural Network (CNN) is a class of neural networks that comprises a lot of layers, each performing a specific task such as convolution, pooling, loss calculation, etc. Below image segmentation methods are built up on the CNN network

to create segmentation models. **2D CNN**

CNN has proved its ability to perform image classification and pattern recognition that many researchers found it as of a good choice for the medical image segmentation as well. The 2D CNN concept consists of segmenting 2D images by applying the 2D filters on the input image. *The research conducted by Zhang* in [20] has shown that performing image segmentation by passing multi-model images to the input layers achieve good accuracy than using a single 2D modality input image.

2.5D CNN

The 2.5D CNN models offer an advantage of applying 2D filters to capture richer spatial information of neighboring pixels, and this process showed to be less computational expensive compared to 3D CNN. The 2.5D approaches are also used with labeled data which are easier to be found than finding 3D images. Moreover, many hardware easily adapt to the 2.5D CNN architecture than 3D that requires the advanced hardware. Splitting the volumetric images into 2D representation help avoid dimensionality problems, which make them preferable. However, 2.5D CNN are just limited to 2D kernels that they cannot apply to 3D image segmentation tasks.

3D CNN

3D CNN models are suitable in extracting the powerful volumetric representation over the axes x, y and z representing 3D images. The 3D network aims at predicting the label of a central voxel depending on the content of the 3D patches surrounding this central voxel. The 3D image segmentation approaches have been motivated by the increase in 3D medical imaging and the advanced hardware architecture. The 3D CNN makes the full use of 3D spatial information from volumetric images to give rich and clear information in all directions whereas 2D CNN approaches provide one view of the information, and the 2.5D CNN approaches provide just three orthogonal views of the information [19].

Fully convolution Network (FCN)

The FCN has been developed by Long [21]. This network differs from the usual CNN in that it replaces the last fully connected layer by a fully convolution layer,

which makes this type of CNN architecture to get the dense pixel-wise prediction. For a better information localization, the model concatenates the high resolution convolution results with the results given by up-sampling process, then this combination is passed to the convolution layer to create accurate output. The FCN has showed the improvement over the usual CNN by creating the pixel-wise predictions from a full-sized image instead of making patch-wise predictions and moreover, it does that in just one forward pass process [13]. FCN is used to segment multiple organs of different sizes. Zhou [22] used FCN in 2.5D multiple organ segmentation models on the 3D CT images and yielded good results on the larger organs (Dice = 93.7%) and poor results on small organs (Dice = 55.3%).

Cascaded FCN (CFCN) The cascaded FCN is a type of network architecture that combines a series of FCNs that each model makes use of the contextual features extracted by the previous model. This type of network has an advantage of increasing the model performance by applying separate filters at each stage of the model training process. **Focal FCN (FFCN)**

The focal FCN provides a model with the ability to reduce the number of false positives that occur due to the unbalanced ratio of the background and foreground pixels in a medical images [22]. **Multi-stream FCN**

In the case of images varying in the modality and in resolution, i.e. images scaled differently; a multi-stream approach is a good choice for this case to facilitate the system to take advantage of multiple forms of an organ constituents. Zeng [23] has used the same approach to maximize the contextual information use from multi-scaled images. He has also applied the multi-modality techniques which improved the model robustness vis-a-vis a wide variety of shape and structure in an organ. The FCN normally has receptive of fixed size which make it difficult to track changes once the object size in an image changes, which affects the model performance. The multi-scale network is applied to alleviate this problem. However, sharing the parameters of the same network on resized images does not provide better results since multi-scaled images require to process different parameters. Hamidian [24] proposed applying the

FCN in a sliding window manner to deal with the fixed-sized receptive fields where the image size is bigger compared to the field of the window.

U-Net

U-Net is one of the most well known architectures for medical image segmentation and has been initiated for its first time by Renneberger et al. [11] using the deconvolution concept introduced by Zeiler et al.[25]. U-Net is defined as a special type of CNN architecture, with the ability to perform a stable training with few samples by combining the pooling operations and up-sampling operations. A U-Net model is built upon the FCN architecture and comprises two important parts, the left part called analysis path or contracting path and the synthesis also named expansive path. It also includes the horizontal connection between the contracting path and expansive path in order to keep the contextual information and localization information of the image under segmentation process. The contacting path follows the structure of the standard CNN while the expansive path includes the the up-sampling layer followed by the deconvolutional layer [19]. The special property of U-Net is its ability to create shortcut connections between layers of equal resolution in the contracting path and the expansive path and these connections offer an advantage for the network to provide essential high resolution features to the deconvolution layer which up-samples the image to the same resolution as the input image. The 3D U-Net is an advanced version of 2D U-Net that has the same implementation as the 2D U-Net by replacing 2D operations by 3D operations. 3D U-Net captures enough spatial information from the input image and applies 3D kernels instead of applying 2D ones.

V-Net

V-Net is a variant of U-Net architecture, proposed by Milleteri in his research study [26] for the purpose of 3D image processing tasks. V-Net has a similar structure as U-Net with few changes. It is built upon the left part called compression path and the right part called decompression path. The compression path of network is divided into various stages, each having one or three convolution layers that operate on the resolution of input image. The input image of each stage is processed by the convo-

lution layers applying an appropriate kernel size followed by the non-linear function, then its results is added to the last convolution layer of the same stage in order to facilitate learning the residual function. As the compression path goes deeper in the network, the image resolution decreases through the convolution operation with 2x2x2 voxels wide kernels and a stride 2. Therefore, the resulting feature maps is halved as it would be under the pooling layers, and the number of feature channels are doubled at each stage. The network replaces the pooling layers with convolution layers for the purpose to use a small memory during the model training as there will not be any need for switches to map the output of pooling layers back to their inputs, i.e. there will not be any need for back-propagation process that consumes the memory [26].

The decompressing path consists of extracting features and expanding the spatial support of the lower resolution feature maps in order to gather and combine the necessary information to provide a two channel volumetric segmentation as output. Each stage of this decompressing path performs a deconvolution operation to increase the size of the inputs, followed by one to three convolution layers halving the number of kernels used in the previous layer. The residual function is also learned in the decompression as in the compression path.

The very last convolution layer with a 1x1x1 kernels computes the two feature maps to produce the outputs of the same size as the input volume. A soft-max voxel-wise function applies to the output feature maps to get corresponding probabilistic segmentation of the foreground and background regions. Note that during the compression path operations, the localization information is lost. Therefore, the feature extracted from early stages of the compression path must be transferred to the decompressing path though the concatenation operation in order to provide the localization information to the right part of the network and improve the quality of the final contour prediction [19]. The V-Net has an advantage of increasing the receptive field though down-sampling and speeding up the model convergence time.

Convolution Residual Networks (CRNs) In the research conducted by He in [27],

he noticed that deeper networks are capable to extract sufficient knowledge from data, but that they have a high chance to suffer from gradient descent vanishing problem and their accuracy tend to degrade as the network goes deeper during the learning process. In the same study, He proposed a residual network for the purpose to ease the network training and solve the vanishing problems often encountered in deeper network training. In the network He proposed, a residual map was fed to few layers acting as skip connections to let the network redirect the derivatives through the network by skipping some layers. Therefore, the introduction of this convolution residual network helped the network achieve great accuracy within deeper designs. Yu et al. [28] has proved in his research on the melanoma recognition and segmentation in dermascopy images that the improved version of CRN he named Fully Convolution Residual Network (FCRN) performs even better than the standard RCN that it has the ability to perform pixel-wise prediction, which is commonly of big interest in many segmentation tasks.

2.2 Prostate segmentation methods

Performing an accurate prostate segmentation is a tedious and challenging task and these difficulties differ according to the imaging modalities such as TRUS, MRI and CT [29]. Generally, based on the given information about segmenting the prostate, prostate segmentation methods are classified as contour and shape based methods, region based, supervised an unsupervised classification based methods and hybrid methods and also based on how suitable these methods can be applicable to imaging modalities. Theoretically, prostate segmentation methods can be categorized in one of the four categories.

Contour and shape based methods: these methods base on the prostate boundary or edge information and include the prior shape information since the prostate edge information does not often provide enough information from TRUS, CT or MR image for a successful prostate segmentation.

Region based methods: these types of methods rely on the local intensity or statistics like mean and standard deviation in an energy minimization framework to perform segmentation. It is worth noting that the methods belonging to this category differ according to the energy minimization used. For example, atlas based methods, region based level sets methods.

Supervised and unsupervised classification based methods: are the types of methods that segment or separate the objects in an image based on features such as intensity or higher dimensional features, then classify the objects as belonging to the foreground or the background region.

Hybrid methods: these methods are robust for the prostate segmentation due to they result from the combination of the contour, shape, region or supervised/unsupervised classification information to achieve promising segmentation results, even vis a vis images with artifacts and noise. Based on the applicability of a method on imaging modalities, the above mentioned prostate segmentation methods can again be classified into different categories.

- **Contour and shape based segmentation**

  These methods refer to the contour features and shape information to perform the segmentation. Under these methods we can have edge based methods, probabilistic filter based methods and deformable model based methods, which can be either active contour models, active shape models, deformable models or level sets and curve based methods.

- **Edge based segmentation methods** consist of extracting edges in an image using the gradient filters like Prewitt, Sobel, Canny, Robert, etc. However, these methods often do fail to detect best edges in the case of noisy images, which can affect the segmentation results. To alleviate this problem, the edge based algorithm is combined with intensity and texture based algorithms for a successful and promising segmentation operation. Theses edge based segmentation methods can be used on TRUS and MR image segmentation but with

supporting strategies since TRUS images contain low contrast, speckles and shadow regions that may produce insufficient edge information, which affect the segmentation results. On the other hand, MR images often produce many false edges due to the high soft tissues contrast. To alleviate these issues in TRUS images, local standard deviation is used to distinguish homogeneous and heterogeneous regions in a multi-resolution framework, then resulting information help detecting reliable prostate boundaries [3], while in MR images the prior information of the prostate shape is used to refine the prostate boundary.

- **Probabilistic filtering:** the prostate segmentation using probabilistic filters such as kalman filter, probabilistic data association filter or particle filter has proven to converge fast in TRUS image segmentation since it does not require the optimization framework that could take time to complete. However, initializing and extending these methods to 3D segmentation is tedious that they are only used in TRUS and not in CT and MRI image segmentation whose images are commonly represented as 3Dimensional images.
- **Deformable model based segmentation:** are methods that base on the theory of geometry, physics and mathematical optimization whereby, geometrical theories serve to create constraints on the model shape while physical theories give information on how the shape gradually changes in space and the mathematical optimization theories dictate a way of fitting the model with available data. For maintain the deformable model stability, these models are submitted under the internal and external energies, of which, the external energies push the model towards the object boundary while internal energies deal with keeping the smoothness of the contours during deformation.

  Active contour models (ACMs) are a type of segmentation method that makes use of the constraints and energy forces in an image for the ROI segmentation and its main purpose is to define the smooth shape in the image and form close contour for the region of interest by separating the foreground from the background of the image. They are a kind of curves moving within images to

minimize the energy function generated by both internal and external forces. The minimized energy functional is the weighted combination of internal and external forces. These models were introduced for the first time by Kass in 1998 [30]. For a successful segmentation process, the active contour models can include various types of contours, being snake model, gradient vector flow snake model, balloon model and geometric or geodesic model. These models are applicable to TRUS and MRI images but due to the low contrast in TRUS images, localizing true image boundaries becomes a challenging task. Zaim introduced the use of difference of Gaussian followed by no-maximal suppression to detect dot patterns that seem to be coherent with the prostate tissues texture [6]. The very importance of active contour models resides in their application in the segmentation and processing of various types of images from 3-D imaging modalities including CT, MRI, PET. Moreover, the active contour models provide a crucial support in the early diagnosis and detection of abnormalities in the ROI of 3-D images; which facilitate the accurate description of the region [31].

