

Musculoskeletal Tissue Segmentation in CT Scan Images

by

Raushan Ranjan

202011009

A Thesis Submitted in Partial Fulfilment of the Requirements for the Degree of

MASTER OF TECHNOLOGY

in

INFORMATION AND COMMUNICATION TECHNOLOGY

to

DHIRUBHAI AMBANI INSTITUTE OF INFORMATION AND COMMUNICATION TECHNOLOGY



August, 2022

Declaration

I hereby declare that

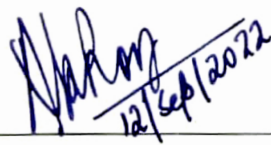
- i) the thesis comprises of my original work towards the degree of Master of Technology in Information and Communication Technology at Dhirubhai Ambani Institute of Information and Communication Technology and has not been submitted elsewhere for a degree,
- ii) due acknowledgment has been made in the text to all the reference material used.



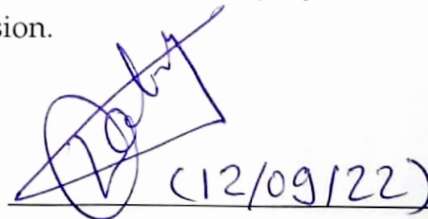
Raushan Ranjan

Certificate

This is to certify that the thesis work entitled Musculoskeletal has been carried out by Raushan Ranjan for the degree of Master of Technology in Information and Communication Technology at *Dhirubhai Ambani Institute of Information and Communication Technology* under our supervision.



Dr. Anil Roy
Thesis Supervisor



Dr. Bakul Gohel
Thesis Co-Supervisor

Acknowledgments

This thesis is a testimony of the support I have gotten from all the people around me. I want to extend my sincerest gratitude to my supervisor Dr. Anil Roy and my co-supervisor, Dr. Bakul Gohel. This thesis would not have been possible without their motivation and unmatched patience. They always encouraged me to explore new and relevant ideas and provided me with the most acceptable directions. This work would not have seen the direction it is currently at without the generous support of Dr. Manish Shah and Shah Hospital, Ahmedabad. His support for providing the dataset and his valuable inputs throughout is much appreciated. I would also like to thank other faculty members who taught me during my masters and the institute, DA-IICT, as a whole for providing me an environment and enormous learning opportunities. Also, I would like to thank all my peers who have supported me throughout my work and studies. A special thanks to my family for their unconditional love and support throughout this program and in general.

Contents

Abstract	vi
List of Principal Symbols and Acronyms	vi
List of Tables	vii
List of Figures	viii
1 Introduction	1
2 Background and related work	5
2.1 Deep Learning	6
2.2 Supervised learning	7
2.3 Convolutional neural networks	7
2.4 Unsupervised and supervised methods	8
2.4.1 Clustering methods	8
2.4.2 Supervised methods	9
2.5 Deep learning approaches	10
2.5.1 CNNs for Medical Imaging Segmentation	10
2.5.2 U-Net	11
2.5.3 Some other Relevant studies	13
2.6 Gray level features methods	15
2.6.1 Fixed threshold	15
2.6.2 Otsu Threshold	16
2.6.3 Region growing and Split and Merge algorithms	16
2.7 Bone segmentation	17
2.8 Related work	18
3 Fundamentals	21
3.1 Computed Tomography images	21
3.2 Segmentation	21

3.3	Evaluation Parameter	22
3.3.1	Precision	22
3.3.2	Recall	22
3.3.3	F1 Score	23
4	Data Set	24
4.1	Data annotation	24
4.1.1	3D Slicer	24
4.1.2	MRicro	27
4.2	Data augmentation	27
5	Landmark detection	29
5.1	UNet Model	29
5.2	Results and Analysis	29
5.2.1	Femur head center	30
5.2.2	Distal Femur	30
6	Segmentation	32
6.1	Model I : Fixed Threshold-based Segmentation	32
6.2	Model II : Otsu’s Segmentation	34
6.3	Model III : 2D UNet	36
6.3.1	Tissue segmentation	37
6.3.2	Femoral Head	38
6.3.3	Distal Femur	38
6.4	Model IV: 3D UNet	39
6.4.1	Femur Head	42
6.4.2	Distal Femur	43
6.5	Results and Analysis	46
7	Conclusion	50
	References	52

Abstract

Osteoarthritis (OA) is a common cause of painful knee joints among senior citizens. Osteoarthritis restricts the range of movement of the body. Senior citizens are particularly vulnerable, as limited mobility reduces their quality of life, leading to depression and isolation. Proper working of joints is an essential health indicator. In early-stage osteoarthritis can be treated mainly based on medication and physiotherapy. When the conventional treatment stops giving relief to patients, total knee arthroplasty (TKA) is the next option available to the patient. The success of TKA depends on proper sizing and positioning of implant, pre-operative and intra-operatively decisions. 3D segmented bone information from the patient's Magnetic Resonance Imaging (MRI) or Computed Tomography (CT) can help doctors make better pre-operative decisions on implant size and positioning plan.

CT scan image is the input for this pre-operative stage of TKA. But the challenge here is that the CT images comes as in slices and it becomes a problem to identify exact location and their physical condition of the bones converging at the knee. The main bones are femur, tibia and fibula. It would become quite helpful for the surgeon if these bones and only bones are visibly available in a 3D view of the given CT scan images. For this segmentation of these bones is desired and if the segmentation be done in an automated way, it would be very less resource and less time consuming. This thesis presents an automated framework for segmentation of the distal femur and femoral head from 3D CT scans. The first stage of this framework consists of 2D U-Net architecture for landmark prediction of the distal femur and femoral head in the original 3D CT scans. The next stage of the framework is to perform local 3D segmentation around the landmark. We also demonstrated and performed tissue and femur bone segmentation using various methods such as fixed threshold segmentation, UNet segmentation, and Ostu's based segmentation. Further focusing on the femur bone of CT scans of the knee joint, we are using deep learning method the UNet as segmentation model.

The test set of CT scans is used to evaluate the methods. The experiments of the results indicate that the method chosen by us, of landmarking and further

segmentation is effectively and thoroughly automatically able to segment femur bone and distal femur from the whole leg CT scans.

List of Tables

3.1	Confusion matrix	22
5.1	Landmark prediction accuracy of Femur head (in mm).	30
5.2	Landmark prediction accuracy of distal femur(in mm).	30
5.3	Landmark prediction accuracy of distal femur in 3D plane and z-axis	30
6.1	Performance matrix score of fixed threshold segmentation	34
6.2	Performance analysis for all bone based on Ostu’s segmentation.	35
6.3	Performance matrix for Knee Joint based on U-Net deep learning model.	37
6.4	Performance matrix for Femur Head based on 2D UNet deep learning model.	39
6.5	Performance analysis for Femur at knee Joint based on UNet deep learning model.	39
6.6	Performance analysis of 3D UNet on Femoral head.	43
6.7	Performance analysis of 3D UNet on Distal Femur.	45
6.8	Performance measure of all our approaches.	46

List of Figures

1.1	Full CT scan slices of size 512 x 512 pixels.	4
2.1	An illustration of how neurons in a convolutional layer are structured in three dimensions.	8
2.2	A schematic of our 2D u-net model.	12
2.3	A schematic of a residual block, as proposed in [28]	14
4.1	A view of 3D slicer application	25
4.2	Example of CT slices in axial, coronal and sagittal format.	26
4.3	3D mask for ground truth.	27
5.1	A schematic of our proposed 2D u-net model.	31
6.1	A flowchart showing the workflow of fixed threshold segmentation method.	33
6.2	3D view of a prediction of bone by fixed threshold value method.	34
6.3	3D view of a prediction of bones using local Otsu’s threshold method.	35
6.4	A schematic of our 2D u-net model.	36
6.5	3D view of a prediction of knee joint obtained by 2D UNet model.	37
6.6	3D view of a prediction of femur head and femur shaft from 2D UNet	38
6.7	3D view of ground truth at hip joint	38
6.8	Segmented Femur Condyle	39
6.9	A flowchart showing the workflow of our proposed method	41
6.10	A schematic view of our proposed 3D U-Net model	42
6.11	Segmentation result for femur head on a sample test patient.	42
6.12	Slice representation of femur head	43
6.13	Slice representation at distal femur.	44
6.14	Segmentation result for femur head on a sample test patient.	45
6.15	Segmentation result of Model [39] on axial slice of Femur head	47
6.16	Segmentation result of 3D UNet model on axial slice of Femur head	48
6.17	Segmentation result of 3D UNet model on axial slice of Distal femur	49

CHAPTER 1

Introduction

A consequence of an aging society is that the older population is prone to chronic diseases. This may lead to increased demands for qualified medical practitioners in the upcoming years. The current healthcare system must be restructured from personalized care to a more decentralized and automated one that emphasizes more accuracy and less time for treatment. Osteoarthritis (OA) is a disease that affects people over the age of 60 years. The degradation of articular cartilage in joints, notably in the knee and hip, characterizes this disorder. When the cartilage wears away, bones scrape against each other, producing stiffness, pain, and discomfort in joint movement. It is the most frequent joint illness and the second most common rheumatologic condition in India, with a prevalence of 22% to 39% of the population. [53]. Women are more likely than men to suffer from it. Over the age of 65 years, nearly 45% of women show symptoms, while 70% amongst those over 65 show the existence of osteoarthritis in the radiology tests [53]. Among the treatment options currently available, total hip replacement (THR) and total knee arthroplasty (TKA) surgery are the most effective.

Different technologies are used to produce high-resolution three-dimensional models of tissues and bones from patient-specific medical imaging, such as Computed Tomography (CT). Medical images provide a way to examine, diagnose, monitor, and eventually treat multiple medical conditions wherein the skillful physicians do not require any instrument insertion in the body. Tissue segmentation is the technique of segregating an image into sections that correspond to distinct types of tissue. These classes are defined in biology as specific tissue types, whole organs, or sub-regions of organs (e.g., bone or muscle segments). Common image modalities which are used in recording image data are X-Ray, computed tomography (CT) [31] and magnetic resonance imaging (MRI). It would be of great help to the medical practitioners if there were a definitive trustworthy system that could segment the input medical image for a speedy analysis and diagnosis from the input medical imagery. The lack of trust in the system hinders accepting

such solutions in the practical world, which remains the focal point of our work. We perform the segmentation and further analyze various models to perform tissue segmentation and further perform bone segmentation, focusing on the femur bone, leveraging the advantages of the deep learning technology.

Deep learning is a commonly used terminology nowadays. From face recognition systems to self-driving cars, profit/costs forecasts, or automatic detection of problems. It appears that deep learning-based autonomous systems will have the capability of transforming the world in the coming years, owing to the development of convolutional neural networks (CNN). CNN's are one of the kind of a deep neural network. CNN recently obtained exceptional performance in tasks involving natural language processing, speech recognition and object recognition [25]. Moreover, in various studies, it is seen that U-Net has been effectively used to analyze medical images and accomplish tasks such as object detection, image classification, segmentation and registration. U-Net is one such convolutional neural network developed for biomedical image segmentation at the Computer Science Department of the University of Freiburg [58]. This study looked into using a convolutional neural network (CNN) to automatically segment bones in CT scans. This is a difficult task, and it would be quite beneficial to assist doctors in their difficult task of making decisions after studying those images. Segmenting an image into its constituent regions or separate objects is known as image segmentation. [24] Segmentation gives us a representation that is not so hard to study further and analyze. A label is assigned to every pixel in 2D and voxel in 3D images during segmentation. Thus, it is essentially a voxel classification task. Voxels with similar characteristics or that belong to the same object have the same label and are thus grouped in the same class. Segmentation has many real-world applications in medical imaging, like locating tumors and other diseases, measuring tissue volumes, surgery planning, and virtual surgery simulation. More traditional computer vision and machine learning algorithms for image segmentation have been developed and successfully executed. There is no general algorithm for semi-automated or autonomous segmentation of medical images, which is always chosen. The choice of the algorithm depends on several things like which part of the body is to be studied, the type of the medical image, and the goal of the study. Even in many circumstances, manual segmentation is still the gold standard despite being a time-consuming and arduous process. Various types of noise caused by the acquisition system and the partial volume effect are the facets that should be considered both in the automatic and manual segmentation methods.

Bone segmentation in CT scans is a critical step in assisting surgeons with vari-

ous medical procedures. The majority of the bones may be easily detected visually in CT scans, while perfect automated segmentation remains a problematic issue. CT scans frequently have low signal-to-noise ratio, a poor spatial resolution, and a variety of artifacts, making it a time-consuming and challenging process.

Image segmentation algorithms are divided into the following three categories:

- Manual
- Semi-automated
- Fully automated

Manual segmentation is perhaps the most consistent and precise method; but, it requires subject knowledge, a significant amount of effort, and is time-consuming. The results achieved from the manual segmentation techniques are often heavily influenced by the expert in charge's analysis. On the other hand, segmentation algorithms that are semi-automatic need some manual initialization, like regions of interest for the structures to be segmented or even some parameters. Finally, the fully automated algorithms must retrieve the elements of interest independently in an autonomous fashion, with no user input. Increased GPU computational capacity is one of the key factors that has allowed deep learning-based approaches to expand so quickly in recent years. Deep learning models in computer vision achieved great results thanks to the latest GPUs, which significantly shortened the training time of complex networks.

In this thesis we present a framework based on U-Net for automated landmark detection of the femur head and distal femur from CT scan modality. Based on these landmarks, we present an automated U-Net-based segmentation method to segment bones and femur bone using very few computational resources. In particular, the landmark of femoral bone is detected in the CT scans, and the next local segmentation of femoral bone from other bones and tissues of the body is performed.

Our work presented in this thesis can be summarized as follows:

- Implementing and understanding the efficiency of different existing, standard segmentation techniques such as fixed threshold-based, Otsu segmentation and U-Net based segmentation for tissue segmentation.
- Studying their limitations.
- Implementing landmark prediction and then merging with U-Net.

- Eliminating the need for manual cropping at the pre-processing stage for finding regions of interest as one can see in Fig. (1.1).
- Performing the data augmentation to increase the training data size for better segmentation.
- Implementing U-net for femur bone segmentation.

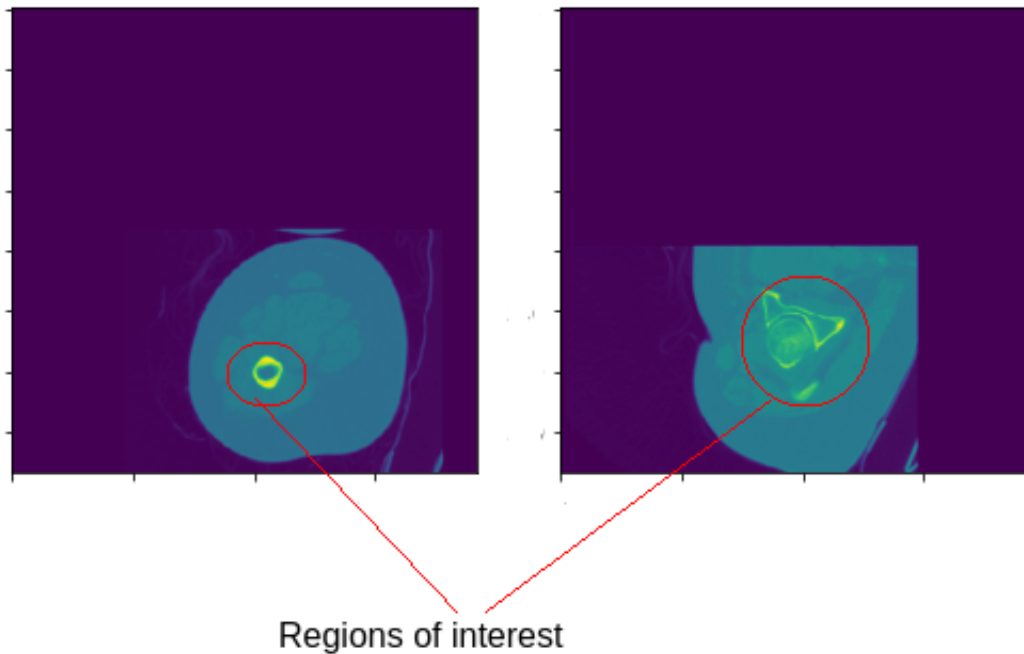


Figure 1.1: Full CT scan slices of size 512 x 512 pixels.

The second chapter of the thesis gives background and previously reported works related to Computed Tomography based segmentation techniques with specific focus on their use in medical images. The third chapter of the thesis consists of backgrounds and fundamentals. The fourth chapter is all about the dataset for this thesis. The fifth chapter is about landmark detection and finding the region of interest. Some of the most frequent segmentation strategies are briefly discussed and analyzed in the sixth chapter, and then more focus is on the deep learning approaches. The final chapter is about the work's conclusion and possible future advancements.

CHAPTER 2

Background and related work

Segmentation of image consists of splitting a picture into significant regions semi-automatically or automatically . The aim of segmentation is to locate the voxels in a region of object interior or its border. In the process semantic segmentation each voxel is given a label, and the resultant image is in which voxels with the very same label have similar qualities or properties [43], [14], [52], [54], [30] [8] . As a result, image segmentation gives us a more generalised representation of image data that holds a lot of meaning and is a critical path in effectively evaluating and diagnosing medical images.

Segmentation retrieves critical information for a quantitative examination of clinical parameters in medical image processing. Segmentation of tissue, organs, bones, or other small portion is an important step in the great majority of computer-aided detection (CAD) systems. Different ways to this problem have been offered in the literature, and in recent years, papers integrating deep learning and medical imaging have been the most common topic [40]. However, it remains a difficult task due to a variety of problems such as image noise, low contrast, artefacts, inhomogeneities, and others. An automatic segmentation is a method with least dependency on the operator and also have features like accuracy, reliability, repeatability and robustness [61].

The most successful medical image segmentation methods are simply categorized into techniques based on supervised and unsupervised, texture features and grey level characteristics methods. These methods are briefly presented in the upcoming sections. Some recent and major publications on this topic are referenced and discussed at the end of chapter.

This contains the U-Net [58], a neural network for medical image segmentation that underpins the Deep-learning method used in this thesis.

2.1 Deep Learning

Artificial intelligence (AI) is among the most widely debated topics in today's society, among both the general public and academics. AI is defined as computers' ability to accomplish tasks that would ordinarily need human intelligence. There are numerous active research subjects on automating some form of function right now. Understanding and recognising self-driving cars, sights and sounds, strategic game systems and digesting complex information and diagnostics in medicine are just a few examples. Furthermore, deep learning is continually being applied to new situations, and its full potential has yet to be fully explored. Initially, AI systems were capable of successfully accomplishing tasks that were considered difficult for humans. AI could be described by a set of mathematical and logical rules, such as proving logical theorems or solving algebra problems. Later on, solving issues that people accomplish effortlessly, such as recognising a given sound or visual pattern, became much more difficult. Making machines learn from their experiences, allowing them to solve problems using what they've already learned, is a prominent strategy right now. They learn about the world by developing a hierarchy of concepts, with each concept represented as a collection of smaller ones. As a result, basic, simpler elements are used to understand abstract and more complicated representations. In this approach, machines may develop complex models using a variety of techniques without the need for anybody to programme what the model should learn. If its hierarchy of concepts were displayed, it would have multiple levels, which is one of the main reason why this method is known as deep learning [25]. Deep learning is a subset of machine learning that comprises all approaches based on data representation learning. The data format that is fed to commonly used machine learning techniques, on the other side, has an impact. A feature is any knowledge included in those representations, and hand-crafted characteristics are those developed by humans. Choosing the right features for a machine learning system is a major step in enhancing its performance. Representation learning, on the other side, is a technique for learning both the representation and the mapping between representation to output. This method eliminates the need for handcrafted features and frequently produces better results. In few cases, obtaining a proper representation selecting proper feature set can be equivalently difficult to solving the actual problem itself. For all of these reasons, deep learning has its roots from many levels of data qualities or concepts, with the higher level ones built on the lower level ones. This renders flexibility and a great power that allows to solve complex real-world

problems. Deep learning networks are made up of multiple layers that are stacked on top of each other. The signal is processed by a specific unit at each layer based on parameters learned during training, and then it is passed to the next layer. The layers between the input and output in a network are known as hidden layers. Various linear and non-linear transformations are applied on these hidden layers. Because of the enormous number of hidden layers, these networks are referred to as deep neural networks. In general, deep learning methods can be partitioned into supervised and unsupervised learning: in supervised learning models, models learn features from labeled data, whereas in unsupervised learning models aim at a pattern analysis.

We have primarily focused on the supervised learning method in this thesis.

2.2 Supervised learning

A supervised learning model's dataset is made up of training data units, each of which comprises input features and the desired output, which are also known as label, target, or mask. The algorithm itself analyze the data, extracts relevant important features from it and provide a function which can predict new unknown data which is also known as test data. While adopting a supervised learning method to solve a problem, few pointers need to be kept in the mind [25]. First and foremost, the training dataset should reflect the distribution of datasets in the real world. Second, the data input sample should be carefully selected. It shouldn't have a lot of information in it, but just enough for the algorithm to figure out an appropriate function. Next, a suitable function is picked and its parameter are manipulated to adjust. Finally, the algorithm's performance should be evaluated using a testing dataset that the algorithm has never seen before. Regression and classification are the two basic areas of issues that supervised learning systems address. A classification model predicts discrete values or data divided into groups, whereas a regression model predicts continuous values. Support vector machines, artificial neural networks, and decision trees are examples of typical algorithms used for both purpose.

2.3 Convolutional neural networks

Convolutional neural networks (CNNs) are the most advanced deep neural network technology for image processing.

CNNs are inspired by the biological visual cortex [64], [6], ANNs aim to mimic

biological neuron activity in a comparable way. CNNs are comparable to traditional neural networks, but they assume that the input data is organised in a grid-like structure. Because images are multidimensional grids by nature, the convolutional layer, in which neurons are organised in three dimensions and a three-dimensional volume input is processed to produce another three-dimensional volume, is a key building component of CNNs. In Fig. (2.1) a scheme of this arrangement is shown.

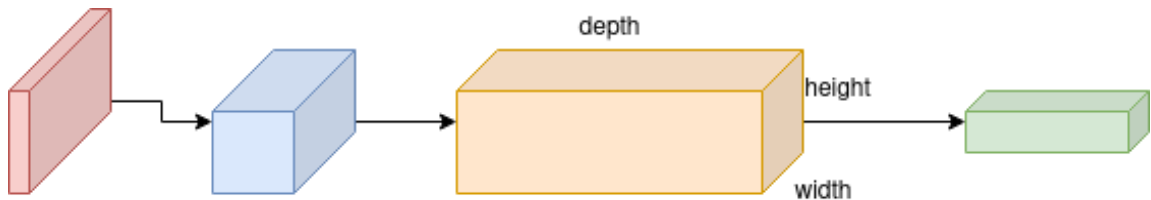


Figure 2.1: An illustration of how neurons in a convolutional layer are structured in three dimensions.

In order to optimise learning, CNN design takes advantage of this characteristic. Because of their large size and the loss of local information, vectorizing multidimensional data such as images is not appropriate. Instead, processing sub regions to derive feature maps is favoured [25], which are vital features of an image such as edges. Instead of utilising element-wise multiplication, the image is convolved with a kernel or sliding window, which has the added benefit of decreasing the network’s complexity through weight sharing. However, instead of convolution, the existing libraries for CNN implementation employ the cross-correlation operator [55], which fulfills the same motive by using symmetric kernels and omitting the “flip” or time reversal that is associated with convolution. The key issue becomes determining the ideal kernel weights that minimise the error function [25].

2.4 Unsupervised and supervised methods

In the following sub sections, few examples from unsupervised and supervised algorithms for segmentation are described.

2.4.1 Clustering methods

Unsupervised methods or clustering divide the data initially in a way that different objects are grouped together on the basis of similarity. They do not even require any type of training datasets with minimal involvement of operator. These

algorithms have the advantage of being able to generalise effectively across a variety of data sets and requiring a minimal amount of time. However, because they are sensitive to noise, they do not always produce the best results. For cluster analysis, K-means is a method that is used frequently. It is mainly composed of a somewhat simple iterative algorithm. k centroids are –initially randomly picked positions in the input space, and then the following steps is followed:

- The Euclidean distance from the centroids is calculated for all objects.
- Every object is allocated to the centroid that is closest to it.
- The centroids are recalculated based on the average value of each cluster.

The prior steps are then reiterated until the components of each group remain unchanged. On the one hand, this approach is simple to use, but it has a few disadvantages, such as the number of clusters which must be selected manually by the operator, as well as its outliers, sensitivity to noise and initial values. K-means clustering algorithms have already been widely used in the medical image analysis sector, for example, in the segmentation of the brain and tumours of brain from MRI scans [34],[63]. Fuzzy C-means is another unsupervised technique that has performed well on medical images. This method is very alike to k-means, except that the each point is assigned a weight to each cluster. As a result, a soft segmentation is created, which means that a point does not fit in to just one cluster and that overlapping areas are possible. This last point is particularly important in medical imaging where tissues are clearly overlapping. Fuzzy C-means has been used successfully in medical images, particularly for segmentation on MRI scans [67]. On MR brain images, a study compared k-means and Fuzzy C-means [45], indicating that one algorithm is not necessarily superior than the other, and that their relative effectiveness varies depending on the specific task.

2.4.2 Supervised methods

For image segmentation, several supervised algorithms have been proposed, often known as classifiers. These are generally based on supervised learning (see Section 2.2), which requires the of a training dataset consisting of objects and their target labels. Segmenting the training dataset manually requires time and effort, and it is a process that is highly dependent on human knowledge and actions. These methods, on the other hand, may adapt to a wide range of tasks and handle complex problems.

The supervised approach k-nearest-neighbour (k-nn) is a very frequently used technique. This technique requires a large amount of labelled data and is classified as a non parametric classifier since it does not consider the statistical structure of the data. Each pixel is classified in the k-nn technique into the class to which the maximum of his k closest neighbors belong, where k is a parameter chosen manually by the operator. The choice of k, the distance measurement, and the mechanism of counting votes all have an effect on the approach's performance.

Another commonly used method is the maximum likelihood methodology, which assumes the data is comprised of statistically independent and identical distribution samples which usually belong to Gaussian distribution. The likelihood principle is then used to estimate the distribution's parameters.

The supervised artificial neural network-based techniques are also included in the supervised methods category. For medical image segmentation, feed-forward and feedback networks have been widely used, feedback is inspired by the human visual cortex mechanism. These convolutional neural network feedback algorithms are based on neuron selection and visual information from neurons. [40]. Several studies using deep learning for medical image segmentation have been conducted in recent years, and the most relevant ones are discussed in the following sections.

2.5 Deep learning approaches

Segmentation approaches that have been based on Deep learning have been successfully applied to a variety of image segmentation in the field of medicine in many challenges in recent years. For instance, skull segmentation [49], brain areas [47], [22], calcifications in coronary arteries, organs, and different substructures [40]. Although all of these research work use CNNs, each one has a different architecture. The transformation of conventional classification CNNs into segmentation networks is briefly explained in the next subsection.

2.5.1 CNNs for Medical Imaging Segmentation

General CNNs are the primary form of classification networks [25]. They take an image as input and gives a vector as an output corresponding to the image's probability of belonging to each potential class. CNNs often deal with the entire assignment pipeline, from feature extraction to learning the desired outcomes. End-to-end training refers to methods that do not use any hand-crafted interme-

diary algorithms. As a result, fully automated techniques reduce manual intervention and have produced amazing results in a variety of applications, ranging from self-driving automobiles to medical imaging segmentation jobs [48].

CNN architectures can be deftly used for segmentation tasks in which each voxel or pixel is given to a class. In the beginning, this was attained using a patch-wise classification method. In this approach, every pixel is categorized using a patch surrounding it, and the patch is then moved in a gliding window pattern until all the pixels in the image have been looked over. However, two significant flaws were discovered [25]. To begin with, because the patch usually only covers a limited area surrounding the pixel, the architecture of the network focuses on learning local properties while not considering the global arrangements. Second, the convolutions are performed redundantly due to several overlapping patches.

To get over these limitations, an alternative technique is given that uses the generic CNN construct and replaces the fully connected layers with convolutions [25]. This technique allows for a multi-class prediction for several pixels with less calculations. Because of the convolutions and pooling layers, the network's output is smaller than the input. Several solutions have been investigated in order to build a segmentation map that is the same size as the original image [20]. Long et al. [43] suggested a method that uses deconvolution procedures to upsample the smaller feature maps. The authors of [43] describe deconvolution as backwards-strided convolution applying it as many times as the pooling layer reinstates the image's original size.

Earlier methods for image segmentation that used CNNs consisted of piling up many convolutional layers. [16]. The reason behind this was that as the number of layers in the image grew, so did the number of features extracted from it; however, the high number of learnable parameters makes the training process slow. An example of this kind of architecture was proposed by Cireşan et al. [16].