Deformable mesh uses a volumetric mesh to represent the segmented shapes of interest, which is the best way to exploit more information in the image and has proven to be less computational expensive regardless of the image content. In addition, this deformable model performs well on the noisy images and incomplete data, due to the customization according to the clinical claim and specialization capabilities [32].

Active shape model (ASM) is a type of deformable models that is used when there is anatomical structure variation of the segmentation results. i.e. when there is no prior shape information to guide the segmentation process. In such case, this model serves to maintain the principal modes of shape variations of anatomical structures of the image under segmentation. ASM offers many applications in different image segmentation problems due to its robustness. For instance, Betroui improved the prostate edge and reduced the noise using a

prior knowledge of the noise in TRUS images then used ASM to produce the prostate segmentation results. Cootes et al. [33] proposed an ASM framework for prostate segmentation in MRI slices.

Edge based level sets have been introduced for its first time as a framework by Osher et al. [34] and have proven to be powerful and efficient for medical image segmentation. The main purpose of edge based level sets is to learn the curvature propagation in higher dimensional image data and learns by expanding from a seed point in a direction normal to the curve area that produces the segmented contour with a speed inversely proportional to the intensity gradient. The framework expansion stops at the point where the intensity difference is the highest in a local neighborhood. Therefore, the curve expansion provides the level sets framework with support in finding an object contour and allows an efficient curve splitting and merging based on the changes in topology. Give that the prostate gland' s anatomical structure is heterogeneous, its segmentation using traditional level sets initialized on gay-scale image is challenging especially from TRUS images. To solve this problem, Kachouie et al. [35] proposed a Gaussian filter, the performed a morphological filtering to classify the mid prostate gland image as belong to the prostate or non-prostate region.

Curve fitting consists of fitting the parametric curves such as splines, ellipses, etc. in the model during prostate segmentation. Parametric curves benefit from the resemblance between the central gland of the prostate and elliptical curve to act as internal forces while the intensity gradient act as external forces to deform the curve towards the prostate edges.

The curve fitting concept was applied in prostate segmentation from TRUS images where Hu proposed using the ellipsoid curvature initialized from the manual delineations of the limits of the axes to perform 3D prostate segmentation.

– **Region based segmentation**

Region based segmentation methods have been developed based on the level of the intensity distributions of the prostate region in various imaging modalities.

Atlas

An atlas is formed from the manual segmentation of the anatomical structures registered to a common coordinate frame, then used as a reference during the segmentation of new image data. Hence, the atlas-based segmentation consists of mapping a pre-segmented atlas image to the target image data. This property made atlas based segmentation robust in the segmentation of the stable structures over large data sets. In addition, atlas based segmentation approach offers a lot of applications in CT and MR image segmentation [36].

Graph partitioning

Graph based segmentation approaches consider pixels or groups of pixels as nodes and the edges between pixels as costs, then segment the graph by minimizing a cost function (distance between pixels) and grouping the closely related pixels into different classes. This approach uses algorithms such as spanning tree, minimum cut or normalized cuts for a successful segmentation.

Region based level sets

These methods use region based statistics for the energy minimization criteria in order to propagate the level set and segment the image data. Chan and Vese in [64] proved these methods to perform better than traditional edge based level sets even on images with weak edges or noisy images from TRUS imaging modalities and from many other imaging modalities.

– **Supervised and un-supervised classification based algorithms**

These methods aim at partitioning the feature space into a set labels for different regions by using a cluster or classier based technique. On one hand classifier based segmentation, the classifiers learn from a set of training data with labeled objects as the prior information to create the predictor that will assign labels to new un-labeled data. On the side of clustering based segmentation, clustering methods base on the given feature vectors to determine the groups of similar objects according to the feature vector linked to each group.

Classifier based segmentation

For the prostate segmentation using classifier based methods, the prostate is considered as the learning problem, then from the training data set each object is correlated with a class label and a feature vector. The training data is used to build the predictor that will assign a class label to the object based on the observed feature vector. These methods have had hard to segment TRUS images due to the inhomogeneities, low contrast texture and noise in images. To alleviate this complexity problem, Zaim proposed texture features, spatial information and gray-scale values in the prostate segmentation [21].

Clustering based segmentation

The methods used for a cluster based segmentation aims at creating essential predetermined number of groups from a set of un-labeled data based on the distance measures where each data is linked with a corresponding feature vector and the groups of similar objects are identified based on a predetermined set of features.

– **Hybrid segmentation methods**

These methods consist of combining various prostate gland information holders such as the priory boundary, shape, region based and feature information to perform a more accurate prostate segmentation. More importantly, hybrid methods have a great advantage of being applicable to the above mentioned imaging modalities.

Many researchers have published various prostate segmentation methods; of which some are mentioned in this work. A range of methods have been proposed to segment the prostate gland by using feature based machine learning. For instance, Maan et al. [37] proposed a segmentation approach using multi spectral MRI such as T1, T2 and proton density weighted images and used the non-parametric and parametric classifiers, Bayesian-quadratic, and the K-nearest neighbor for prostate segmentation.

Habes et al.[38], proposed an automatic prostate segmentation in the whole-body MRI scans for epidemiological studies. This method based on the SVM to detect the

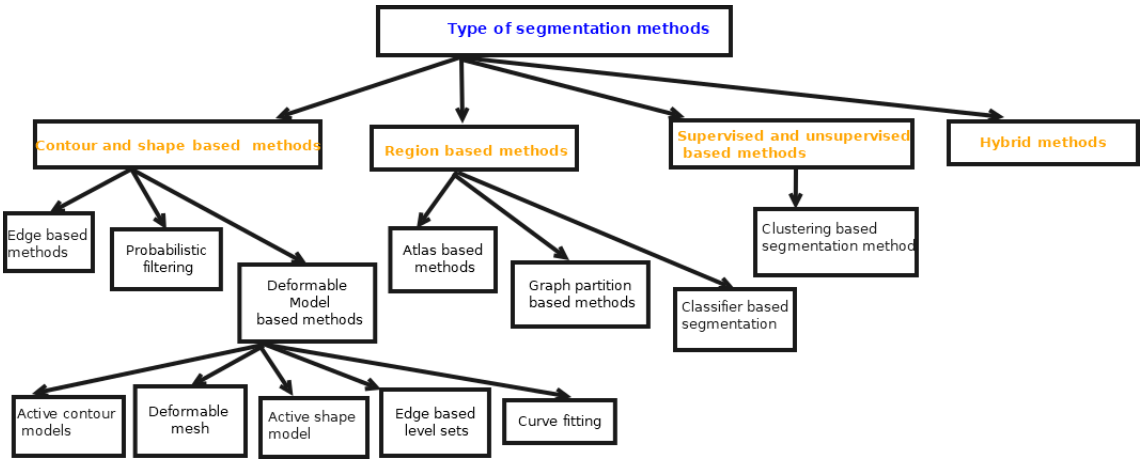


Figure 2.2: Summary of prostate segmentation methods

prostate on MRIs and used SVM binary classifier to segment the 3D MRI voxels. Then the resulting automatic 3D features such as median, gradient, anisotropy and Eigen values of the 3D tensor were used to create the classification binary mask for segmentation process.

Later on, deep learning methods emerged to solve more complex and huge data. Wang proposed classic Generative Adversarial Network-based automatic segmentation method named SegGAN to segment the prostate from MRI. Tian et al. [39] proposed an end-to-end prostate CNN-based segmentation method he named PSNet to segment prostate from MRI. The method was able to create an inference for the pixel-wise segmentation task and has achieved satisfying results. However, the prediction results found that the number of foreground voxels were less than the ones of the background voxels, which may cause the model to get trapped in the local minima with the use of soft-max loss function. Thus this author used the weighted cross entropy loss function to prevent the model from classifying the prostate voxels as background. Aldoj proposed a new neural network for automatic segmentation of prostate segmentation and its zones, with a purpose to alleviate the problem of segmenting the complex anatomical structure of prostate. This network is called Dense-Unet and was built upon the combination of DenseNet and U-Net and proved

the hypothesis that the network can learn well even for inaccurate labels.

Liao et al. [40] used the independent analysis in the deep learning framework that he proposed to learn the effective features in the automatic segmentation of the prostate from MRI. In the study by Guo in [41], a deformable automatic prostate segmentation method was proposed and developed by putting together the deep feature learning and the sparse patch matching for better results. Cheng et al.[3] proposed the enhanced holistically nested networks (HNN) as a full automatic prostate segmentation method with the ability to handle significant heterogeneity observed in prostate T2 MR images through using the salient edge map to learn the hierarchical features of prostate MRIs.

Chapter 3: Methodology

In this chapter, the MRI, CNN and U-Net that are involved in making the implementation of the target prostate segmentation model from MRI possible will be introduced in details. On the side of MRI, a fundamental overview of the basics of magnetic resonance imaging is reviewed whereby a background in physics and engineering is discussed. The goal is to concentrate and keep it as intuitive and simple as possible on the principles of magnetic resonance imaging. The direction of the B_0 magnetic field is described as the direction of the z-axis. The basic overview of the CNN architecture is detailed and finally the U-Net architecture used in the implementation of the proposed automatic segmentation model is described in detail.

3.1 Magnetic Resonance Imaging (MRI)

Medical imaging is defined as a collection of methods and mechanisms of generating visual representations of the body' s inner organs for clinical analysis and medical intervention as well as the visual representation of the role of certain tissues and organs. Medical imaging attempts to expose internal structures hidden by the skin and bones as well as for disease diagnosis and for treatment planning. Medical imaging also creates a normal anatomy and psychology database to facilitate the identification of abnormalities. Computer Tomography (CT), Magnetic Resonance Imaging (MRI) and Positron Emission Tomography (PET) are the key modalities in medical imaging [29].

Magnetic resonance imaging (MRI) has become a central pillar in modern medicine practices. MRI is an advanced imaging technology which produces soft tissue contrast of high spatial resolution. The images resulting from MRI have a tomographic 3D presen-

tation with the ability to show dynamic physiologic changes. From the alterations made during MR images acquiring, various images can be generated to represent a huge array of physical phenomena based on the rich physics of nuclear magnetic resonance (NMR) [42], [43]. The MR images can be created with different contrast reflections such as proton density, T1 and T2 relaxation time, tissue susceptibility variations, diffusion, temperature, tissue perfusion, oxygen levels, etc.

The MR images contrast

The image contrast is key important in image analysis. Therefore, a good image contrast should be taken into consideration with a purpose to get enough information for a successful image analysis. However, an MRI image is affected by tissues' characteristics and other various components, among others being T1, T2 and T*2 relaxation, spin density, susceptibility effects and flow effects.Relaxation is a mechanism by which the energy obtained from an RF pulse is released by spins. The main characteristics of biological tissues have an impact on MRI signal strength, brightness which determine the image contrast [44].

- The proton density i.e. the number of excitable spins per unit volume, determines the maximum signal that can come from a tissue. The resulting images are called weighted proton density images, since they are generated by emphasizing the proton density while minimizing T1 and T2.

- A tissue' s T1 time defines the time it takes to recover the excited spins and to be available for the next excitation. Therefore, T1 impacts the signal amplitude and can vary randomly. The contrast images which are calculated by T1 are referred to as T1-Weighted images (T1W).

- After the excitation, the T2 time mainly describes the easiest possibility for an MRI signal of fading. And the contrast images that are primarily calculated by T2 are called T2-Weighted images (T2W).

The pertaining characteristics of biological tissues are proton density, T1 and T2 times and differ from one tissue to the other. The resulting images also vary in their tissue-tissue

contrast based on which of these parameters is highly considered in the MRI series. This justifies the MR images ‘charm soft-tissue discrimination and diagnostic potentials. For example, tissues that are practically indistinct on computed tomography (CT) scans can be differentiated by MRI without contrast medium administration on the basis of their particular differences in terms of the above mentioned three parameters.