Ronneberger et al. [58] built on this previous research and created the U-Net architecture for biomedical segmentation of images. It comprises of a diminishing route, which is made of a generic CNN, and a growing path, which uses deconvolution to restore the image's original size. Because of its topology, this network is also known as an encoder-decoder.

2.5.2 U-Net

The U-Net is the most well-known "CNN architecture for medical image segmentation" [58], and it is the one on which my own thesis is based. The fully-convolutional network architecture, which resembles a convolutional autoencoder,

is shown in Fig. (2.2). The core idea of the deconvolutional layers was borrowed from [43], but there is one significant difference: there are so-called skip connections between feature maps at the same depth level of the contracting and expanding path. The feature maps are then concatenated, giving the U-Net a substantial advantage over patch-wise techniques in terms of allowing global features to propagate to higher resolution layers. As a result, the skip connections feature can be found in a variety of other segmentation works, with the characteristics mapping merged in various ways.

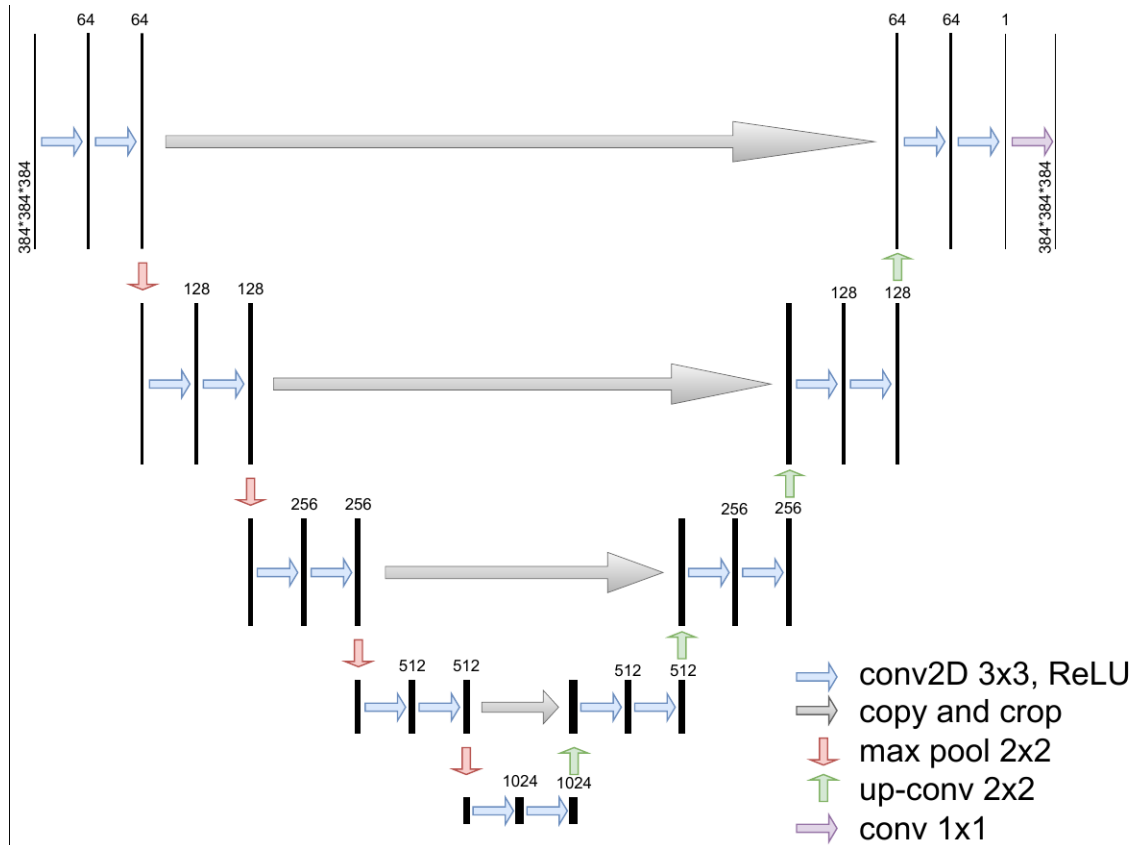


Figure 2.2: A schematic of our 2D u-net model.

There are five layers of depth in our network architecture. Each level in the contracting path performs two 3×3 convolutions, followed by a rectifier linear unit (ReLU) activation function and a 2×2 max pooling operation that decreases the feature map by half. The feature channels start at 64 and double with each subsequent level. The expanding path is similar to the contracting path, with the exception that the feature maps from the contracting section are concatenated and a 2×2 up-convolution is used instead of max pooling. At each layer, this last process doubles the image's size. Following the last expanding layer, the final segmentation map is generated using a convolution with 1×1 filters. The U-Net

was created in Caffe [33], a popular deep learning framework, and trained using stochastic gradient descent as the optimization approach in the original research. On the last feature map, a pixel-wise softmax is applied, and the dice loss is used as the loss function. A weight map is incorporated in the dice loss calculation to correct for the frequency mismatch across classes. When we talk about medical image segmentation, class imbalance is a common problem, and loss function weighting could be a solution. Another option is to use the Dice coefficient as a metric [48], which takes into account the overlapping regions of the source image and the segmentation mask.

In this work, data augmentation is a key strategy. Obtaining a huge dataset on which medical imaging analysis can be performed is generally difficult, and this is a simple and effective way to solve this issue. With only a few training photos available, the U-Net was trained end-to-end and obtained exceptional results in a variety of biomedical image segmentation tasks. It won the ISBI cell tracking challenge for transmitted light microscopy images in 2015, outperforming the previous best technique in the ISBI challenge for segmentation of neural structures in electron microscopic stacks.

2.5.3 Some other Relevant studies

Cicek et al. [15] extended the U-Net framework to 3D implementations. Their network generates rich volumetric segmentations after being trained with 2D annotated slices. Since a large percentage of medical data belongs to 3D, with same sizes in every dimension, and a slide-by-slide application of 2D convolutional procedures can be costly, 3D implementation is an interesting evolution of the initial U-Net. Inspired by the U-Net, Milletari et al. [48] introduced the V-Net, a fully-convolutional neural network for 3D image segmentation which is volumetric. The V-Net solves the vanishing gradient problem by performing 3D convolutions using volumetric kernels and using residual blocks. The residual block is a learning architecture that was recently created to make image recognition job optimization easier [28]. At the end of these blocks, the signal, which has been processed with convolutions and non-linearities, is aggregated with the stage's original input. Convergence can be reached in less time this way. Fig. (2.3) depicts a residual learning building block. The objective function is another important distinction between the V-Net and the U-Net. The cross-entropy is no longer used, and an objective function which is based on the Dice coefficient is presented to deal with the unbalanced classes problem. This strategy has the advantage of not requiring any hyperparameters (like the weight map, for example) to maximise the dice

similarity coefficient. Given the relative manual ground truth annotations, the V-Net was trained from the first to the end on a dataset of prostate scans in MRI in the original work. Based on the Dice coefficient, it was able to achieve a quick and accurate segmentation.

In a recent study, the feasibility of using the same CNN for multiple segmentation tasks in images obtained with different modalities was also investigated [51]. A single convolutional network was proposed in this study that executed nicely in the segmentation of tissues in pectoral muscle in MR breast images, MR brain images and coronary arteries in cardiac CTA. Since of their adaptability and learning capabilities, CNNs could potentially solve a variety of segmentation challenges in the future without any prior task-specific training.

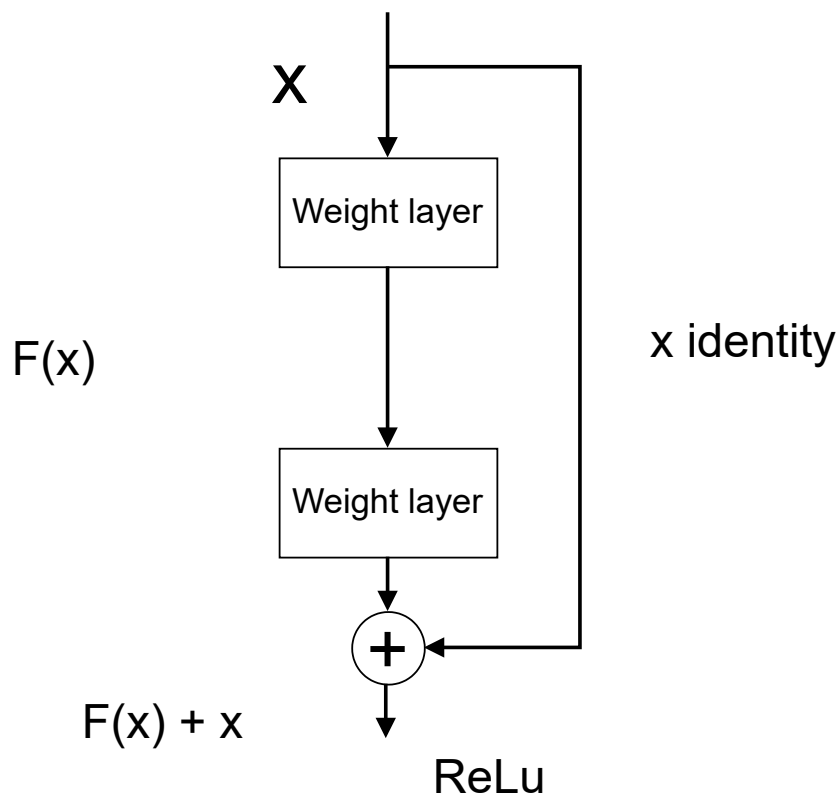


Figure 2.3: A schematic of a residual block, as proposed in [28]

A new novel segmentation strategy based on adversarial training is also worth highlighting. In 2014, Goodfellow et al. [26] proposed generative adversarial networks, a new construct for estimating generative models using an adversarial approach. Two networks are simultaneously trained in this experiment, one that creates data and the other that learns to distinguish whether such a sample arrives from the training data or the generative model. The generative network improves its output as a result of the competition between the both models, providing data

that the classification function cannot differentiate from the original. Adversarial networks have recently been used in semantic segmentation [44]. This approach has been shown to be effective in correcting inconsistencies between both the ground truth and a segmentation CNN's segmentation map. In the field of medical imaging, this method has yielded promising results in the segmentation of prostate cancer in MRI [38], organs in chest X-Rays [17], and brain MRI [50]. In comparison to earlier methods based on a single segmentation net, all of these investigations indicate improved segmentation performance. Further examination of this architecture, including additional challenges, could be an intriguing future deep learning research area.

To summarise, numerous medical image segmentation challenges have been solved using various deep learning algorithms in recent years. The approaches that receive the entire image or a big subregion as input have beaten out the "slide window" techniques, but both could be improved, and developing a particular architecture for a certain task might not be the only effective solution.

2.6 Gray level features methods

A screening tool for femur bone can be created by combining the grey scale features of CT scans with the appropriate segmentation mask. However, because proximal femur segmentation currently requires manual intervention, these approaches have yet to be adopted for clinical use, limiting their utility in a clinical setting.

In these methods segmentation is performed on the intensity values of the pixels. These techniques can be classified as follows, fixed threshold [35], [9]; Otsu Threshold, watershed-based techniques [66], [2], [5], [27]; graph-cut [56], and model-based approaches [37], [68], region growing, region split and merge, edge based segmentation and a few more. The most common methods for dealing with medical photographs are explained in the following subsections.

2.6.1 Fixed threshold

The most basic segmentation approach is thresholding, which is based on the notion that distinct pixels/voxels in an image have different intensity values. It splits pixels into foreground and background, with foreground having a greater intensity value and background having a lower intensity value. Therefore it can be defined mathematically as follows:

$$f_{k,j} = \begin{cases} 1, & \text{if } i_{k,j} \geq t \\ 0, & \text{otherwise} \end{cases}$$

where t is the threshold value, $i(k,j)$ is the initial value of the pixel at coordinate (k,j) and $f(k,j)$ is the result after applying the threshold value.

Thresholding based method works well when only two segmentation classes is desired and if the image and each class's pixels have similar intensities and minimal noises. A multi threshold based method can also be applied when we have multi class segmentation requirement.

2.6.2 Otsu Threshold

Segmentation based on threshold technology is essential in image analysis and segmentation. Otsu's method is widely used in calculating stable, simple, and compelling analysis [41]. A binary image or black and white image is beneficial in many image analysis or processing problems as it reduces the processing complexity. An essential step in obtaining the binary is to figure out the optimal threshold that can be used to convert the greyscale image into a binary image; this can be done by using global and local thresholds [65]. The global threshold is a single threshold value used for the whole image, and the local threshold is a single unique value used for partitioned sub-images of the whole image.

2.6.3 Region growing and Split and Merge algorithms

Region growing is a technique for extracting a region of interest from a scalar image or volume using predetermined growing criteria (intensity distribution, connectedness, and image edges)[29]. It incorporates information about neighbouring regions and is aimed to locate homogenous zones that have a higher likelihood of corresponding to anatomical features than thresholding techniques. For each object to be segmented, this approach necessitates at least one manual seed point, from which all the pixels or voxels that correspond to that object are selected based on homogeneity criteria. As a result, the fundamental disadvantage is that it is very dependent and sensitive to initialization, and that noise and partial volume effects have a significant impact.

Split and merge algorithms are very much similar to region growing algorithm. It dose not requires seed points [46]. Similarly, it splits the regions into subregions and merges the ones that meet the necessary requirements, based on the manual predefined criteria. The fundamental problem of this approach is that

it requires a dataset with a pyramidal grid structure, which makes it computationally expensive and unsuitable for today's massive data sets.

Thresholding-based techniques are majorly used in the segmentation of tissue in CT scans. Because bones have higher intensity values than other surrounding tissues, this method of segmentation is usually successful. Due to shallow space between the femur and Tibia and weak and diffused boundaries between all bones in the knee area, it's rare to find a threshold that's less than the the intensity value of all bones yet more than that of surrounding tissues. The local thresholding strategy can prevent this problem, however due to intensity inhomogeneity, it may not always perform effectively. All of these approaches are notoriously noise-sensitive, resulting in over-segmentation and a failure to detect important areas with low-contrast bone borders. All these methods can only segment tissues; these methods cannot be able to segment different bones.

To produce precise, acceptable masks for finite element modelling, all segmentation algorithms now require a user-in-the-loop parameter which adds manual intervention to the segmentation. Their processing time is so long that it cannot be used for clinical usage. This lack of robustness is costly and requires a trained professional to correct the segmentation manually and time taking. As a result, these methods are not as effective at processing huge cohorts as fully automated systems. Deep learning in image segmentation has received a lot of interest in recent years. The U-Net architecture provides a direct result of the use of deep learning as a feasible choice for biomedical image segmentation by Ronneberger et al. [58]. This research demonstrates how the U-Net architecture may provide accurate and rapid segmentations without the use of massive training data sets. Deep learning has become the state-of-the-art approach in medical imaging, notwithstanding the necessity for ground truth masks. [60, 32]. A segmentation prediction of data can be generated fully automatically (i.e., without human intervention) in a few seconds, which is often faster than preceding methods that do not use deep learning.

2.7 Bone segmentation

In many medical treatments, segmenting bones from CT images is a crucial step. It has a variety of applications in image based computer aided orthopaedic surgery; it is critical for finding fractures and diagnosing bone diseases, as well as assisting the radiologist in the overall medical decision-making process. They also provide some reliable resources for examining and segmenting other body sections, such as organs. As a result, a lot of time has been spent looking for an automatic seg-

mentation algorithm. Despite extensive research, current systems do not provide reliable image segmentation in the presence of non - uniformity or noise, and frequently requiring algorithm setup or parameter tuning. In a clinical situation, if a plan to make radiologist work easier is essential, this can be a big disadvantage. Furthermore, the majority of these methods are unable to discriminate between distinct bone structures and classify them. They can only segment a small number of bone structures, and their lack of generality is another issue.

The bones are difficult to segment for a variety of reasons. To begin with, osseous tissue and soft tissue regions may not always appear to be easily differentiated on CT images. A segmentation based purely on the intensity value would be inaccurate because the HU values of the bone and other surrounding tissues overlap significantly. Second, the bone structures themselves do not have consistent density qualities, hence the HU within a single bone might vary greatly. To make matters even more complicated, several disorders (such as Osteoarthritis, Rheumatoid Arthritis, and Osteoporosis) can affect bone density in different regions. Finally, CT scans' limited resolution and poor quality can be an issue. The presence of noise or artefacts impacts the automatic segmentation performance, as adjacent bones appear to be in direct touch.

2.8 Related work

The four primary techniques to bone segmentation in CT scans are region-based, edge-based, intensity-based and deformable [26]. Atlas-based segmentation is another technique that has recently shown promising results. The most basic segmentation approaches are intensity-based techniques, which are based on local or global thresholding as explained in the preceding sections. They're usually simple to set up, but their main drawback is that they're not adaptable because they require the regions to be separated into groups with comparable intensity values. This flaw can be addressed in part using an adaptive threshold [69]. In this 2010 study, an automated 3D adaptive thresholding technique for bone segmentation in CT scans was developed. It produces good results and takes less than 10 seconds each 2D slice on average. In more recent studies, thresholding has become an integral part of the segmentation algorithm's pre-processing pipeline.

Edge-based approaches use a variety of edge detection operators to locate and segment the boundaries of the bones. For example, this study [35] used various edge detection approaches to produce effective segmentation of the proximal femur, knee, and cranium in CT scans. It also simply requires a very basic ini-

tialization. It does, however, require a site-specific separation of particular bones in order to be applied to other anatomic locations. In general, edge-based approaches are sensitive to noise and do not always produce the best results.

Region-based techniques partition an image into sections based on a homogeneity criterion. In the previous chapter a few more details are explained. Finally, when using deformable techniques, research refers to model with active contours or surfaces. Those curves change during the segmentation process due to various forces, yet the model stays smooth. Those approaches have been the most extensively studied in recent years for bone segmentation in CT scans, with multiple studies yielding positive results [12]. They do, however, have some drawbacks, such as an extreme sensitivity to the beginning conditions.

Hybrid methods that combine the aforementioned methodologies have also been studied, with mixed results. The authors of [13], proposed an approach that consists of image contouring, pre-processing, and label filtering.

Atlas-based image segmentation algorithms are an innovative approach to image segmentation. They use a collection of expert-labeled images and hope to apply what they've learned to new, unknown data. These approaches compute a transformation that registers the atlas to the image being studied in order to segment a new image (i.e establishes a point-to-point correspondence). To segment the patient image, the mask is deformed from the atlas onto it [21]. Atlas-based algorithms, among the previously outlined methods, have recently achieved better accuracy and robustness in many segmentation tasks. It's worth highlighting this 2017 study [23] on human skeleton segmentation in CT images. For the segmentation of individual bones, it presents a unique atlas and articulated registration approach. This can segment 62 main bones with pinpoint accuracy, including 24 vertebrae and 24 ribs. The average Dice coefficient is 0.90, which is better than prior studies' results.

Moreover, in the near future, deep learning-based algorithms will undoubtedly play a significant role in bone segmentation. The most popular topic in publications applying deep learning to medical imaging is segmentation [40]. Recent studies have described successful methods based on deep learning based pelvic bone segmentation [42], proximal femur segmentation [18] [7] in Ct images. The proximal femur is automatically segmented in quantitative CT images using a deep learning-based U-Net inspired architecture in this [7]. The network implementation is identical to the one presented in this thesis, except the operator chooses the region of interest. The method reported is intriguing because it uses a local otsu method as pre-processing to detect bone regions in CT scans and elimi-

nate background voxels, which reduces image size. They achieve a mean DSC of 0.990 ± 0.002 and a mean HD95 of 0.981 ± 0.040 mm using this method. There would definitely be more studies on the subject in the coming few years.

This thesis proposes an automated approach for detecting regions of interest in CT images and using the U-Net architecture to segment the femur bone to the best of my knowledge.

CHAPTER 3

Fundamentals

3.1 Computed Tomography images

Computed tomography (CT) is a medical imaging method that creates cross-sectional images of the patient's body using a combination of X-rays. Where a person is exposed to sequential X-ray radiation along with the region of interest. These X-ray images result in a 3D volume [11]. After post-processing, it creates a cross-sectional image (slices). The original image can be reconstructed using algebraic methods like Fourier-based techniques or statistical methods, giving grayscale values as in Hounsfield units (HUs) defined by:

$$HU(x, y, z) = 1000 \frac{\mu(x, y, z) - \mu_w}{\mu_w} \quad (3.1)$$

where $HU(x,y,z)$ represents the Hounsfield units at location (x,y,z) , (x,y,z) is the corresponding avg. linear attenuation coefficient at the same location, and μ_w is the linear attenuation coefficient for water at the specific conditions used in the procedure [10].

These CT slices are DICOM files with the file type (.dcm).

3.2 Segmentation

Segmentation is a method in which an image is divided into many subclasses called image segments. Voxels from 3D imaging modalities such as Computed Tomography (CT), Micro-Computed Tomography (micro-CT or X-ray), or Magnetic Resonance Imaging (MRI) scanners are labelled during 3D image segmentation to isolate regions of interest as a separate category. In this thesis, we have done tissue segmentation and bone segmentation. In tissue segmentation, we differentiate voxels in a different types of tissue. In contrast, bone segmentation

means differentiating voxels of different types of bones.

3.3 Evaluation Parameter

In the task of the segmentation generally we classify the prediction into the four categories: "TP", "FP", "TN" and "FN". With this we compute the pixel wise model result and get the model performance it can be seen in the Table 3.1

The result "TN" stands for True Negative, and it displays the number of negative occurrences that were correctly identified. Similarly, "TP" stands for True Positive, which refers to the number of positive events that have been correctly identified. False Positive and False Negative values are denoted by the letters "FP" and "FN," respectively. False Positive value (FP) is the number of actual negative cases categorised as positive, whereas False Negative value (FN) is the number of actual positive instances classified as negative.

Table 3.1: Confusion matrix

		Predicted	
		Negative	Positive
Actual	Negative	True Negative (TN)	False Positive (FP)
	Positive	False Negative (FN)	True Positive (TP)

The F1 score, precision, and recall are the measures that are used to track the training and evaluate the test dataset. F1 score which is also known as Dice Score [57].

3.3.1 Precision

Precision shows the purity of the predicting result. It is a ratio between the positively predicted pixel that matched ground truth to the all the positive predicted pixel.

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}} \quad (3.2)$$

3.3.2 Recall

Recall shows the completeness of the predicting result. It is a ration between the positively predicted pixel that matches ground truth to the all correct pixel in the ground truth.

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}} \quad (3.3)$$

3.3.3 F1 Score

The overlap index, also known as the Dice coefficient/score or F1 score, is a statistical metric that shows how similar two samples are. Its range of values is 0 to 1, with 0 indicating that the samples do not overlap at all and 1 indicating that they are identical.

It's also known as mask, and it's widely used in image segmentation jobs to compare an algorithm's output to its relative ground truth image. It is the most frequently used statistic to validate medical volume segmentations [62].

$$\text{F1 Score} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (3.4)$$

Another common index in these tasks is the Jaccard, which has the following relationship with the F1 score:

$$\text{JAC} = \frac{\text{F1 score}}{2 - \text{F1 Score}} \quad (3.5)$$

Because the two metrics evaluate the same overlapping qualities, monitoring them both offers no new information [62], and only the F1 Score is taken into account in this research.

CHAPTER 4

Data Set

The Shah Hospital, Ahmedabad, provided us with CT scans of both men and women aged between 45yrs - 82yrs old. This dataset of high-quality CT scans contains full-leg CT scans of 80 patients of which 50 are used for training, 10 are used for validation and 20 are used for testing. The in-plane resolution of each CT scans ranges from 500×500 to 512×512 voxels with 812- 1211 slices and $0.9 \times 0.9 \times 0.99$ to $0.97 \times 0.97 \times 1 \text{ mm}^3$ voxel size.

4.1 Data annotation

Building an artificial intelligence or machine learning based model that acts like a human requires many training data. To train a model, data and their label are required. Data annotation is the categorization and labeling of data.

4.1.1 3D Slicer

3D slicer [1, 36] is an open-source software that gives us a platform to edit, analyze and segment parts of CT scans. We used 3D Slicer to make our ground truth/lable our data. Even when done by professionals, labelling takes time, and the outcomes are dependent on the operator's skill. A variety of processes were performed to the initial scans in this work in order to create ground truth images. For this purpose, 80 CT scans were used to creating the ground truth.

The original scans (.dcm files) were preprocessed and checked with the 3D slicer tool for any type of corrupted or missing .dcm files. After preprocessing stage, a fixed threshold is decided by slowly increasing the threshold value from -200 till a threshold value at which only bone pixels whose threshold value is greater than fixed threshold value are remaining. Segment editor tool of 3D slicer is used to select or deselect pixels of any labels. 3D view of 3D slicer is used to verify, identify any noise and eliminating noisy voxels. In Fig. (4.1 one can see, in

right pane three axial sections of CT images is shown with bone highlighted at a fixed threshold value of 144.01.

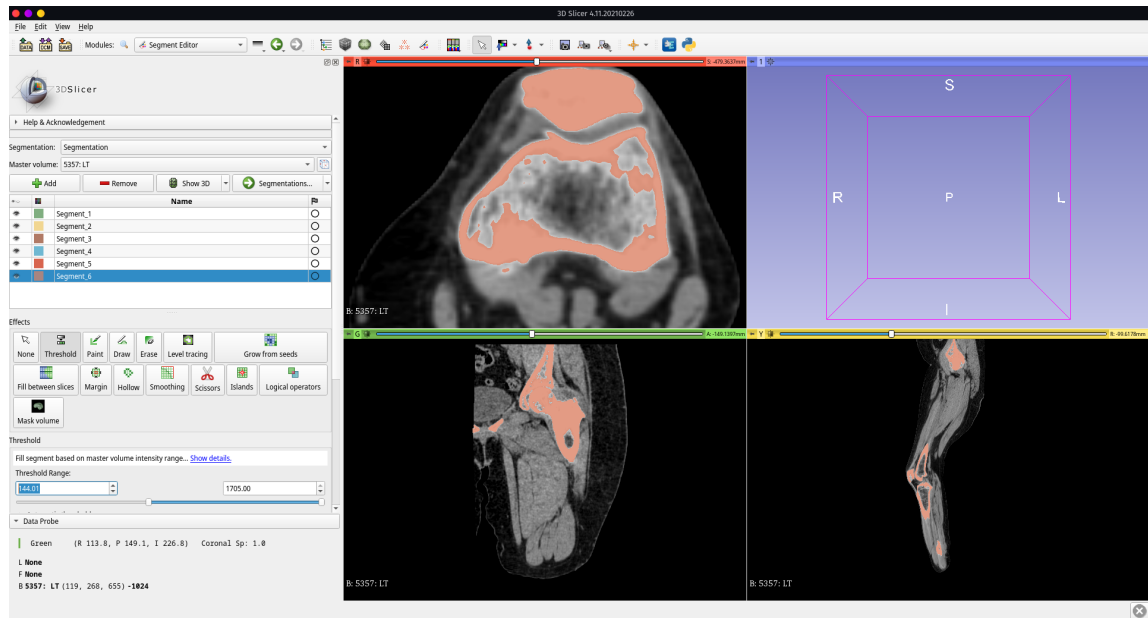


Figure 4.1: A view of 3D slicer application

These annotations are mostly made up of plain binary data that is saved slice by slice. Data visualization and file reading for U-Net training requires the creation of unique classes and methods. This, combined with the lack of homogeneity in data resolutions, encouraged us to save the data in a more standard format than DICOM. For the following reasons, we chose the NIfTI format:

- Folder organisation was simplified. NIfTI data is organised by volume. As a result, per patient scan, only two files are required: one for CT and one for CT ground-truth values. When compared to DICOM format, this significantly reduces the total quantity of files.
- There are visualisation tools (3D Slicer) as well as many reading and writing packages for NIfTI files available.
- All slices have the same pixel spacing, rotation, and orientation.

With the available metadata from the first slice, we generated the new NIfTI files. Obtains the origin coordinates, voxel dimensions and orientation for CT images, creating the affine matrices required by the NIfTI format. We used the same affine matrix for the CT ground-truth volumes as we did for the relevant CT images, and based on the annotated slice information, we assigned the following voxel values:

- The background label is labelled with a number of zeros.
- 1 is for the Femur bone.
- 2 is for the Patella bone.
- 3 is for the Tibia bone.
- 4 is for the Fibula bone.
- 5 is for the Pelvis bone.
- 6 is for all the bones in the foot.

These numbers from 0-6 are the nomenclature for different classes in our ground truth. Fig. (4.2a). shows the original CT image in Axial format, Fig. (4.2b) shows the original CT image in Coronal format, Fig. (4.2c) shows the original CT image in Sagittal and Fig. (4.3) shows masked image of the same images shown in the rest of the three figures.