Key parameters influencing MR image generation

Repetition Time (TR)

In the MR image generation process, a slice is excited and the resulting signals are recorded several times to create an MR image. The interval between two consecutive excitation steps of the same slice is called repetition time (TR) and has a great impact on the contrast of T1. More excited spins rotate back into the Z-plane when TR is long, and leads to the longitudinal magnetization growth. The image contrast is greatly influenced by T1 if a short repetition time is considered.

Note that biological tissues with a short T1, relax quickly and give a large signal after the next RF pulse under this condition and thus appear bright on the image. on the other hand, tissues with a long T1, relax slowly between two RF pulses and therefore, less longitudinal magnetization will be available when the next excitation pulse is applied. Hence, these tissues emit less signal than tissues with a short T1 and do not look nicer on the image. Tissues with a short T1 such as fat, bones look brighter since during the TR interval they recover much of their longitudinal magnetization and thus give a stronger MRI signal, whereas tissues with a long T1 such as muscle, cortex appear darker because they do not recover much of their longitudinal magnetization during the TR interval, and therefore, generate a weaker signal.

Echo Time (TE)

When the refocusing pulse is applied, the previously mentioned effects must be reversed before obtaining an appropriate MR signal. The signal induced after phase coherence is restored in the receiver coil and is known as a spin echo and can be measured. The interval between the application of the excitation pulse and the compilation of the MRI signal

is called Echo Time (TE). The echo time has an impact on the effect of T2 on the image contrast in that, a short echo time that is less than 30msec, causes little signal difference between tissues. This is because T2 relaxation has only just begun and there was only a small signal decay at the time of echo collection. This implies a low T2 weighting in the resulting image. in other words, tissues with a short T2 that have lost most of their signal appear darker in the image while tissues with a long T2 generate a stronger signal and therefore, appear light. To make the image generation more successful, the operator can control the degree of T2 weighting of the resulting MR image by choosing an echo time (TE).

Additional Image Quality considerations

1. **The power of magnetic field**

   The application of a higher magnetic field strength enhances the longitudinal magnetization since this causes a large number of protons to converge around the main axis of the magnetic field, resulting in an increase in SNR. The enhanced SNR obtained with high-field systems can be used to generate images with high resolution or to carry out fast imaging.

2. **Bandwidth of the slice thickness and receivers**

   Thinner slices contain a lot of noise and this implies less SNR. The increase in partial volume results also contributes to the slice thickness. The receive bandwidth is defined as the spectrum of frequencies gathered during the encoding by an MR system. The frequencies gathering include much of noise, therefore, a large receiver bandwidth allows for quicker data acquisition but again reduces the SNR. On the other hand, there is a possibility that there can be more motion objects with a narrow bandwidth and the number of slices that can be acquired for a given TR reduced.

3. **Sending and receiving coil collection (RF coil)**

   Selecting a suitable radio frequency (RF) coil provides a better way to improve the SNR without needing to increase voxel size or lengthen the scan time. The RF coil should usually be as similar to the anatomy being imaged as possible and surround the target organ. The closer the coil can be put under inspection to an organ, the better

the sign will be. It is possible to use RF coils either to transmit RF or receive the MR signal or only as receiver coils.

4. **Field of View (FOV)**

   The FOV defines the size of pixels in an image. for unchanged matrix, a smaller FOV results in a smaller pixel size, and the pixel size is of paramount importance for the spatial resolution of the MR images. In other words, for the same FOV, a higher spatial resolution benefits a lot from a higher matrix. Contrarily, for a constant FOV, a coarser matrix results in a lower spatial resolution.

5. **The size of the image matrix**

   A matrix is made up of a grid of rows and columns that is in 2-Dimensional. Each individual square of the grid is a pixel that is given a value corresponding to the strength of the signal. Therefore, each MR image pixel provides information on a corresponding 3D unit of volume, named a voxel. The voxel scale specifies an MR image' s spatial resolution.

6. **Parameters of imaging**

   The TR, TE and flip angle are the important parameters of MR imaging. The choice of imaging parameters affects the SNR in that the SNR rises with TR but at longer TRs, the T1 effect is also lost. Contrariwise, the TE growth leads to the SNR decay. The T2 comparison is lost in a short TE. For this reason, only T1-Weighted sequences have a choice of shortening TE to enhance SNR.

7. **The number of excitation steps**

   The number of excitation steps shortened as NEX, defines how many times a signal is calculated from a given slice. The SNR increase improves proportionally to the NEX increase. However, the time for scanning increases linearly with NEX.

3.2 Convolution Neural Network (CNN)

Deep learning has attracted many researchers' interest in recent years and the CNN has proved the most outperforming algorithm among different deep learning models [10]. CNN is a class of artificial neural network that has attracted many researchers' attention in computer vision tasks due to the highly promising results shared in the object recognition competition named ImageNet Large Scale Visual Recognition Competition (ILSVRC) in 2012 [45],[46]. In addition, CNN has achieved great results in various medical research fields. For instance, Gulshan et al.[47], Esteva [48], and Ehteshami Bejnordi et al. [**?**] proved the potential of deep learning for retinopaphy screening, skin lesion classification and lymph node metastasis detection respectively. Moreover, many publications have been made in several areas among others being lesion detection [49], classification [50], segmentation [51], image reconstruction [52], [53]], and natural language processing [22]. CNN is designed to automatically and adapt to learning spatial hierarchies of features from low to high level patterns. Mathematically, CNN is defined as an architecture built upon three main building blocks/layers: convolution, pooling and fully connected layers. The convolution and pooling layers play an important role of extracting features from input data while the fully connected layer maps the extracted features into the final output. At the convolution layer, a bunch of operations are performed on the 2D or 3D images depending on whether the network is a 2D- or 3D-CNN. These operations are followed by a small grid/an array of parameters named kernel that acts as an optimizing feature extractor from image at each position. These two operations, i.e. convolution and kernel make the CNN an efficient method for image processing because a feature can occur anywhere in the image. Along with feeding each layer's output as input to its next layer, the number of extracted features grow progressively and make the learning process complex. The parameter optimization process is performed in order to minimize the difference between outputs and ground truth through the back-propagation and gradient descent.

3.2.1 Reasons for choosing CNN

The key difference between CNN and traditional radio-mics studies resides in that the latter mostly use the hand-crafted feature extractor methods including texture analysis, followed by traditional machine learning classifiers such as Random Forest and Support Vector Machine [54],[51] whereas CNN does not require any hand-crafted feature extraction. Moreover, CNN does not require human expertise for segmenting an organ or tumor and it is computationally expensive due it learns from huge amount of data. Therefore, CNN models require Graphical Processing Units (GPUs) for increasing the model training speed.

3.2.2 CNN Architecture

The CNN architecture is made up of repetitive stacks of different convolution layers and a pooling layer, followed by one or more fully connected layers. The step at which the input data is transformed into output is called forward propagation. According to the loss between the output and label, the back-propagation is used to learn the convolution kernel coefficients and weights.

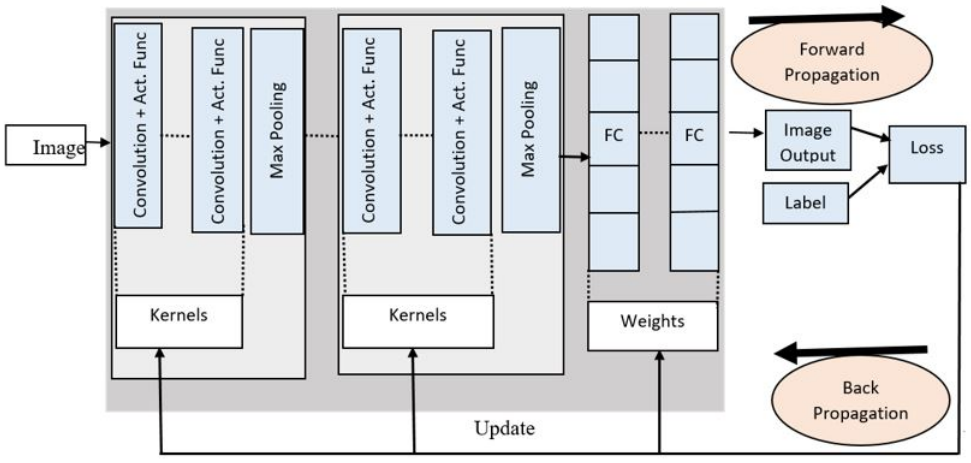


Figure 3.1: CNN basic architecture

- **Convolution layer**

  The convolution layer is the basic building block of CNN. Its main function is to perform feature extraction on the input image through the combination of the linear and non-linear operations: convolution operation and activation function. The idea of stacking a bunch of convolution layers in a CNN model allows the layers close to the input to learn low-level features (e.g, lines) and the layers deeper in the network to learn abstract/high-level features such as shapes.

  Convolution operation

  The convolution operation applies the learned filters known as kernels to extract features from the input image, with a purpose to create feature maps that richly summarizes the features present in the input. At each tensor location, the element-wise product between each kernel element and the input tensor is determined and summed to obtain the output value at the corresponding output tensor location referred to as the function map. This process is performed repeatedly by applying multiple kernels to form an arbitrary number of feature maps representing different input tensor characteristics. Different kernels may therefore be regarded as different extractors of features. Note that the size and number of the kernel are two primary hyper-parameters that describe the convolution process. Frequently, the former is 3x3 and the latter is random, defining the depth of maps of output features [10].

  Weight sharing is the main feature of a convolution operation by which kernels are shared across all positions in the image and generate the characteristics among others being: 1) making local feature patterns extracted by translation of kernels invariant as kernels move through all image positions and recognize local learned patterns. 2) by down-sampling in combination with a pooling operation, learning spatial hierarchies of feature patterns, resulting in an increasingly wider field of view, and 3) increasing model performance by reducing the number of parameters to be learned compared to fully connected networks.

  Briefly, the training of a CNN model with regard to the convolution layer aims at finding the kernels that perform optimally based on a given training data set for a

given task. The kernels are the only parameters that are automatically learned in the convolution layer during the training process, on contrary, the size of the kernels, the number of kernels, padding and stride are the hyper-parameter that are to be set before starting the training process [10].

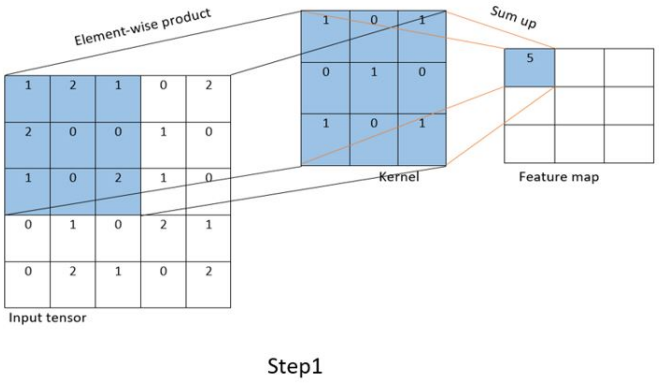


Figure 3.2: Examples of Convolution step1

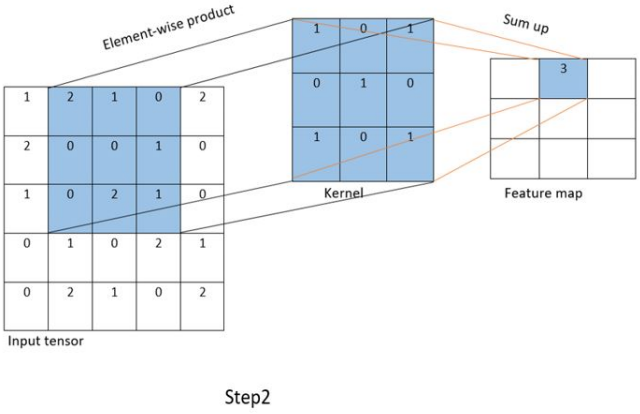


Figure 3.3: Examples of Convolution step2

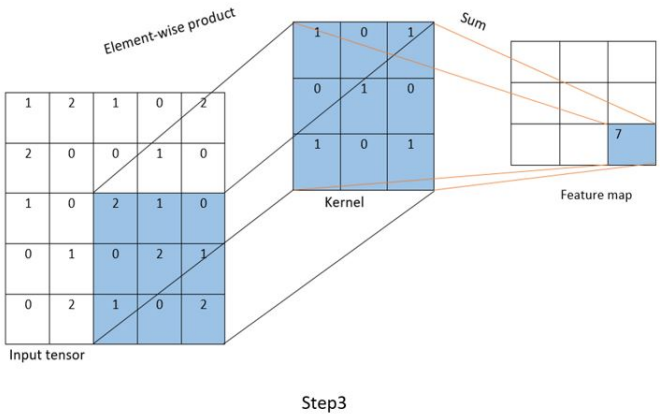


Figure 3.4: Examples of Convolution step3

3.2.3 **Activation functions**

The activation function is a non-linear function that operates on the outputs it receives from the linear operation such as convolution. Various non-linear functions including sigmoid or hyperbolic tangent function have been used in the past decades but the Rectified Linear Function (ReLU) got a lot of popularity over other activation functions in deep learning since it does not activate all neurons at the same time, which speeds up the training process.