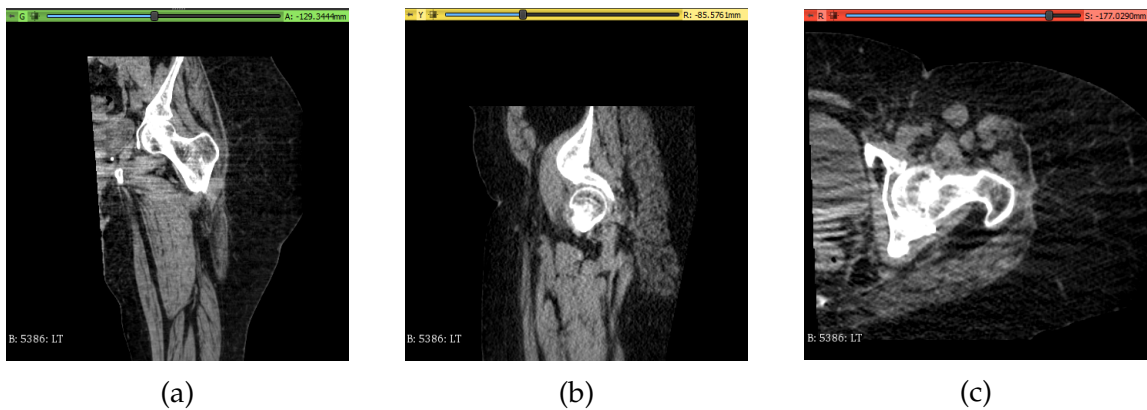


Figure 4.2: Example of CT slices in axial, coronal and sagittal format.

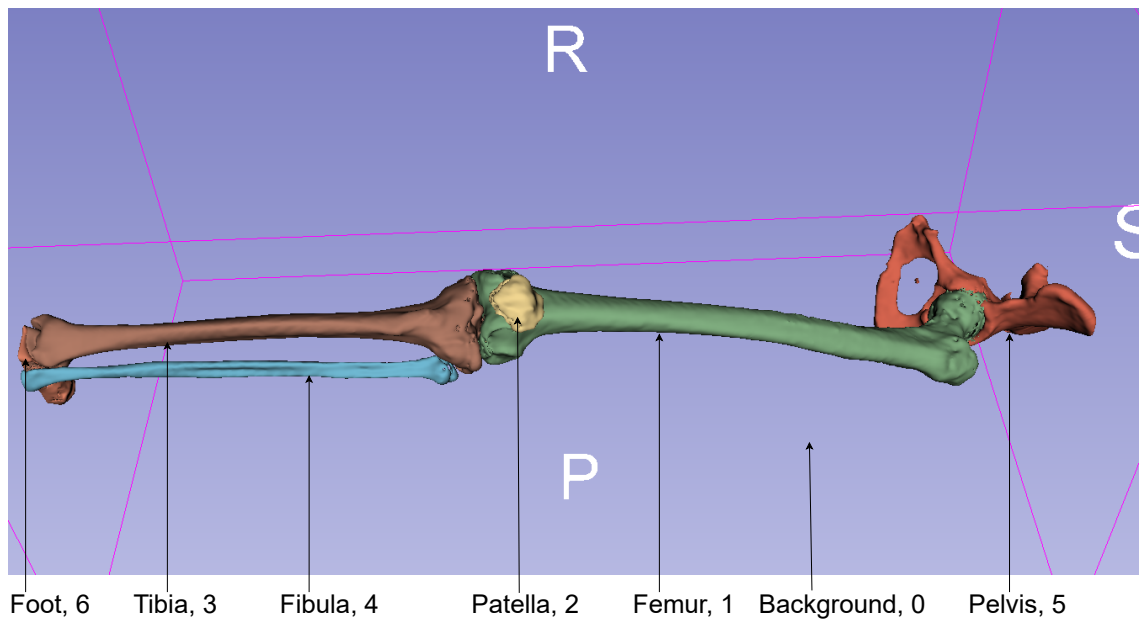


Figure 4.3: 3D mask for ground truth.

These data annotation techniques and annotated data are verified by Dr. Manish Shah, Shah Hospital, Ahmadabad.

4.1.2 MRIcro

MRIcro [59] is an application that allows Windows and Linux computers to view and export medical images. We have used MRIcro to label Landmark co-ordinate of the following:

- Femur head centre.
- Femoral centre.

4.2 Data augmentation

The performance of most machine learning models and deep learning models, in particular, depends on the quantity and quality of training data. However, insufficient data is one of the most common challenges in implementing deep learning models. Collecting and preprocessing data is time-consuming and costly.

Data augmentation is a set of techniques to artificially increase the amount of data by creating new data points from existing data and by making some small changes to the data by rotation, scaling and translation. The augmented data are similar to the real data for models.

Here we have done data augmentation two times, by following factors:

- Rotation up to 15° .
- Transnational radius = 10 mm.
- Scaling = 0.2.

These factors are aligned with the ideal condition of a CT Scan taken, with different patients having legs slightly rotated and always facing up. We have increased our data set of 50 CT scans to 150 CT Scan images by augmenting for the training.

CHAPTER 5

Landmark detection

Identifying femoral bone landmarks is essential and used mainly in orthopedics and biomechanics. A fully automated detection overcomes the drawbacks of labor-intensive manual identification. Landmarks give us the location of objects in the 3D plane, which will be used to identify regions of interest. This thesis aims to automatically identify the femur head center and distal femur landmark, which will be used with 3D UNet to segment the femur and distal femur with very few resources and less time.

5.1 UNet Model

For the landmark prediction, data upto 20mm in radius around the landmark coordinate from CT ground truth values is kept as it is and diluted rest of the data. We selected 2D U-Net architecture for this model to segment the landmarks. Fig. (5.1) shows the architecture of our 2D UNet.

For this implementation, I used the Tensor-flow based Keras framework for the U-Net. For training the model we have used newly created NIfTI files, we selected the 41 slices keeping landmark slice in between, and 55 slices randomly from the each CT scan of all patients. The training was performed on Google Collab Pro, to avoid surpassing the available memory, the training ran for 2700 epochs for 25 CT images with a batch size of 4, resulting in a total number of iterations of 8,10,000 (2700 epochs \times (25*96) slices/ batch size of 8)). We have performed this experiment for femur head and distal femur.

5.2 Results and Analysis

Following training, the model is tested using the remaining annotated volumes. Data was fed into the network to build a probability map for each of the pixels,

and afterwards the values were converted to discrete values to generate a binary 3D image. We have taken centroid of this predicted image as predicted center.

5.2.1 Femur head center

Table 5.1 shows the accuracy of the femur head's center prediction in the 3D plane and z-axis(Superior axis), which can be considered good as per our requirement to locate the femur head in 3D CT Scan volume.

Table 5.1: Landmark prediction accuracy of Femur head (in mm).

Patient Id	3D plane	Superior axis
78	7.07	0.67
82	11.35	0.02
88	9.88	1.35
81	13.74	0.96
102	11.67	0.54
103	4.13	1.1

5.2.2 Distal Femur

Table 5.2 shows the accuracy for predicted co-ordinate center of distal femur in 3D plane and z axis(Superior axis). Which can be considered good as per our requirement to locate distal femur in 3D CT Scan volume.

Table 5.2: Landmark prediction accuracy of distal femur(in mm).

Patient Id	3D plane	Superior axis
78	6.55	2.01
82	8.14	3.23
88	6.26	5.56
81	12.26	4.04
102	4.8	3.61
103	10.35	4.09

Table 5.3: Landmark prediction accuracy of distal femur in 3D plane and z-axis

Landmarks	3D Plane	z axis(slice)
Femur Head center	9.64 ± 3.18	0.7 ± 0.01
Distal Femur	6.5 ± 2.5	3.75 ± 0.43

In table 5.3 one can see the average accuracy of the femur head and distal femur in the 3D plane is under 10 mm distance, and in slice or z-axis, it is up to

1mm in the case of the femur head and 4mm in the case of the distal femur. Our model fulfills our requirement of predicting the location of the femur head and distal femur location in the whole CT Image of the leg.

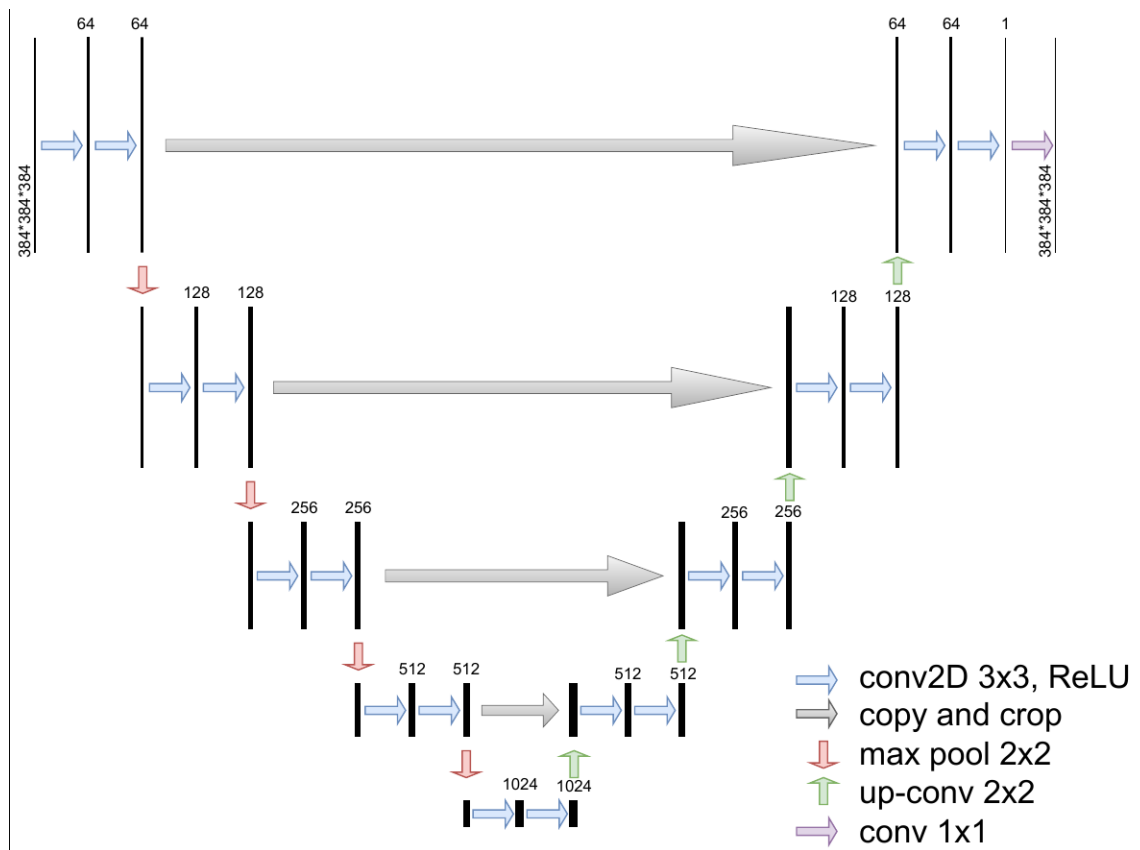


Figure 5.1: A schematic of our proposed 2D u-net model.

CHAPTER 6

Segmentation

Tissue Segmentation [3] is not something that has been recently introduced. There are architectures such as UNet [58], and Segnet [4] which have been used for image segmentation, which has also been used and referred to during medical imagery segmentation. Also, tools such as 3D slicer provide a workspace for a user to perform segmentation in a non-AI approach. Many diagnostic and treatment planning is dependent on the successful segmentation of bony structures in the CT-scan images.

6.1 Model I : Fixed Threshold-based Segmentation

The Hounsfield unit (HU) [19] is a computed tomography images are interpreted using a relative quantitative measurement of radio density. The attenuation or observation coefficient of radiation within a tissue is used to produce a grayscale image. The attenuation and observation of the X-RAY beam are proportional to the physical density of the tissue. Where distilled water is arbitrarily defined as 0 HU, Air is -1000 HU, bones range from 400 HU and can go up to 2000 for dense bones like cochlea and more than 3000 for metals like iron and silver. [19].

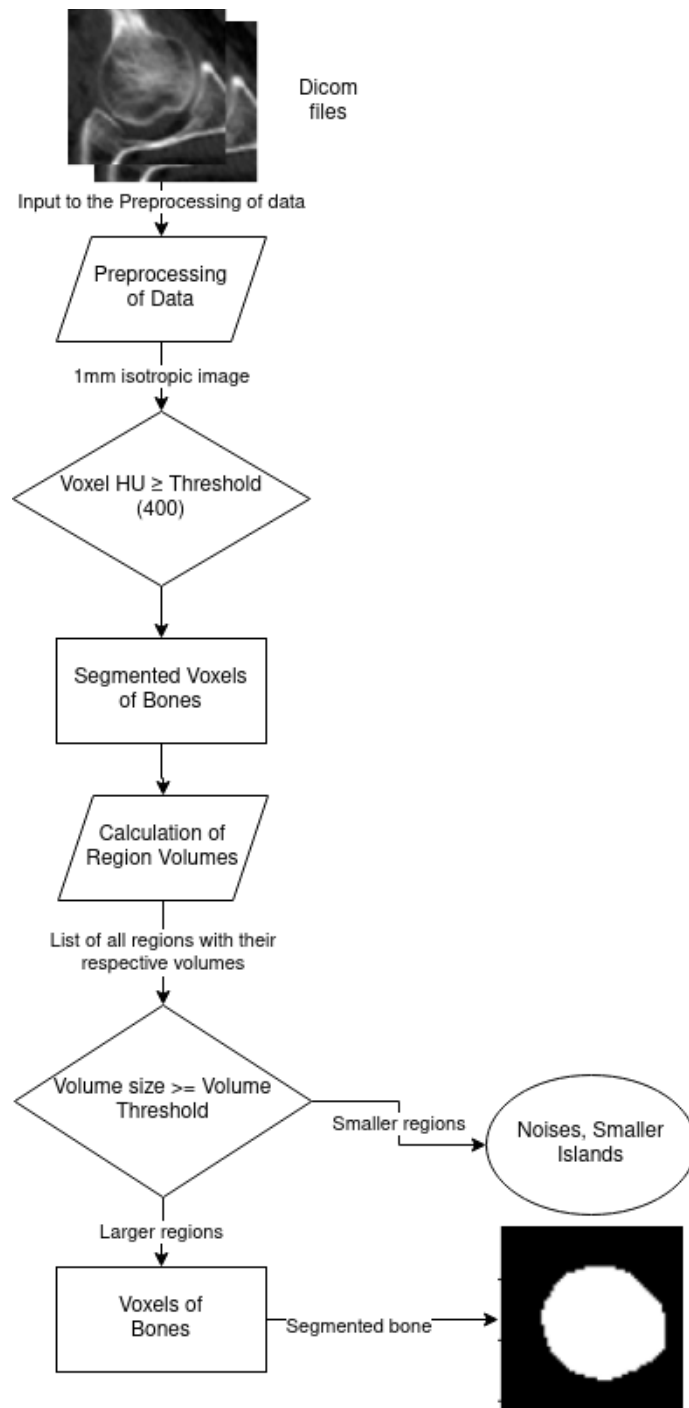


Figure 6.1: A flowchart showing the workflow of fixed threshold segmentation method.

Fig. (6.1) shows the flowchart depiction of the algorithm for the approach for doing HU-based standard segmentation. CT scans have different resolutions, so every CT scan data is processed to have the exact 1mm resolution in the preprocessing stage. Then, we select voxels based on HU value, voxels that have HU value in the range of bones are selected, but many voxels are noise and have HU

value in the range of bone. Furthermore, connected voxels which form a region are selected, and these regions are selected based on volume size that indicates the possibility of a bone. 6 Patients' CTs are taken for this experiment. Fig. (6.2) depicts a test scan segmentation image at knee for the sole purpose of visualisation. The prediction is visualised in three dimensions. We have calculated precision, recall, and F1 score to mask (Ground Truth as mentioned in Chapter 4) for fixed threshold-based segmentation. These scores of the performance matrix are given in table 6.1

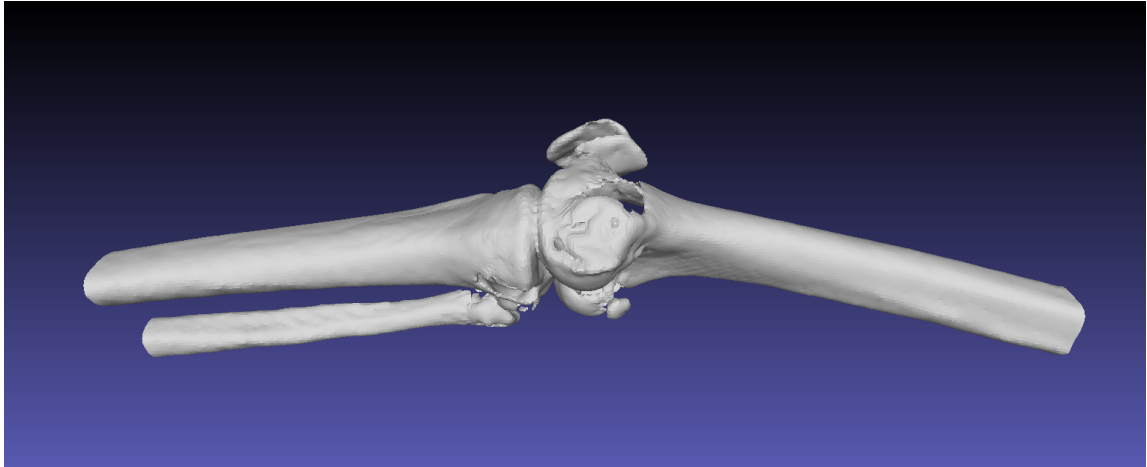


Figure 6.2: 3D view of a prediction of bone by fixed threshold value method.

Table 6.1: Performance matrix score of fixed threshold segmentation

Patient ID	Precision	Recall	F1 Score
78	0.863	0.9951	0.924
82	0.875	0.9891	0.929
88	0.908	0.9867	0.945
81	0.856	0.9931	0.919
102	0.851	0.9859	0.913
103	0.917	0.9844	0.949

6.2 Model II : Otsu's Segmentation

Segmentation based on threshold technology is essential in image analysis and segmentation. Otsu's method is widely used in calculating stable, simple, and compelling analysis [41]. A binary image or black and white image is beneficial in many image analysis or processing problems as it reduces the processing complexity. An essential step in obtaining the binary is to figure out the optimal

threshold that can be used to convert the greyscale image into a binary image; this can be done by using global and local thresholds [65]. The global threshold is a single value applied to the entire image and the local threshold is a single unique value used for partitioned sub-images of the whole image. The local threshold technique has been used here to segment bones on the Knee Joint of size 384mm in length. Fig. (6.3) depicts a test scan segmented femur, tibia, fibula, and patella bones for the sole purpose of visualisation. The prediction is visualised in three dimensions.

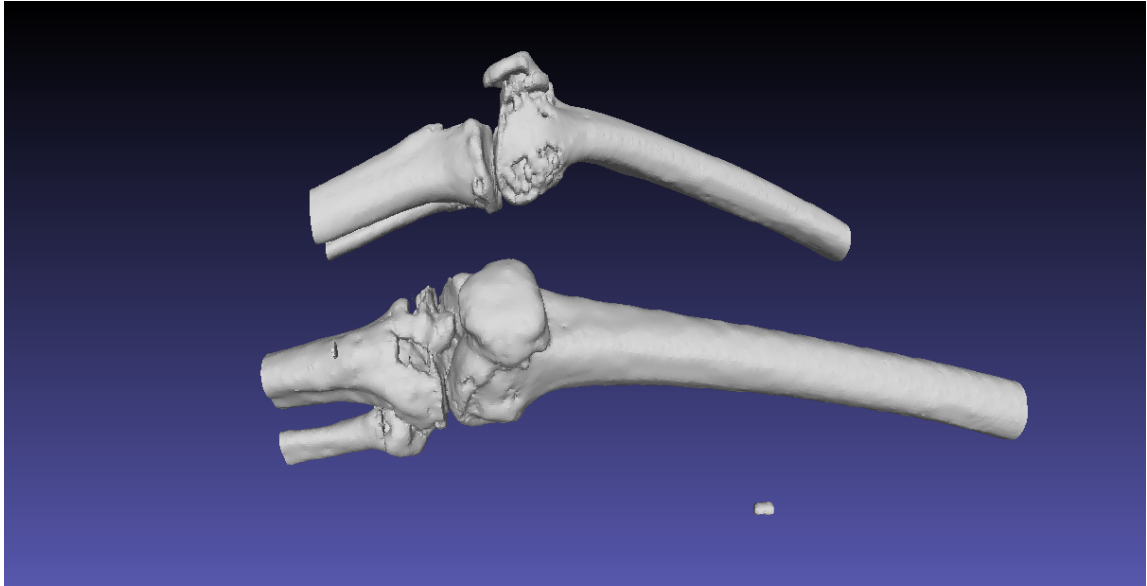


Figure 6.3: 3D view of a prediction of bones using local Otsu’s threshold method.

We have calculated precision, recall, and F1 score to mask (Ground Truth as mentioned in Chapter 4) for local Otsu’s threshold based segmentation. The scores of the performance matrix are given in table 6.2

Table 6.2: Performance analysis for all bone based on Otsu’s segmentation.

Patient ID	Precision	Recall	F1 Score
78	0.49	0.98	0.65
82	0.44	0.97	0.61
88	0.51	0.97	0.67
81	0.45	0.98	0.62
102	0.47	0.97	0.63
103	0.55	0.98	0.71

6.3 Model III : 2D UNet

UNet [58] was first designed and applied in 2015 to process bio-medical images. It has evolved from a traditional convolution neural network. A conventional neural network is generally focused on classifying images, where input is an image and output is a label. However, in biomedical cases, it not only requires us to distinguish abnormal images from disease but also requires localization of the area of abnormality. The consensus is that an extensive data set with thousands of annotated training samples is required to train deep networks successfully. Also, UNet can work efficiently and gives a good result with solid use of data augmentation [58].

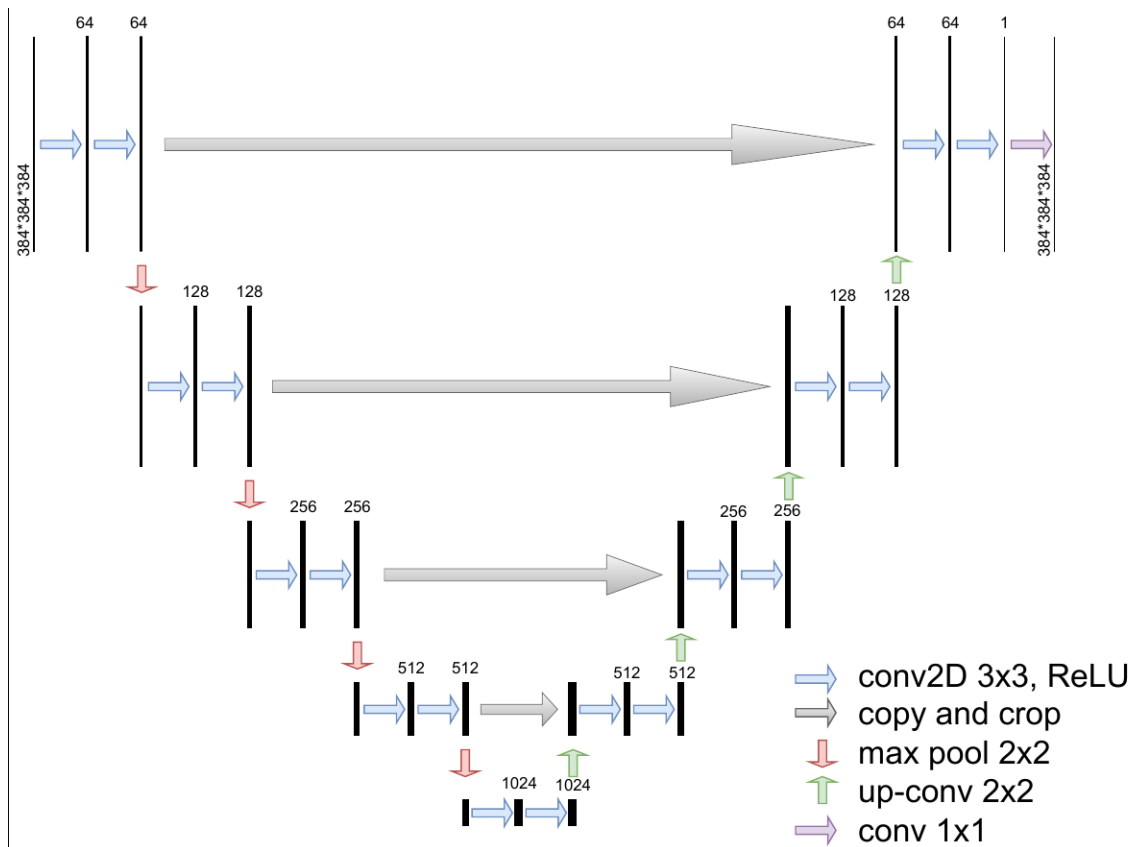


Figure 6.4: A schematic of our 2D u-net model.

The Fig. (6.4) explains our UNet architecture diagrammatically. Generally, CT of the leg is of size 512mm in height, 512mm in width, and 700mm to 1200mm in depth. Computational resources to process this massive data with 3D UNet are too high. Therefore, we have used 2D UNet and treated the third dimension as Channels. After that, Data is converted to 2mm resolution per voxel. We had used nineteen CT samples for training, five CT samples for validation, and six

CT samples for testing. After performing augmentation two times, there were 57 Samples for Training and 15 for validation. For this implementation, I used the Tensor-flow based Keras framework for the U-Net. For training the model we have used newly created NIfTI files, The training was performed on Google Collab Pro, Training was carried out for 7000 epochs for 57 training samples with 256 2D slice images each, with a batch size of 2 to avoid exceeding the available memory, resulting in a total of 5,10,72,000 iterations(7000 epochs \times (57 \times 256) slices/ batch size of 2)). We have performed this experiment for tissue segmentation, femur head segmentation and distal femur segmentation.

6.3.1 Tissue segmentation

While executing the U-Net segmentation, the network tries to understand the generalized anatomy of bone among all tissues and does the segmentation of bones. Fig. (6.5) depicts a test scan segmented femur, tibia, fibula and patella for the sole purpose of visualisation. The prediction is visualised in three dimensions. Table 6.3 shows the performance measure concerning ground truth.

Table 6.3: Performance matrix for Knee Joint based on U-Net deep learning model.

Patient ID	Precision	Recall	F1 Score
78	0.90021	0.83439	0.86605
82	0.8678	0.7237	0.78923
88	0.88072	0.81759	0.84798
81	0.83922	0.66723	0.74341
102	0.88703	0.80286	0.84285
103	0.9175	0.8213	0.86664

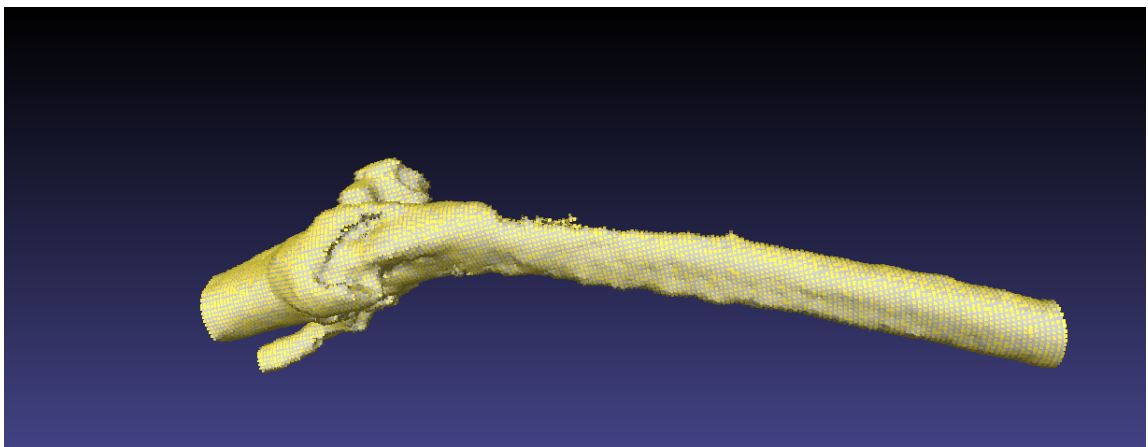


Figure 6.5: 3D view of a prediction of knee joint obtained by 2D UNet model.

6.3.2 Femoral Head

In this experiment, we are segmenting the femur head concerning Fig. (6.7), which contains the femur head and hip joint bone. In Fig. (6.11), one can see the segmented femur head obtained from the 2D UNet model.