- **Sigmoid function**

  The sigmoid function often referred to as the logistic function, is a non-linear activation function which is preferably used in the feed-forward neural networks. However, the sigmoid function has a downside of producing a slow model convergence, causing sharp damp gradients during the propagation. As shown in Figure3.6, Sigmoid function computes the function:

$$f(x) = \frac{1}{1+e^{-x}} \tag{3.1}$$

  and is defined in interval [0,1].

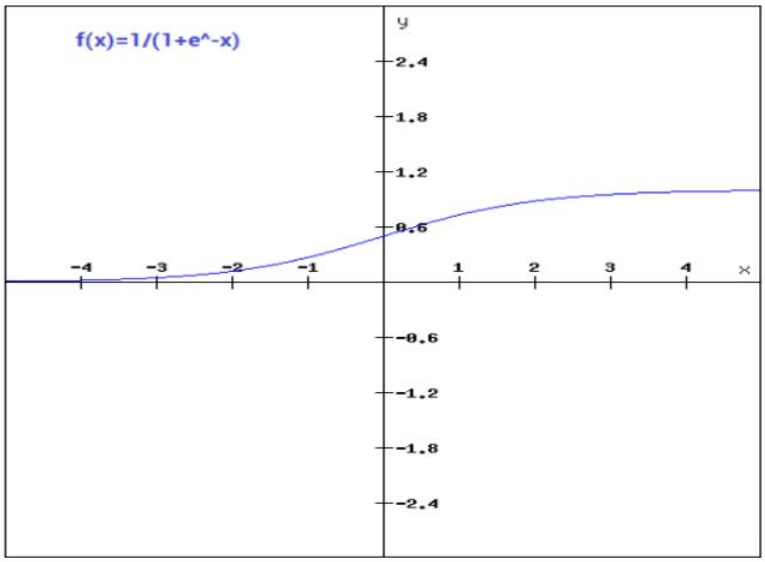


Figure 3.5: Sigmoid Function

- **Softmax function**

  The softmax function is a type of activation function also used in neural computing, especially in multi-class models where it consists of producing the output whose probability ranges between 0 and 1 with a total sum of the probabilities being equal to 1. The softmax is given by the relationship:

$$f(x_i) = \frac{exp(x_i)}{\sum_j exp(x_j)} \tag{3.2}$$

  Among the probabilities returned by softmax, the target class is displayed with the highest probability. The softmax function is mostly applied to the output layers and differs from the sigmoid function in that it is applied in multivariate classification problems while sigmoid function is applied in binary classification tasks.

- **Rectified Linear Unit (ReLU)**

  Proposed by Nair and Hintson 2010, ReLU is the most widely used activation function in deep learning areas where it has proved to achieve great accuracy. ReLU has

many advantages over its counterparts, including being the fastest learning activation function and offers the best performance and the model generalization ability. ReLU has its unique property of representing a nearly linear function, which allows it to preserve the properties of linear models; this made them easy to optimize using gradient descent algorithms. In addition, ReLU has the ability to alleviate the vanishing gradient problem observed in other activation function mentioned in above [10].
The ReLU activation function performs a threshold operation to each input element where values less than zero are set to zero as plotted in Figure3.8, ReLU computes the function [55]:

$$f(x) = max(0,x) = \begin{cases} x_i, & \text{if } X_i \geq 0 \\ 0, & \text{if } X_i \leq 0 \end{cases} \tag{3.3}$$

This means that the neurons will only be deactivated if the output of the linear transformation is less than 0 since this function count only values that are greater than or equal to zero.

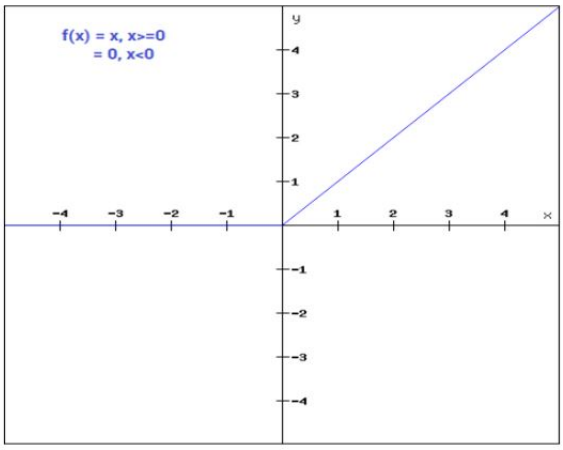


Figure 3.6: Rectified Linear Unit (ReLU) function

- **Pooling layer**

  The pooling layer is a new layer that comes just after the non-linear function like ReLU to reorder the layers present in the CNN that may be repeated one or many times in a model. The pooling layer is added in the network to operate upon every feature map separately in order to create a new set of the same number of pooled feature maps. The pooling operation is chosen much like a kernel to operate on the feature maps with a purpose to reduce its size. The most commonly used pooling operation is of kernel size 2x2 applied with a stride of 2 pixels. It is of note that contrary to the filter/kernel, the pooling operation is specified rather than learned. Average and maximum pooling are the common function used in the pooling operation.

  Let the below matrix be one of the 4x4 pixels feature maps from the convolution layer.

Table 3.1: A 4x4 feature map

20	10	20	182
1	95	78	40
10	14	10	4
12	12	40	6

Maximum pooling

Maximum pooling shortened as max pooling extracts patches from the input feature maps by getting the maximum value in each patch and ignoring all other values, down-sampling the height and width but keeping the depth dimension of the feature map unchanged.

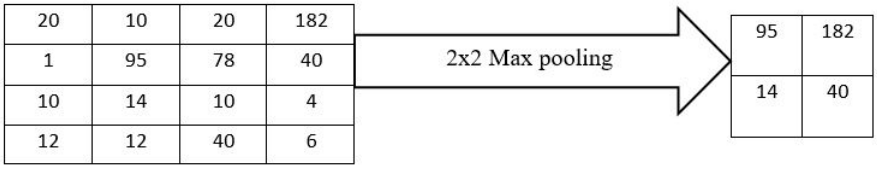


Figure 3.7: Max pooling example

Average pooling

Also referred to as global average pooling, the average pooling operates on the input feature maps to extract patches from the extreme feature maps whereby, the feature map is down-sampled into a 1x1 array of numbers, considering the average of all elements in each feature map, and leaving the depth of feature maps unaltered. The average pooling is preferably used towards the end of the network, just before the fully connected layers where it offers a great advantage to the network of decreasing the number of learnable parameters and allowing the CNN to receive inputs with variable sizes. The down-sampled or pooled feature maps resulting from the pooling

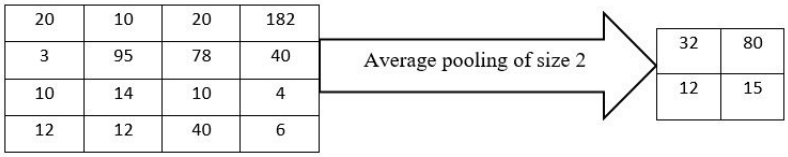


Figure 3.8: Average pooling example

operation make a summarized version of the features detected in the input, and are of paramount importance since small changes in the location of the feature in the input detected by the convolution layer will give a pooled feature map with the feature in the same location thanks to the invariance translation introduced by the pooling operation. i.e., the input translation by a small change does not change the values of the pooled outputs.

- **Loss Function**

  During the model training process, the optimizer, i.e. Adam optimizer in our case, repeatedly tries a set of parameters known as weights and biases until it finds the optimal values for the model to produce the accurate output image. The loss function comes in to help the optimizer assess whether the it is trying good values and if the model is achieving good accuracy with promising results as the training goes on.

  The loss function also called error function or cost function, is a key concept in model training where it is calculated as the difference between the predicted result and the

target value, i.e. the ground truth. Therefore, the choosing and configuring a loss function for a model should be done attentively since if it is incorrectly done, it will break the model performance and finally affects the model results.

A large number of loss functions such as: Binary Cross-Entropy for binary classification tasks, Weighted Binary Cross-Entropy, Hausdorff Distance Loss, Balanced Cross-Entropy, etc as shown in Figure 3.9, have been established in deep learning to help gauge how well the model learns from the data set along the whole learning process.

Depending on the problem in hands, these loss functions can be grouped into 4 types: Distribution-based, Region-based, Boundary-based and Compound-based loss functions [56].

In the case of our proposed algorithm, we used the Dice Loss function, a region-based loss function introduced in computer vision by Milletari et al.[26] in 2016 for 3D medical image segmentation to determine the similarity between the manual segmented image, i.e. the ground truth and the automatic segmentation image output. The Dice Loss function is given by the following formula:

This Dice Loss function is inspired by the Dice Coefficient, which is a measure of similarity between two samples and whose value ranges from 0 to 1. The greater the value the better the similarity. when the two samples are perfectly similar, the dice coefficient takes its value to 1, otherwise it takes it to 0, and meaning that the two samples are completely different. Thus, the Dice Loss function is calculated by the formula: 1-Dice coefficient f or it to maximize the similarity between the samples.

The Dice Coefficient (DSC) is defined as:

$$DSC = \frac{2\sum_i^N p_i g_i}{\sum_i^N p_i^2 + \sum_i^N g_i^2} \tag{3.4}$$

The Dice Loss Function is in turn given by the following formula:

$$D_{Loss} = 1 - \frac{2\sum_i^N p_i g_i}{\sum_i^N p_i^2 + \sum_i^N g_i^2} \tag{3.5}$$

Where: p_i & g_i represent the pairs of pixel values of prediction and ground truth, respectively.

Note that the Dice Loss function has proven to outperform other loss functions in medical image segmentation because it has its uniqueness of considering the loss information both locally and globally, which is very important for a model to achieve a high accuracy. Below is the figure showing some examples of Loss functions used in semantic segmentation.