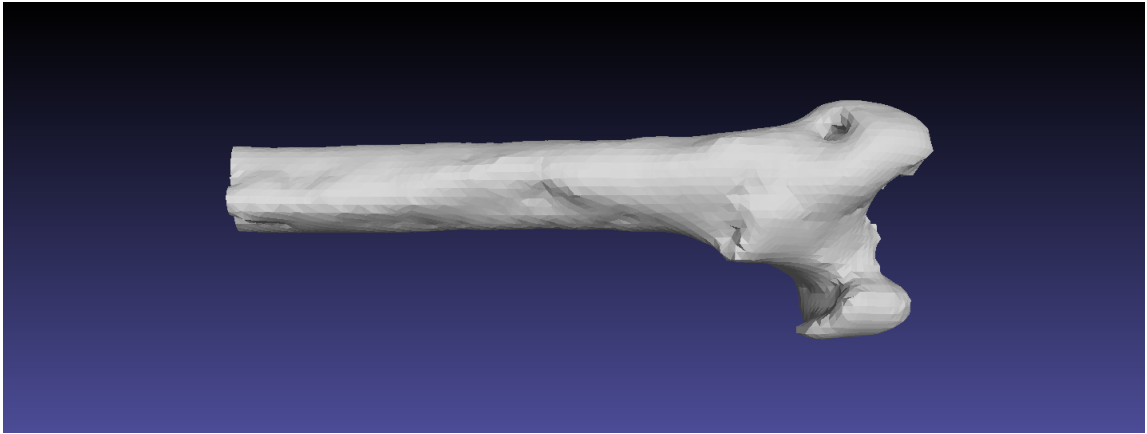


Figure 6.6: 3D view of a prediction of femur head and femur shaft from 2D UNet

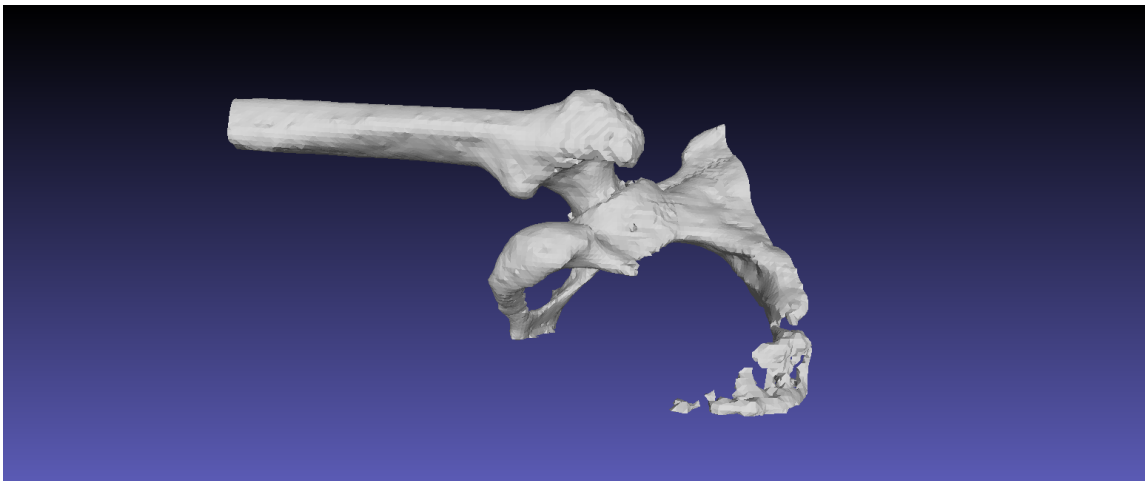


Figure 6.7: 3D view of ground truth at hip joint

Table 6.4 shows performance measure with respect to ground truth for femur head segmentation within the entire femur bone image.

6.3.3 Distal Femur

The distal femur is more complex than the femur head, as, in the distal femur, there are more bones and different angles between bones. Fig. (6.8) shows segmented femur bone obtained by 2D UNet, and one can see it does not have any

Table 6.4: Performance matrix for Femur Head based on 2D UNet deep learning model.

Patient ID	Precision	Recall	F1 Score
78	0.8786	0.8139	0.84504
82	0.873	0.7499	0.8068
88	0.885	0.7825	0.8321
81	0.8688	0.6656	0.7537
102	0.8859	0.7757	0.8272
103	0.8919	0.8196	0.8543

voxels of other bones. It shows model can segment the femur from other bones. Table 6.5 shows performance matrix of 2D UNet on Distal femur.

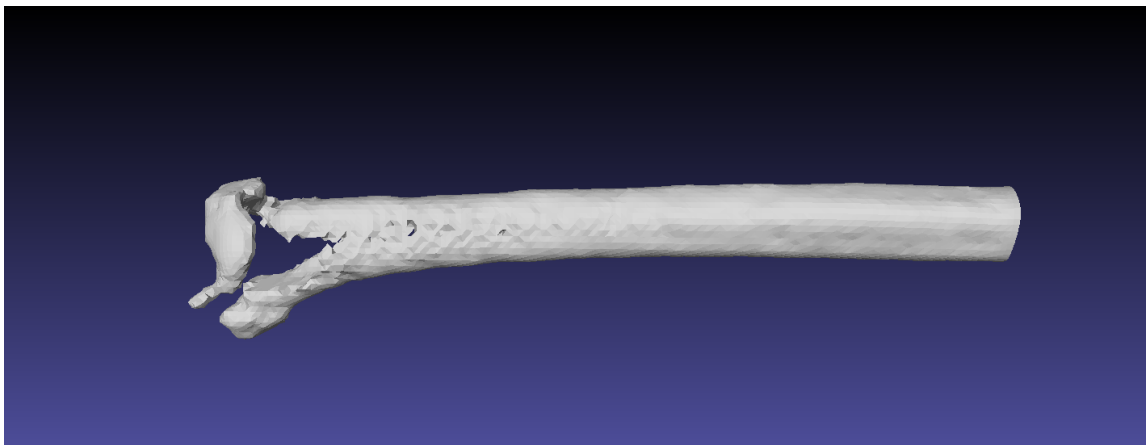


Figure 6.8: Segmented Femur Condyle

Table 6.5: Performance analysis for Femur at knee Joint based on UNet deep learning model.

Patient ID	Precision	Recall	F1 Score
78	0.8561	0.724	0.7845
82	0.7645	0.6238	0.6871
88	0.838	0.74577	0.78921
81	0.7397	0.49336	0.5919
102	0.8274	0.7093	0.7638
103	0.9042	0.6865	0.7805

6.4 Model IV: 3D UNet

3D-UNet is a contractive and expanding path composition that aims to build a bottleneck in its centermost part through convolution and pooling operations. After this bottleneck, the image is formed through convolutions and upsampling.

The computational requirement of 3D UNet is very high. For the work to be completed successfully, the developed segmentation algorithm method has multiple steps to automatically crop background voxels to reduce the image size around the desired area. The first step to find landmark is developed in Chapter 5. Fig. 6.9 shows the flowchart depiction of the algorithm for the approach to do figure shows our proposed model, in which 3D slices volume is first processed through landmark prediction model, after locating landmarks 3D volume will be automatically cropped around the location to decrease the size of the image, then 3D segmentation is performed on cropped data using 3D UNet, Fig. (6.10) represent our 3D Model architecture. We have performed 3D UNet on the Femur head and Distal femur.

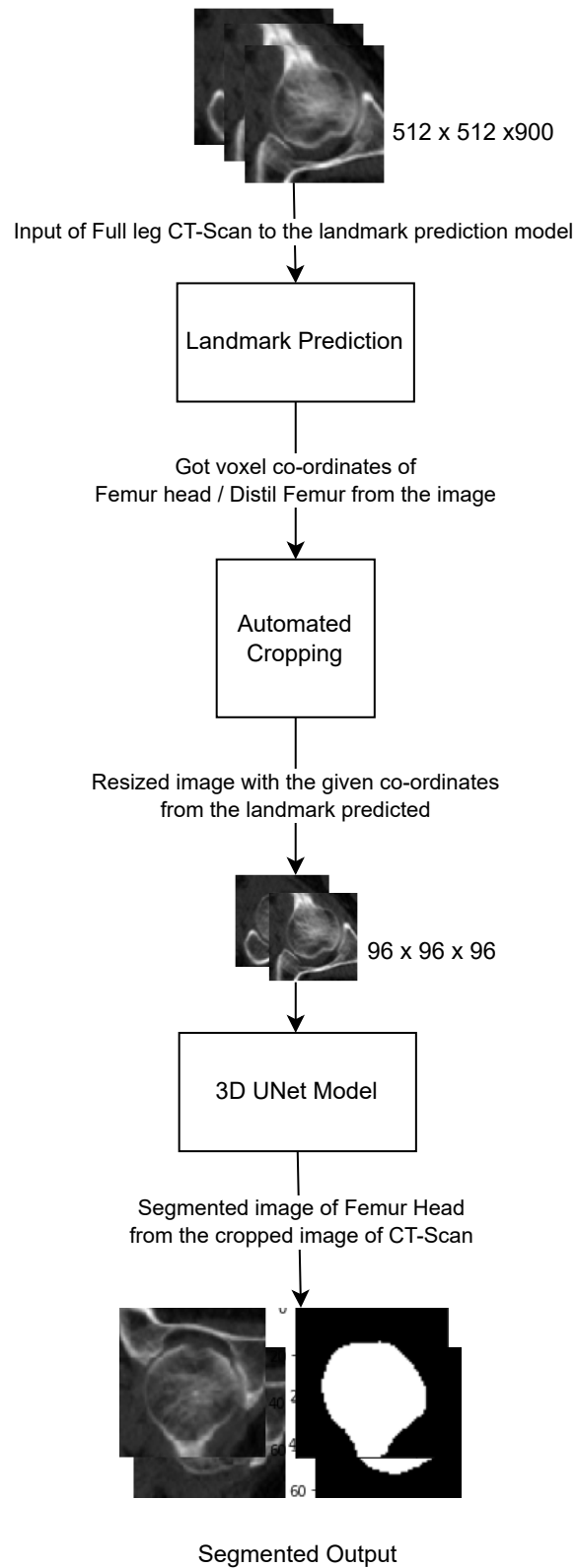


Figure 6.9: A flowchart showing the workflow of our proposed method

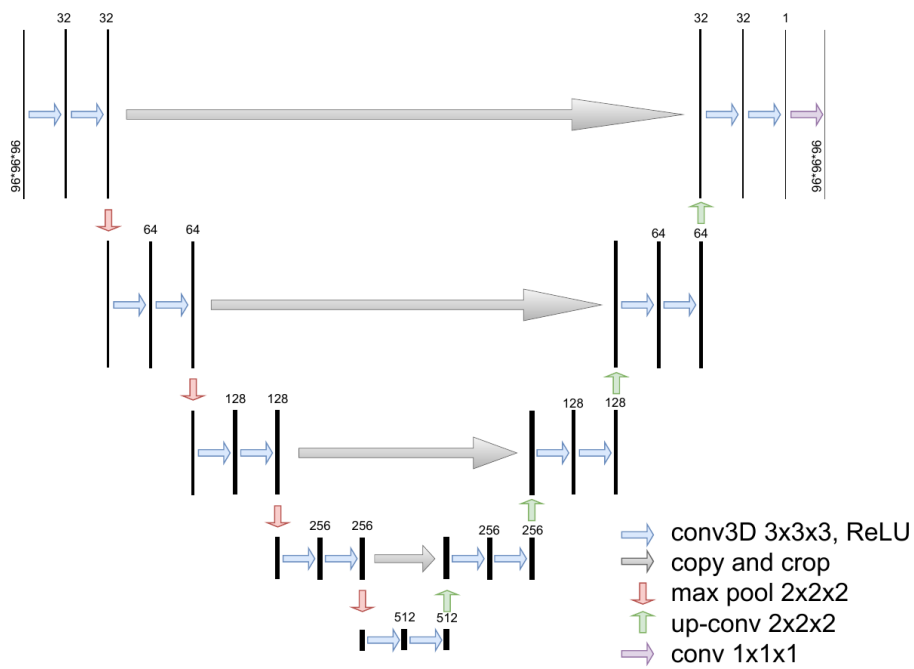


Figure 6.10: A schematic view of our proposed 3D U-Net model

6.4.1 Femur Head

The distal femur is more complex than the femur head, as, in the distal femur, there are more bones and different angles between bones. In Fig. (6.12), 1st column is of CT scan slices, 2nd column is of ground truth(mask), and the last column is the segmented output from 3D UNet, where one can see the hip bone is not segmented. A 3D rendering of the segmented structure with 3D U-Net is shown in Fig. (6.11).

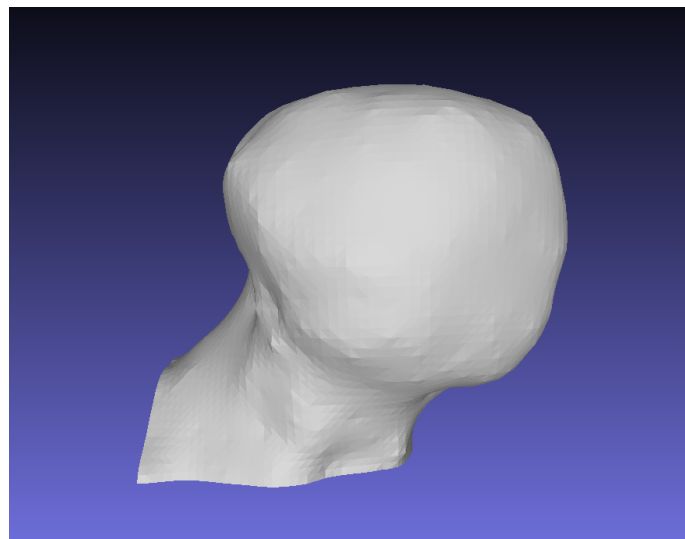


Figure 6.11: Segmentation result for femur head on a sample test patient.