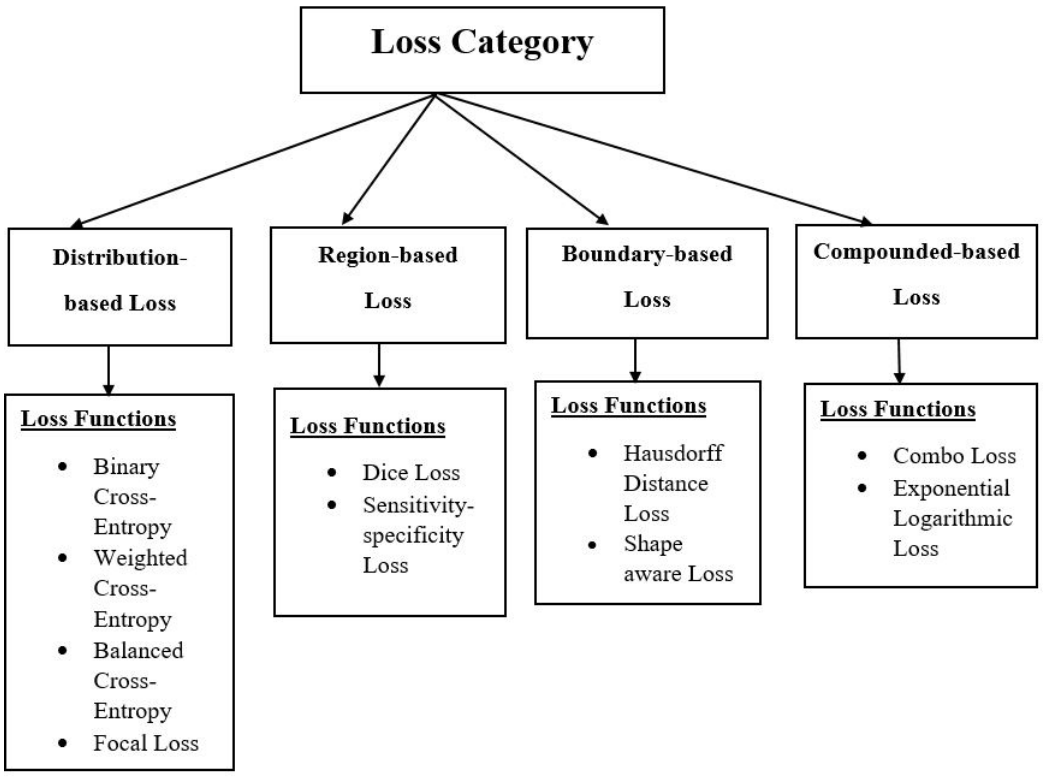


Figure 3.9: Example of Loss functions used in semantic segmentation

3.3 3D U-Net

The developed U-Net architecture comprises two main parts for it to successfully segment the prostate from MRI. Those parts are the contracting path also referred to as encoder path and expansive path referred to as decoder path. To train and evaluate the performance of the model itself, the 10 fold cross-validation approach is used then compute the mean dice score over all the model dice scores of the 10 fold cross-validation. The proposed method is also compared to other segmentation methods presented in the literature by comparing their qualitative and quantitative metrics.

Given that the prostate volume vary from patient to patient and that 3D network requires a common volume size, there should be a selection of a fixed number of total slices around the predicted volume. The total number of slices used in our model is 64. U-Net has an elegant architecture built upon the contracting path and its symmetric path known as expansive path.

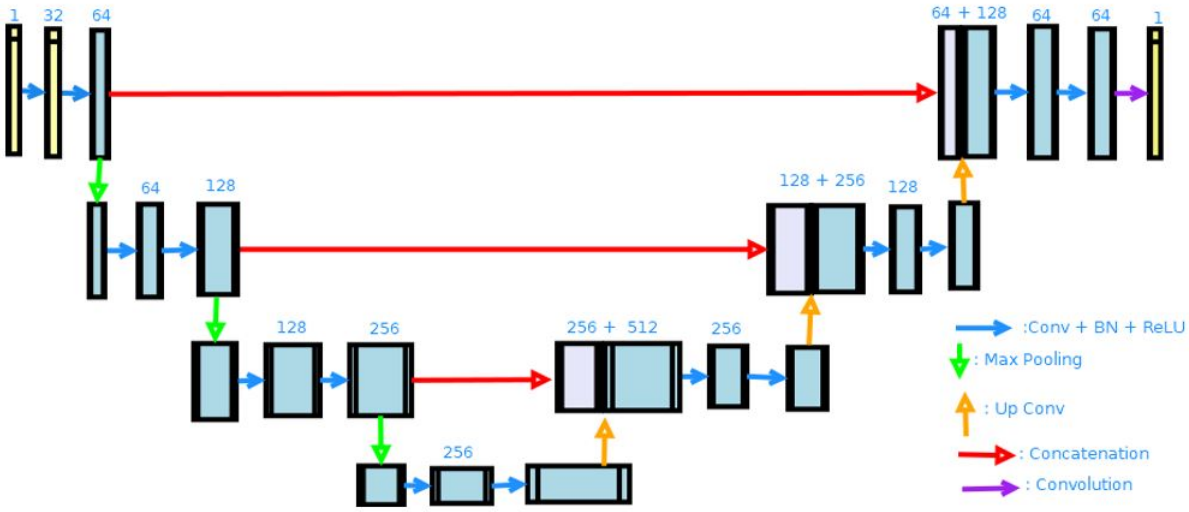


Figure 3.10: The 3D U-Net architecture

The architecture comprises two main paths:

- The left path commonly called contracting path or encoder whose task is to down-

sample the input image by extracting features from it and reduce its resolution using an appropriate stride. The contracting path follows the standard CNN architecture. It consists of repeated application of Convolution, Batch normalization with Activation function, followed by a 2x2x2 max pooling operation with a stride of 2 for down-sapling. It is key to note that at every stage in the contacting path the number of feature channels is doubled while the spatial dimensions, height and width are halved.

- The right path known as expansive path or decoder aims at up-sampling the feature map and halves the number of feature channels until the output image has the same size as the input image. The first operation performed on the expansion path is up-sampling, which makes the size of feature map to increase gradually until it reaches the original size of the input. The 2x2x2 up-convolution for up-sampling is followed by two 3x3x3 convolution operations.

  After every convolution operation, border pixel information is lost and this can affect the segmentation accuracy. Therefore, after each 3x3x3 up-convolution a concatenation operation is performed to forward the features extracted from early stages of the contracting path to the expansive path for a better information localization. Then after localizing the information, the segmented image is displayed through the last 1x1x1 Convolution layer.

  Batch normalization is performed during the training process to avoid the bottleneck and facilitate the fast network convergence. Batch normalization is a method of normalizing the data directly during the neural network training after the weighted sum operation and before passing data through the activation function. In this study, each batch is normalized during the training process for all data to be on the same scale.

1. **Layers architecture at the down-sampling path:**

   Convolution layer #1 applies 32, 3x3x3 kernel filter, Batch Normalization followed by ReLU Activation Function

   Convolution layer #2 applies 32, 3x3x3 kernel filter, Batch Normalization followed by ReLU Activation Function

Pooling #1 applies a max pooling with 2x2x2 kernel filter

Convolution layer #3 applies 64, 3x3x3 kernel filter, Batch Normalization followed by ReLU activation function

Convolution layer #4 applies 64, 3x3x3 kernel filter, Batch Normalization followed by ReLU activation function

Pooling layer #2 applies a max pooling with 2x2x2 kernel filter

Convolution layer #5 applies 128, 3x3x3 kernel filter, Batch Normalization followed by ReLU activation function

Pooling layer #3 applies a max pooling with 2x2x2 kernel filter

2. **Layers architecture at the bridge:**

   Convolution layer #7 applies 256, 3x3x3 kernel filter, Batch Normalization followed by ReLU activation function

   Convolution layer #8 applies 256, 3x3x3 kernel filter, Batch Normalization followed by ReLU activation function

3. **Layers architecture at the up-sampling path:**

   Transposed convolution layer #1 applies 512, 3x3x3 kernel filter, Batch Normalization followed by ReLU

   Up-convolution layer #1 applies 768, 3x3x3 kernel filter, Batch Normalization on 256, followed by ReLU activation function

   Up-convolution layer #2 applies 256, 3x3x3 kernel filter, Batch Normalization followed by ReLU activation function

   Transposed convolution layer #2 applies 256, 3x3x3 kernel filter, Batch Normalization, followed by ReLU activation function

   Up-convolution layer #3 applies 384, 3x3x3 kernel filter, Batch Normalization on 128, followed by ReLU activation function

   Transposed convolution layer #3 applies 128, 3x3x3 kernel filter, Batch Normalization, followed by ReLU activation function

   Up-convolution layer #5 applies 192, 3x3x3 kernel filter, Batch Normalization on 64,

followed by ReLU activation function

Transposed convolution layer #4 applies 64, 3x3x3 kernel filter, Batch Normalization, followed by ReLU activation function

Up-convolution layer #6 applies 64, 3x3x3 kernel filter, Batch Normalization, followed by ReLU activation function

The last convolution layer applies 64, 1x1x1 kernel filter with a sigmoid function, then give the segmented image output.

Chapter 4: Evaluation

This chapter describes the implementation of the proposed method and its evaluation.

4.1 Experimental details

4.1.1 Database description

The data set used, Promise12 consists of 50 T2-Weighted three dimensional Magnetic Resonance images (3D MRIs) made publically available by MICCAI Grand Challenge Team for prostate segmentation purposes. The data set includes both patients with benign diseases (example: benign prostatic hyperplasia) and prostate cancer. For the purpose of generalization ability of the algorithm, the data was collected from multiple centers and multiple MRI device vendors with different acquisition and scanning protocols. The figure 4.1 illustrates some examples of MR images acquired with various protocols, and from different areas for a better model generalization. The data set was split into training and test

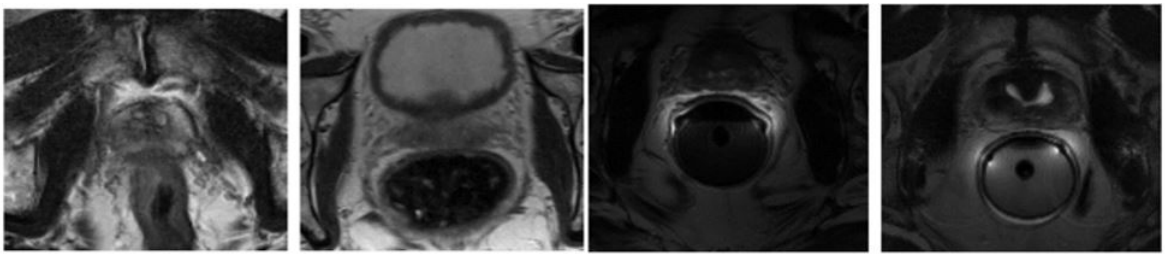


Figure 4.1: Examples of prostate 3D MR images from Promise12 data set

data sets whereby ninety percent was used for training the model and other remaining ten percent were used for testing the model. The ten-fold cross validation approach was used for training and evaluate the model.

4.1.2 Data pre-processing

Data pre-processing and augmentation are important steps in the model construction but before performing any processing on data, we have to assess if the application under development is going to benefit from it in order to make sure that the application really needs that type of data processing. In the pre-processing process of our data, we unify the spatial resolution of each volume to 1.0 x 1.0 x 1.5 millimeter. All volumes processed by the network have a fixed size of 64 x 128 x128.
Normalization is an operation carried out on the data before or during the process of training the network that aims to convert all the data to the same scale within a range of [0-1].

4.1.3 Experimental design

The U-Net architecture proposed in this study is the extension of the U-Net architecture from Ronnerberger et al. and replaces all the 2D operations with their matching 3D counterparts. Our 3D U-Net architecture for prostate segmentation from MR images performs segmentation on the 3D T2-Weighted images using 3D operations including 3D convolutions, 3D max pooling and 3D up-convolution layers.
The algorithm implementation was done using Python programming language on the open source PyTorch framework and the computations were performed with NVIDIA Tesla V100 32G.

4.1.4 Experimental parameters and hyper-parameters

Model parameters are variables estimated automatically from the data and are also optimized automatically during the training process while hyper-parameters are set manually beforehand and help in the parameters estimation [10].In this part, a set of parameters and hyper-parameters used for the proposed model is presented whose performance is evaluated depending on those parameters because the latter play important roles in defining the general learning capacity of a model. The proposed network architecture aims at mapping the input image to its corresponding segmented image output with efficiency. The loss func-

tion plays a crucial role of measuring how well the output matches the ground truth data based on the weights and biases. In other words, the loss function allows to characterize the quality of a set of weights w. In many neural networks, the weights are optimized through gradient descent methods such as back-propagation gradient descent.

Gradient descent is an optimization technique that is commonly used in machine learning and deep learning to train models with a purpose to adjust the weights, which minimizes the loss function. Minimizing the loss function is the best way to find the optimal parameters that yield the outperforming model. In these experiments, Adam optimization algorithm the extension of classic gradient descent is used. Adam optimizer have proved to provide best performing models in computer vision tasks including medical image segmentation.

To optimize the modal parameters, the gradient descent starts with initializing parameters randomly, finds the error learning iteration, then keep on updating the model parameters along the training process in order to get the values that lead to minimal cost. Mathematically, to update the modal weights, the gradient descent algorithm performs multiplication operation of the slope by a scalar called learning rate, also called step size to determine the next point that the weight will be updated to. For instance, the gradient dw to update the parameter w for performs the computation $w = w - lr * dw$ (11)

Learning rate expresses the amount that the weights are updated during the training of neural networks for the model for mapping correctly the inputs to outputs in the training data set. The learning rate will help control the rate or speed at which the model learns. It is worth noting that a too large learning rate allows a model to learn faster but may cause it to diverge and the optimizer can overstep the minima due to high changes in weights. On the other hand, a smaller learning rate allows a model to learn a more optimal set of weights but take a lot of time to converge due to steps towards the minimum loss function are tiny. In other words, smaller learning rate leads to training model over-fitting, a larger one allow the training model generalization whereas a too large learning rate causes the model to diverge. The initial learning rate is 0.002 which decreases by a decay of 0.001 after 10, 30, 60 epochs.

Batch size describes the number of training samples used in one iteration. The batch size

choice highly depends on the hardware' s memory and on the execution time of the training. The batch size is 2.

Weight decay is a concept that helps generalize a model not on the side of data things but on the side of model things. During the model training process in the real world practice, there may be a bunch of data that cannot be fit well with a straight line and the idea of increasing the degree of polynomial make a model become more complex at some point then start over-fitting. To prevent the model from facing this problem of complexity, the number of parameters has to decrease.

However, this is does not provide a solution in the real world application. The introduction of weight decay solves the problem by providing the ability to add a lot of parameters to the loss function and at the same time prevent the model from getting complex. Some of the added parameters can be negative and others positive, which can lead to the huge loss function. The weight decay is multiplied with the sum of squares of the added parameters and this process help generalize the model even better than using data augmentation approach.

Mini-batch size expresses the number of samples of the data that are fed to the network from a huge data set. Feeding the whole data set to the network (referred to as batch gradient descent) would be practically inefficient since it is time consuming and causes the model to get stuck before completing the learning process. Therefore, the mini-batch (mini-batch gradient descent) is practically efficient as it consists of involving various random samples that are greater than 1 and less than the whole data set into the model variable updating process. These samples of data are referred to as batches. The typical batch size is 32 and setting a too big batch size leads to an overgeneralized model, which will not provide good results when new data are presented. Briefly, the batch size means the whole data set fed into the training process, while the mini-batch size consists of a number of small training samples that are involved in the training iteration.

Epoch defines how many times the algorithm trains on the data set as a whole. Deciding on the number of epochs depends on the kind of data in hands and the task to accomplish. Different possibilities of choosing the number of epochs for training a model could be either imposing a condition for a model to stop the training process once the error start tending to

zero or to start with a quite small number of epochs and increase it along with the training progress, keeping track of important evaluation metrics such as accuracy. The number of epoch is 100. Below is the table with a brief summary of hyper-parameters and their values defined in our experiment. The table 4.1 lists the the model hyper-parameters used in our experiments.

Table 4.1: The model Hyper-parameter values

Hyper-parameter	Value
Optimizer	Adam
Initial Learning	0.002
Weight decay	0.001
Epochs	100
Batch size	2

4.1.5 **Evaluation metrics**

The evaluation metrics are an important key to the development of successful algorithms since they provide developers with a way to estimate how well the developed algorithm outputs best results for a specific task, how it validates its performance on data and how outperforming it is against other approaches. In image segmentation tasks, the popular approach used to evaluate the model is comparing its results with gold standards referred to as manual segmentation results. In the evaluation of the proposed segmentation algorithm, the overlap based method named Dice' s similarity Coefficient (Dice) and a Surface Distance Based measure, Hausdorff Distance are used to evaluate the performance.

The qualitative evaluation consists of comparing the gold standard which we consider as the actual image, resulting from the manual segmentation performed by expert physicians with the automatic segmentation results regarded as the prediction got from the automatic segmentation by the proposed method. Then comment on how similar they look like on the figures.

Dice Similarity Coefficient (DSC)

Dice Coefficient is a statistical tool which measures the similarity between two finite sets of data. It has become conceivably the most broadly used tool in the evaluation of image

segmentation algorithms. The dice coefficient is a scalar or numeric vector that range between [0, 1], where a similarity of 1 means that the automatic segmentation results and the corresponding gold standard are a perfect match. Mathematically, the Dice Coefficient is computed as:

$$DSC = \frac{2|X \cup Y|}{|X \cap Y|} \tag{4.1}$$

Where X and Y are the segmented result and gold standard, respectively. **Hausdorff Distance (HD)**

In pattern recognition and computer vision problem solving, the challenging task is to determine how much shapes differ from each other. The shape comparison must obey to metric properties. The Hausdorff Distance is an evaluation metric that aims at measuring the degree of mismatch between two finite sets by measuring the distance of the point of X that is farthest from any point of Y [57]. The Hausdorff Distance is computed as:

$$H(X,Y) = max(h(X,Y), h(Y,X)) \tag{4.2}$$

With x and Y represent segmentation results and ground truth respectively. And $\|.\|$ is the norm of the sets X and Y. $h(X,Y)$ plays a role of identifying a pixel xX that is far from any pixel of Y and measures the distance from this x to its nearest neighbor in Y using the norm; then ranks each pixel of X based on its distance to the nearest point of Y. It considers the largest ranked pixel as the distance the most mismatched pixel of X.

4.1.6 Experimental results

To start the experiment, the T2-Weighted 3D MR image is fed into the U-Net architecture. At the beginning of the training process, the training loss was high since the model had just learned from very few data. For instance, at the first 10 epochs, the train loss was high while after 30 epochs the learning rate decreased to 0.00000001, see below Figure 4.2, which plots the learning rate variation along with the model training process.

The train loss variation figure 4.3 shows how high the train loss is at the beginning of the training process and the way it decreases along with the training progress. While the

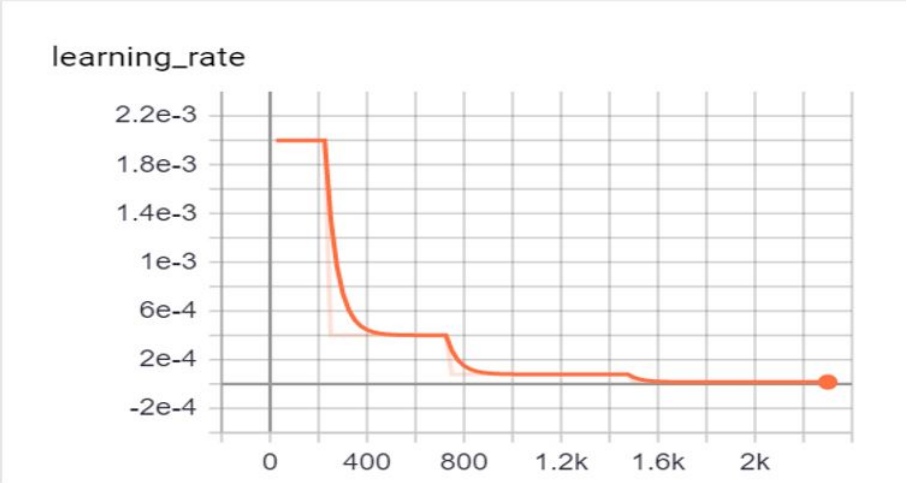


Figure 4.2: Learning rate variation

figure 4.4 shows the train evaluation score increasing with the learning process.

The training evaluation score is low when the mode starts the training and as the training progresses learning from a lot of data the evaluation score increases and when the satisfying accuracy is reached, the score increases slow and tends to remain constant.

Quality evaluation results

The figure 4.5 in below, shows the prostate images manually segmented with their mapped automatically segmented. The manual segmentation gold standard is in blue color and the automatic segmentation result is in red color. For the quality evaluation of the proposed model, we compare the gold standard which we consider as the actual image with the automatic segmentation results regarded as the prediction. The images in figure 4.5 show that the automatic segmentation has been performed successfully since both actual value and predicted value look almost the same.

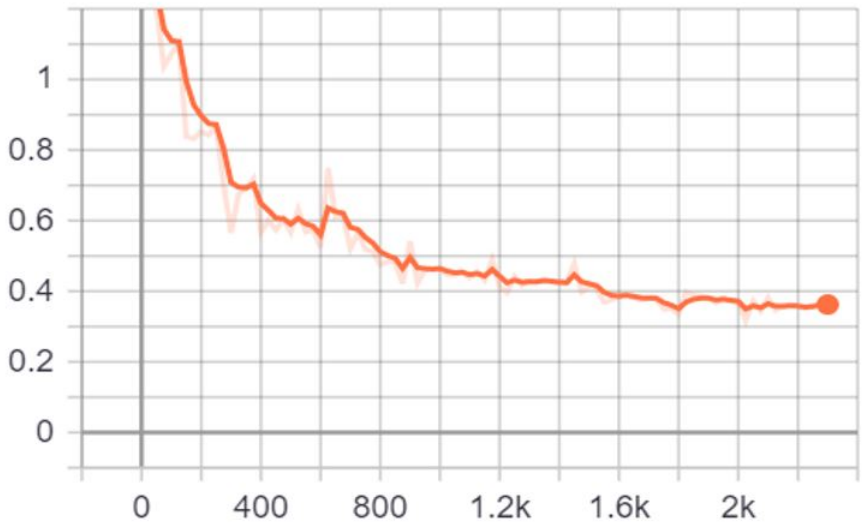


Figure 4.3: Train loss variation

Quantitative evaluation results

To evaluate the quantitative performance of the proposed segmentation algorithm, the calculated Dice Similarity Coefficient gives the maximum of 89.04% and the minimum of 70.75% and the mean DSC of 84.92% as shown in the table Table 4.2; while the HD gives the maximum of 28.4, the minimum of 1.96 and the mean Hausdorff of 5.3mm as plotted in the figure 4.6.

Table 4.2: The quantitative results, DSC values

Fold	1	2	3	4	5	6	7	8	9	10	Mean
DSC (%)	83.64	**70.75**	87.12	88.3	86.54	87.96	83.85	87.29	84.79	**89.04**	**84.92**

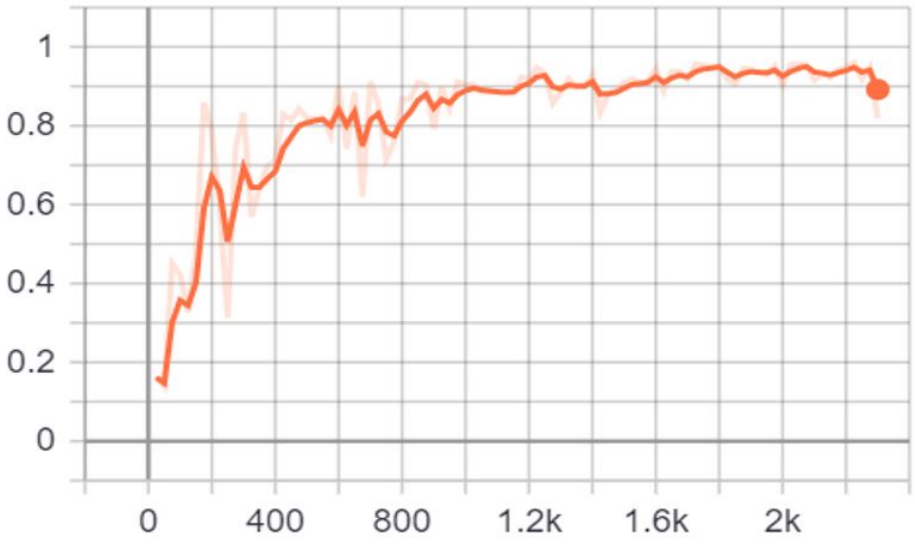


Figure 4.4: Train evaluation Score

4.1.7 **Comparison of the proposed methods with other methods**

As we have shown in the Table 4.2, the proposed algorithm was evaluated with a maximum Dice score of 89.04% and a mean Dice score of 84.92%. Maan et al. [37] proposed an automatic prostate segmentation algorithm from transversal T2-Weighted images based on 3D Active Appearance Models (AAM). The model performance was evaluated to the mean Dice Score of 78%. Zhang et al. [58] introduced an automatic pipeline to segment prostate in diffusion MRI, separating the prostate gland from the surrounding organs. The separation of the prostate gland was followed by a post-processing phase through active contours. This model proposed by Zhang was trained and evaluated on 25 clinical images of patients and achieved an overall accuracy with a mean Dice score of 84%. In the study done by Rundo et al. [29] proposed a new whole prostate gland segmentation algorithm based on U-Net Architecture by introducing squeeze-and Excitation (SE) blocks in the decoder and encoder of the U-Net architecture. During the study conducted by Khan et al. [9] to evaluate four encoder-decoder CNNs in the segmentation of the prostate gland in T2-Weighted MR images, he used the patch-wise DeepLabV+3 where the researcher achieved

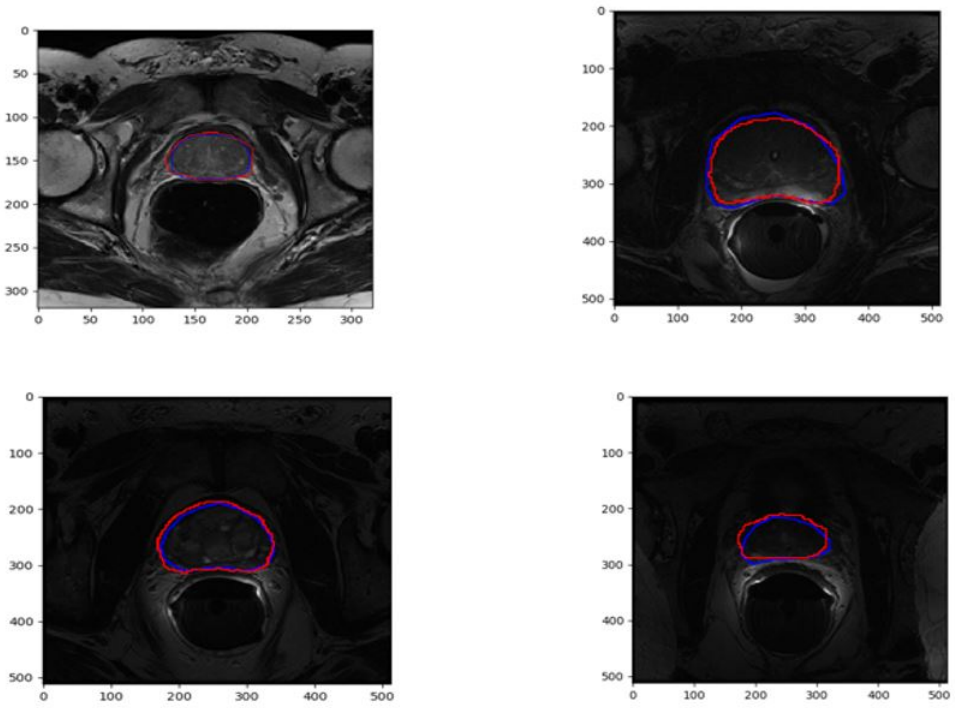


Figure 4.5: Prostate manually segmented, colored in (blue) and automatically segmented, colored in (red)

an accuracy with Dice score of 78.9% to segment the whole prostate gland. The Table 4.3 summarizes the comparison of our proposed segmentation algorithm with other published algorithms.

Table 4.3: Comparison of the proposed model with other published models

Model proposed by:	**Dice score (%)**
Khan et al.[10]	78.9 (%)
Rundo et al.[30]	76 (%)
Maan at al. [38]	78 (%)
Zhang et al.[59]	84 (%)
Our proposed model	**84.9 (%)**

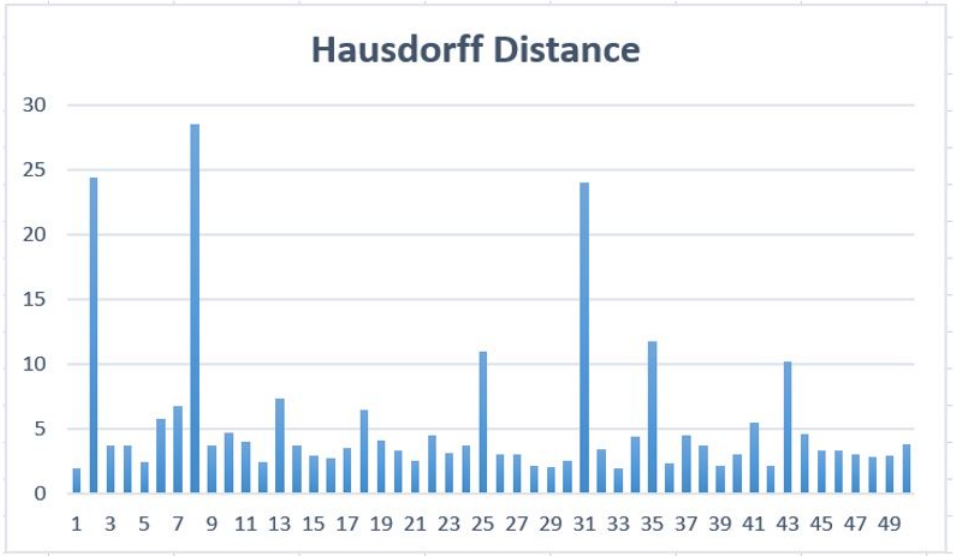


Figure 4.6: Graphical representation of Hausdorff distance

4.1.8 **Discussion of results**

Table4.3 shows the Dice Score for each of the 10 folds of the data set with the overall Mean Dice score averaged over all the presented 10 folds. As it is commonly accepted that a Dice score value greater than 0.7 (70%) represents a good agreement, the Mean Dsc for our segmentation whose value is 84.92% proves that the proposed model achieved good performance with promising results.

Note that, although our Mean DCS value (i.e 84.92%) is greater than values reported in the research conducted by Khan et al.[10] (Mean Dsc: 78.9%), Rundo et al.[30] (Mean Dsc: 76%), Maan et al.[38] (Mean Dsc: 78%) and Zhang et al.[59](Mean Dsc:84%); the proposed model is also lower compared to some other models, especially models builts upon the combination of different methods, for instance, Cheng et al.[4] proposed an enhanced HNN model for the automatic segmentation of prostate in MRI, which has achieved good results with a mean Dsc of 89.77%, Wang et al.[2] proposed an automatic segmentation of prostate magnetic resonance imaging using generative adversarial network which

has achieved promising results with a highest Dsc of 91,66%.

Chapter 5: Conclusion

5.1 Conclusion

Prostate cancer is the most common type of cancer often diagnosed in men all over the world. However, when diagnosed at its early stage, this PCa can be cured. The main goal of this research study was to propose a deep learning based automatic segmentation method of prostate from MRI, as a new tool that can provide assistance to radiologists and medical doctors for the prostate cancer diagnosis and treatment decisions.

In this work, the proposed method was developed by using U-Net architecture, was trained and evaluated on promise12 data set that consists of 50 T2-Weighted three dimensional magnetic resonance images (3D MRIs) made publically available by MICCAI Grand Challenge team for prostate segmentation purposes. The 10-fold cross-validation approach was used to train and evaluate the proposed method. The Dice Score Coefficient (DSC) and Hausdorff Distance (HD) were used as quantitatively evaluation metrics. The results showed the lowest DSC of 70.75 and the highest DSC of 89.04, then a mean DSC of 84.92. The computed HD is 5.3mm. The proposed method was also evaluated by comparing it with other segmentation methods in the literature where it showed to be outperforming than them.

As a conclusion, the proposed prostate segmentation method achieved good accuracy with promising results even better than other segmentation methods due to using U-Net model that has the ability to achieve best accuracy on medical image segmentation with relatively small data sets.

5.2 Future work

The purpose of the proposed method was to perform automatic segmentation of the prostate from MRI. A long with the research study, we have noticed that various directions can help for the future improvement of this study. Performing the same study on large data set and combining U-Net with other segmentation methods can make the used method more robust and achieve higher accuracy.

For the future work, we will implement the proposed method on patients' image data in Rwanda to assist radiologists for prostate cancer diagnosis and treatment planning that can include surgery decisions. We will also create an interface for this proposed method, whereby the users will be able to load images and perform the segmentation on the interface. This will help to save time taken by manual segmentation and reduce errors observed during manual segmentation. The implementation of automatic prostate segmentation method in Rwanda will also help the country to save money that are paid to external experts who help perform manual segmentation in the archaic system. In addition, the patients will no longer spend a whole day queuing at the hospital waiting for medical result because the system will fasten the process. Thus, saving citizens' life though reducing prostate cancer death cases in the country.

Bibliography

[1] G. Litjens, O. Debats, J. Barentsz, N. Karssemeijer, and H. Huisman, "Computer-aided detection of prostate cancer in mri," *IEEE transactions on medical imaging*, vol. 33, no. 5, pp. 1083–1092, 2014.

[2] A. Ziaei, "Advances in medical imaging technology for accurate detection of prostate cancer," in *Prostate Cancer*, IntechOpen, 2018.

[3] R. Cheng, H. R. Roth, N. S. Lay, L. Lu, B. Turkbey, W. Gandler, E. S. McCreedy, T. J. Pohida, P. A. Pinto, P. L. Choyke, *et al.*, "Automatic magnetic resonance prostate segmentation by deep learning with holistically nested networks," *Journal of medical imaging*, vol. 4, no. 4, p. 041302, 2017.

[4] S. Ghose, A. Oliver, R. Martí, X. Lladó, J. C. Vilanova, J. Freixenet, J. Mitra, D. Sidibé, and F. Meriaudeau, "A survey of prostate segmentation methodologies in ultrasound, magnetic resonance and computed tomography images," *Computer methods and programs in biomedicine*, vol. 108, no. 1, pp. 262–287, 2012.

[5] S. Ghose, *Robust image segmentation applied to magnetic resonance and ultrasound images of the prostate*. PhD thesis, Université de Bourgogne, 2012.

[6] P. Nieminen, "Applications of medical informatics and data analysis methods," 2020.

[7] G. Benurugo, E. Munyambaraga, G. Chironda, and E. Bisanukuri, "Awareness on prostate cancer and screening practices among men attending outpatient at a referral hospital in kigali, rwanda: A quantitative study," *International Journal of Africa Nursing Sciences*, vol. 13, p. 100241, 2020.

[8] F. Rubagumya, A. Costas-Chavarri, A. Manirakiza, G. Murenzi, F. Uwinkindi, C. Ntizimira, I. Rukundo, P. Mugenzi, B. Rugwizangoga, C. Shyirambere, *et al.*, "State of cancer control in rwanda: Past, present, and future opportunities," *JCO Global Oncology*, vol. 6, pp. 1171–1177, 2020.

[9] Z. Khan, N. Yahya, K. Alsaih, S. S. A. Ali, and F. Meriaudeau, "Evaluation of deep neural networks for semantic segmentation of prostate in t2w mri," *Sensors*, vol. 20, no. 11, p. 3183, 2020.

[10] R. Yamashita, M. Nishio, R. K. G. Do, and K. Togashi, "Convolutional neural networks: an overview and application in radiology," *Insights into imaging*, vol. 9, no. 4, pp. 611–629, 2018.

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

[12] K. Suzuki, "Survey of deep learning applications to medical image analysis," *Med Imaging Technol*, vol. 35, no. 4, pp. 212–226, 2017.

[13] H. Seo, M. Badiei Khuzani, V. Vasudevan, C. Huang, H. Ren, R. Xiao, X. Jia, and L. Xing, "Machine learning techniques for biomedical image segmentation: An overview of technical aspects and introduction to state-of-art applications," *Medical physics*, vol. 47, no. 5, pp. e148–e167, 2020.

[14] S. Yuheng and Y. Hao, "Image segmentation algorithms overview," *arXiv preprint arXiv:1707.02051*, 2017.

[15] D. D. Patil and S. G. Deore, "Medical image segmentation: a review," *International Journal of Computer Science and Mobile Computing*, vol. 2, no. 1, pp. 22–27, 2013.

[16] T. Rohlfing, R. Brandt, R. Menzel, D. B. Russakoff, and C. R. Maurer, "Quo vadis, atlas-based segmentation?," in *Handbook of biomedical image analysis*, pp. 435–486, Springer, 2005.

[17] H. Kalinic, "Atlas-based image segmentation: A survey," *Department of Electronic Systems and Information Processing, Universiy of Zagreb*, pp. 1–7, 2008.

[18] R. Cuocolo, M. B. Cipullo, A. Stanzione, L. Ugga, V. Romeo, L. Radice, A. Brunetti, and M. Imbriaco, "Machine learning applications in prostate cancer magnetic resonance imaging," *European radiology experimental*, vol. 3, no. 1, pp. 1–8, 2019.

[19] M. H. Hesamian, W. Jia, X. He, and P. Kennedy, "Deep learning techniques for medical image segmentation: achievements and challenges," *Journal of digital imaging*, vol. 32, no. 4, pp. 582–596, 2019.

[20] W. Zhang, R. Li, H. Deng, L. Wang, W. Lin, S. Ji, and D. Shen, "Deep convolutional neural networks for multi-modality isointense infant brain image segmentation," *NeuroImage*, vol. 108, pp. 214–224, 2015.

[21] X. Zhou, R. Takayama, S. Wang, T. Hara, and H. Fujita, "Deep learning of the sectional appearances of 3d ct images for anatomical structure segmentation based on an fcn voting method," *Medical physics*, vol. 44, no. 10, pp. 5221–5233, 2017.

[22] M. C. Chen, R. L. Ball, L. Yang, N. Moradzadeh, B. E. Chapman, D. B. Larson, C. P. Langlotz, T. J. Amrhein, and M. P. Lungren, "Deep learning to classify radiology free-text reports," *Radiology*, vol. 286, no. 3, pp. 845–852, 2018.

[23] G. Zeng and G. Zheng, "Multi-stream 3d fcn with multi-scale deep supervision for multi-modality isointense infant brain mr image segmentation," in *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, pp. 136–140, IEEE, 2018.

[24] S. Hamidian, B. Sahiner, N. Petrick, and A. Pezeshk, "3d convolutional neural network for automatic detection of lung nodules in chest ct," in *Medical Imaging 2017: Computer-Aided Diagnosis*, vol. 10134, p. 1013409, International Society for Optics and Photonics, 2017.

[25] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *European conference on computer vision*, pp. 818–833, Springer, 2014.

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

[27] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

[28] L. Yu, H. Chen, Q. Dou, J. Qin, and P.-A. Heng, "Automated melanoma recognition in dermoscopy images via very deep residual networks," *IEEE transactions on medical imaging*, vol. 36, no. 4, pp. 994–1004, 2016.

[29] L. Rundo, C. Han, Y. Nagano, J. Zhang, R. Hataya, C. Militello, A. Tangherloni, M. S. Nobile, C. Ferretti, D. Besozzi, *et al.*, "Use-net: Incorporating squeeze-and-excitation blocks into u-net for prostate zonal segmentation of multi-institutional mri datasets," *Neurocomputing*, vol. 365, pp. 31–43, 2019.

[30] C. Bong, C. Liew, and H. Lam, "Ground-glass opacity nodules detection and segmentation using the snake model," in *Bio-Inspired Computation and Applications in Image Processing*, pp. 87–104, Elsevier, 2016.

[31] R. Hemalatha, T. Thamizhvani, A. J. A. Dhivya, J. E. Joseph, B. Babu, and R. Chandrasekaran, "Active contour based segmentation techniques for medical image analysis," *Medical and Biological Image Analysis*, vol. 4, p. 17, 2018.

[32] P. Lenkiewicz, M. Pereira, M. M. Freire, and J. Fernandes, "The whole mesh deformation model: a fast image segmentation method suitable for effective parallelization," *EURASIP Journal on Advances in Signal Processing*, vol. 2013, no. 1, pp. 1–17, 2013.

[33] T. F. Cootes, A. Hill, C. J. Taylor, and J. Haslam, "Use of active shape models for locating structures in medical images," *Image and vision computing*, vol. 12, no. 6, pp. 355–365, 1994.

[34] S. Osher and J. A. Sethian, "Fronts propagating with curvature-dependent speed: Algorithms based on hamilton-jacobi formulations," *Journal of computational physics*, vol. 79, no. 1, pp. 12–49, 1988.

[35] N. N. Kachouie, P. Fieguth, and S. Rahnamayan, "An elliptical level set method for automatic trus prostate image segmentation," in *2006 IEEE International Symposium on Signal Processing and Information Technology*, pp. 191–196, IEEE, 2006.

[36] O. Acosta, J. Dowling, G. Cazoulat, A. Simon, O. Salvado, R. De Crevoisier, and P. Haigron, "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 *International Workshop on Prostate Cancer Imaging*, pp. 42–51, Springer, 2010.

[37] B. Maan and F. van der Heijden, "Prostate mr image segmentation using 3d active appearance models," *MICCAI Grand Challenge: Prostate MR Image Segmentation*, vol. 2012, 2012.

[38] M. Habes, T. Schiller, C. Rosenberg, M. Burchardt, and W. Hoffmann, "Automated prostate segmentation in whole-body mri scans for epidemiological studies.," *Physics in medicine and biology*, vol. 58 17, pp. 5899–915, 2013.

[39] Z. Tian, L. Liu, Z. Zhang, and B. Fei, "Psnet: prostate segmentation on mri based on a convolutional neural network," *Journal of Medical Imaging*, vol. 5, no. 2, p. 021208, 2018.

[40] S. Liao, Y. Gao, A. Oto, and D. Shen, "Representation learning: a unified deep learning framework for automatic prostate mr segmentation," in *International Confer-*

ence on Medical image computing and computer-assisted intervention, pp. 254–261, Springer, 2013.

[41] Y. Guo, Y. Gao, and D. Shen, "Deformable mr prostate segmentation via deep feature learning and sparse patch matching," *IEEE transactions on medical imaging*, vol. 35, no. 4, pp. 1077–1089, 2015.

[42] D. Gillespie, C. Kendrick, I. Boon, C. Boon, T. Rattay, and M. H. Yap, "Deep learning in magnetic resonance prostate segmentation: A review and a new perspective," *arXiv preprint arXiv:2011.07795*, 2020.

[43] D. B. Plewes and W. Kucharczyk, "Physics of mri: a primer," *Journal of magnetic resonance imaging*, vol. 35, no. 5, pp. 1038–1054, 2012.

[44] S. Currie, N. Hoggard, I. J. Craven, M. Hadjivassiliou, and I. D. Wilkinson, "Understanding mri: basic mr physics for physicians," *Postgraduate medical journal*, vol. 89, no. 1050, pp. 209–223, 2013.

[45] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, *et al.*, "Imagenet large scale visual recognition challenge," *International journal of computer vision*, vol. 115, no. 3, pp. 211–252, 2015.

[46] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in neural information processing systems*, vol. 25, pp. 1097–1105, 2012.

[47] V. Gulshan, L. Peng, M. Coram, M. C. Stumpe, D. Wu, A. Narayanaswamy, S. Venugopalan, K. Widner, T. Madams, J. Cuadros, *et al.*, "Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs," *Jama*, vol. 316, no. 22, pp. 2402–2410, 2016.

[48] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, "Dermatologist-level classification of skin cancer with deep neural networks," *nature*, vol. 542, no. 7639, pp. 115–118, 2017.

[49] P. Lakhani and B. Sundaram, "Deep learning at chest radiography: automated classification of pulmonary tuberculosis by using convolutional neural networks," *Radiology*, vol. 284, no. 2, pp. 574–582, 2017.

[50] K. Yasaka, H. Akai, O. Abe, and S. Kiryu, "Deep learning with convolutional neural network for differentiation of liver masses at dynamic contrast-enhanced ct: a preliminary study," *Radiology*, vol. 286, no. 3, pp. 887–896, 2018.

[51] P. F. Christ, M. E. A. Elshaer, F. Ettlinger, S. Tatavarty, M. Bickel, P. Bilic, M. Rempfler, M. Armbruster, F. Hofmann, M. D'Anastasi, *et al.*, "Automatic liver and lesion segmentation in ct using cascaded fully convolutional neural networks and 3d conditional random fields," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 415–423, Springer, 2016.

[52] K. H. Kim, S. H. Choi, and S.-H. Park, "Improving arterial spin labeling by using deep learning," *Radiology*, vol. 287, no. 2, pp. 658–666, 2018.

[53] F. Liu, H. Jang, R. Kijowski, T. Bradshaw, and A. B. McMillan, "Deep learning mr imaging–based attenuation correction for pet/mr imaging," *Radiology*, vol. 286, no. 2, pp. 676–684, 2018.

[54] M. I. Razzak, S. Naz, and A. Zaib, "Deep learning for medical image processing: Overview, challenges and the future," *Classification in BioApps*, pp. 323–350, 2018.

[55] A. L. Maas, A. Y. Hannun, and A. Y. Ng, "Rectifier nonlinearities improve neural network acoustic models," in *Proc. icml*, vol. 30, p. 3, Citeseer, 2013.

[56] S. Jadon, "A survey of loss functions for semantic segmentation," *2020 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB)*, Oct 2020.

[57] D. P. Huttenlocher, G. A. Klanderman, and W. J. Rucklidge, "Comparing images using the hausdorff distance," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 15, no. 9, pp. 850–863, 1993.

[58] J. Zhang, S. Baig, A. Wong, M. A. Haider, and F. Khalvati, "Segmentation of prostate in diffusion mr images via clustering," in *International Conference Image Analysis and Recognition*, pp. 471–478, Springer, 2017.

Acknowledgements

"First and foremost, I am grateful and thankful to the Almighty LORD for giving me a healthy life, covering me with his favor, wisdom and courage throughout my thesis work".

Many thanks go to MOFCOM for granting me a full scholarship covering every needs for my Master Degree program.

I also would like to thank the whole Nankai University team, especially the College of Software and International Students Office for all kinds of support provided for the success of my studies.

I would like to express my special thanks to my supervisor Doctor Ling Ma and my colleague Lei Tao who were always there for a considerable guidance and help throughout my whole work. Sincerely, I would not have made it without their support.

My sincere thanks go to my lovely husband Mutijima Sabin and our daughter Ineza Mutijima Pison for being morally and spiritually supportive and so caring in this period of my studies. They are behind my success for real.

I cannot forget to thank my precious biological and spiritual parents, siblings and friends for being so supportive and encouraging along this whole journey of my studies.

Finally, I would like to thank my fellow classmates for being helpful, inspiring and stress cure to me for the good two years we had spent together.

Curriculum Vitae

• Basic Information

Name: MUHAYIMANA ODETTE
Gender：Female
Date of Birth：12/30/1989
Nationality：Rwanda
Research Area：Medical Image Segmentation
Supervisor：A/Prof. Ling Ma

• Education Background

2019.09 –2021.06

Nankai University Software Engineering Master

Sample：

2010.10–2013.08

National University of Rwanda Computer Science and Systems Bachelor

• Publications

None