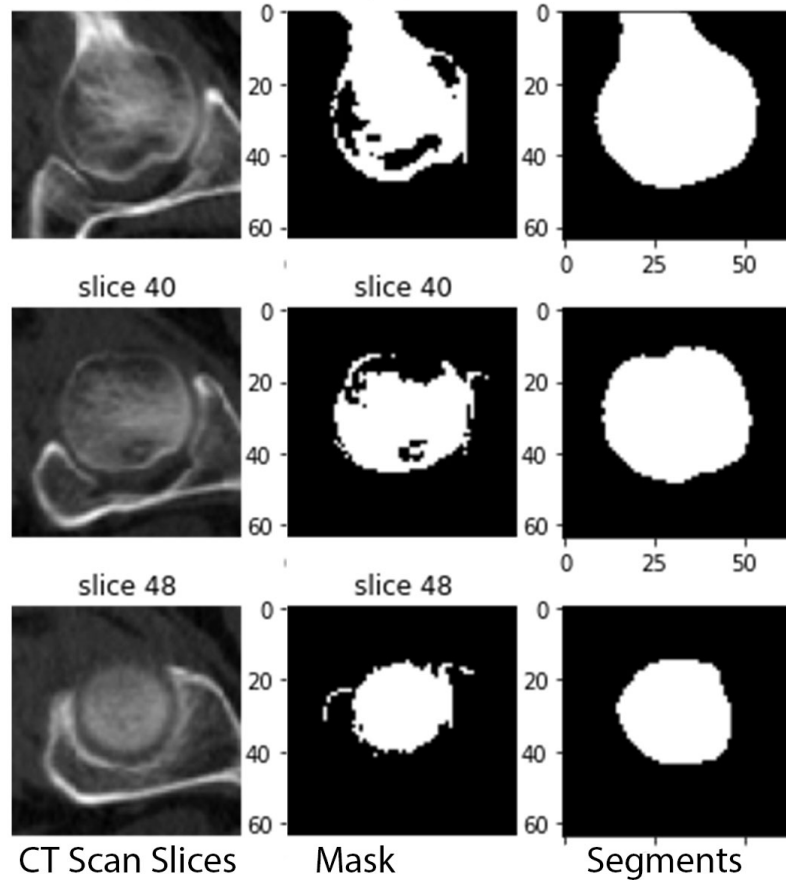


Figure 6.12: Slice representation of femur head

Table 6.6 One can see the performance matrix of the 3D UNet model on the Femur head. Due to hole filling, False positive increases, which leads to less Precision. Higher Recall shows higher True positives.

Table 6.6: Performance analysis of 3D UNet on Femoral head.

Patient Id	Average Precision	Recall	F1 Score
78	0.63	0.94	0.76
82	0.66	0.91	0.77
88	0.75	0.8	0.77
81	0.51	0.93	0.66
102	0.56	0.93	0.7
103	0.8	0.93	0.86

6.4.2 Distal Femur

The distal femur is more complex than the femur head in the distal femur, and there are more bones and different angles between bones. In figure 6.13 1st column is of CT scan slices, 2nd column is of ground truth(mask), and in the last

column is the segmented output from 3D UNet, where one can see patella bones is not segmented. Table 6.7 shows the performance matrix of 3D UNet on the Distal femur. A 3D rendering of the segmented structure with 3D U-Net is shown in Fig. (6.14).

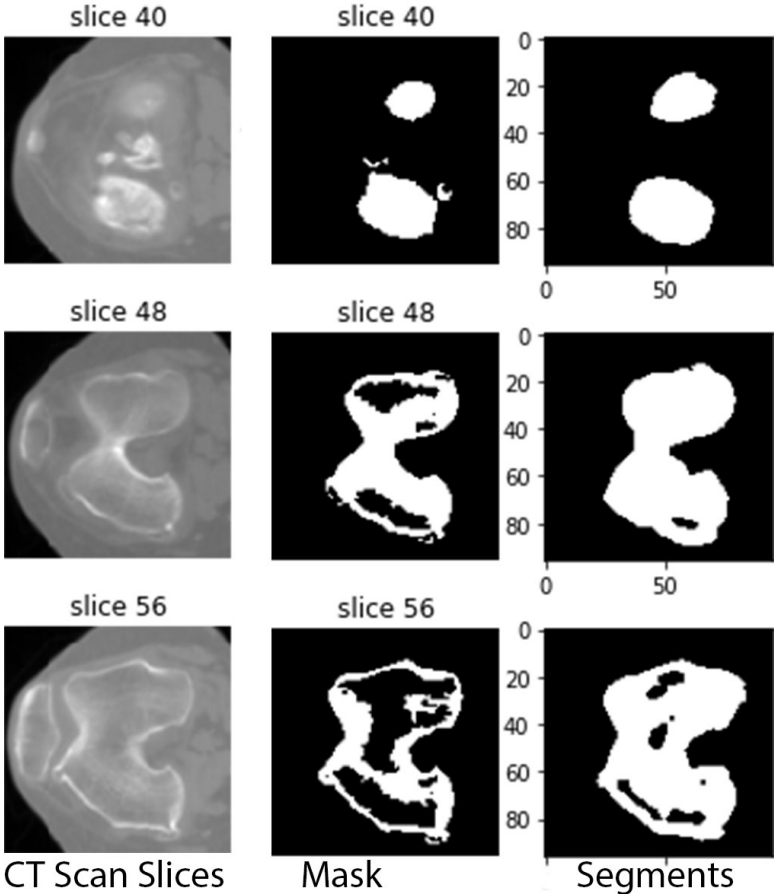


Figure 6.13: Slice representation at distal femur.

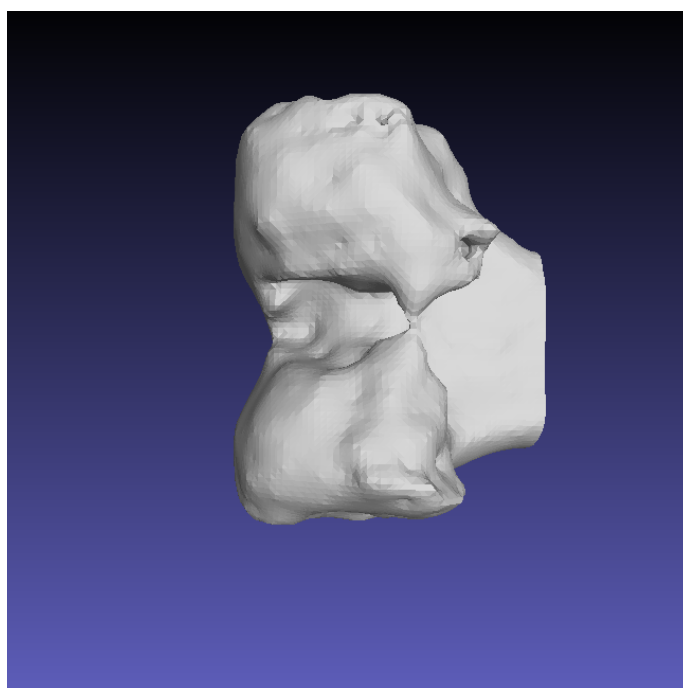


Figure 6.14: Segmentation result for femur head on a sample test patient.

Table 6.7: Performance analysis of 3D UNet on Distal Femur.

Patient Id	Average Precision	Recall	F1 Score
185	0.77	0.84	0.80
200	0.84	0.92	0.88
192	0.62	0.84	0.71
207	0.81	0.92	0.86
224	0.85	0.93	0.89
205	0.76	0.88	0.81
198	0.86	0.92	0.89
190	0.83	0.96	0.89
175	0.85	0.86	0.86
206	0.83	0.92	0.88
167	0.70	0.94	0.81
193	0.84	0.9	0.87
182	0.82	0.92	0.87
171	0.65	0.81	0.72
204	0.81	0.92	0.87
199	0.81	0.88	0.84
201	0.78	0.95	0.86
172	0.76	0.9	0.83
197	0.84	0.92	0.88

6.5 Results and Analysis

The inner part of bones is less dense than the outer layer of bones, so in fixed threshold methods, voxels with less dense bones are not considered as bones and are treated as background, whereas deep learning-based methods treat these less dense bones voxels as bone.

The results in Table 6.8 show a comparison with all our approaches. Here, the F1 score of Standard segmentation is the best of all our models. Our data annotation strategy is based on fixed threshold techniques, and standard segmentation is also based on fixed threshold techniques, which decreases the chance of false positives. Other methods have more false positives as many voxels that are background as per our ground truth, but methods predicted these voxels as bones, resulting in less F1 score. In Fig. (6.13), it can be visually verified that the middle column is the ground truth which has many background pixels in between the bones, and the rightmost column, which is of 3D U-Net model prediction, have the whole part as bone, which leads to many false positives by which lower F1 score.

Table 6.8: Performance measure of all our approaches.

Methods	Average Precision	Recall	F1 Score
Standard Segmentation	0.878 ± 0.025	0.989 ± 0.004	0.930 ± 0.013
Local Otsu Segmentation	0.485 ± 0.037	0.975 ± 0.005	0.648 ± 0.034
2D UNet Bone	0.882 ± 0.025	0.778 ± 0.061	0.826 ± 0.045
2D UNet Distal Femur	0.822 ± 0.055	0.664 ± 0.085	0.733 ± 0.072
2D UNet Femur head	0.881 ± 0.008	0.768 ± 0.051	0.820 ± 0.033
3D UNet Distal Femur	0.52 ± 0.12	0.89 ± 0.03	0.65 ± 0.1
3D UNet Femur head	0.81 ± 0.04	0.90 ± 0.03	0.84 ± 0.03

Fig. (6.15) is a visual representation of the femur bone segmentation by the approach of [39]. They claim that the femur is segmented from at least one of its adjacent bones in 57% of the cases. Due to their unavailability of the dataset, we could not try our approach on the same data set. However, using our best approach 3D UNet it is possible to segment the femur from all the adjacent bones in all cases. Fig. (6.16) and Fig. (6.17) is a visual representation of the femur segmentation using our approach, and one can see that our model can segment femur from other bones(patella, tibia, pelvis bones). Our 3D U-Net approach shows a recall value of 0.9 as compared to $\text{TPR} > 0.85$ of model [39], which shows our model can discriminate the actual femur voxels with more satisfactory precision correctly.

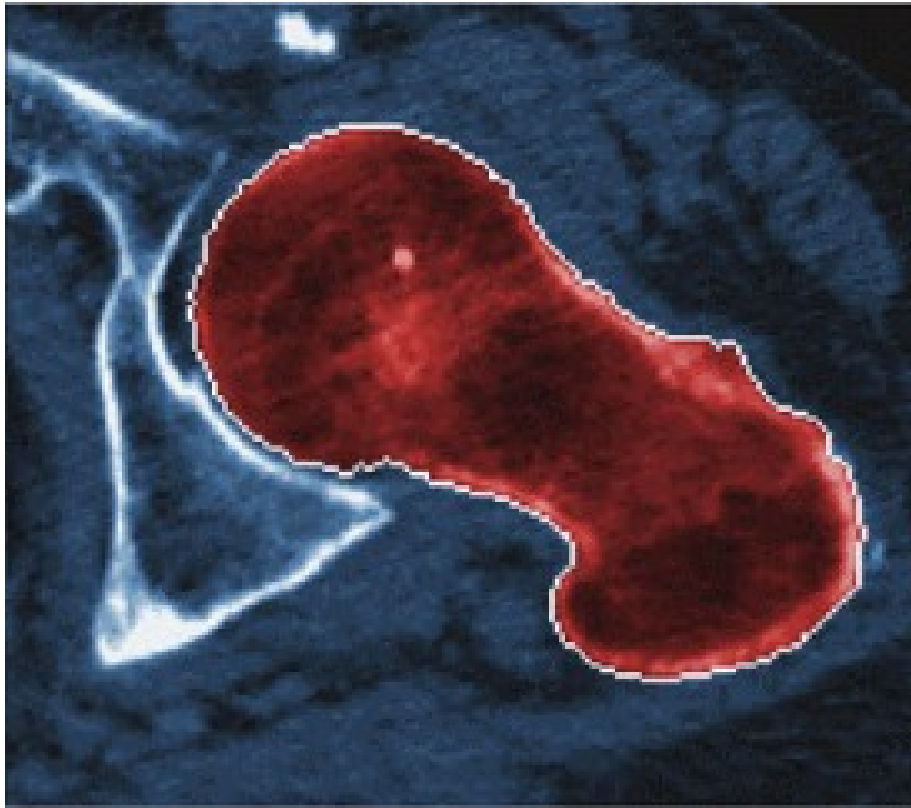


Figure 6.15: Segmentation result of Model [39] on axial slice of Femur head

In [7] the authors employed the Local Otsu approach to identify background voxels and allow automatic bone cutting from a 3D image to lower the size and hence speed up neural network execution. We have used the Landmark prediction model to roughly identify the different parts of the femur. Furthermore, allow the automatic cropping of any part of the femur, by which we can even decrease the size of the 3d image even more than [7], which will reduce the time of execution and computational requirement.

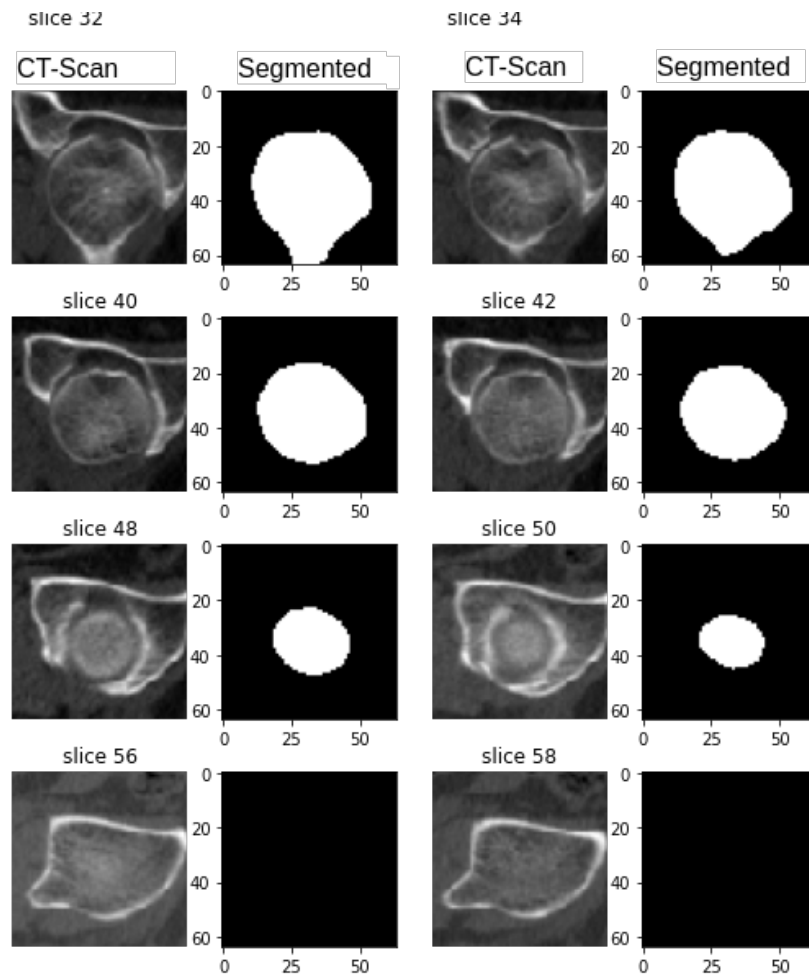


Figure 6.16: Segmentation result of 3D UNet model on axial slice of Femur head

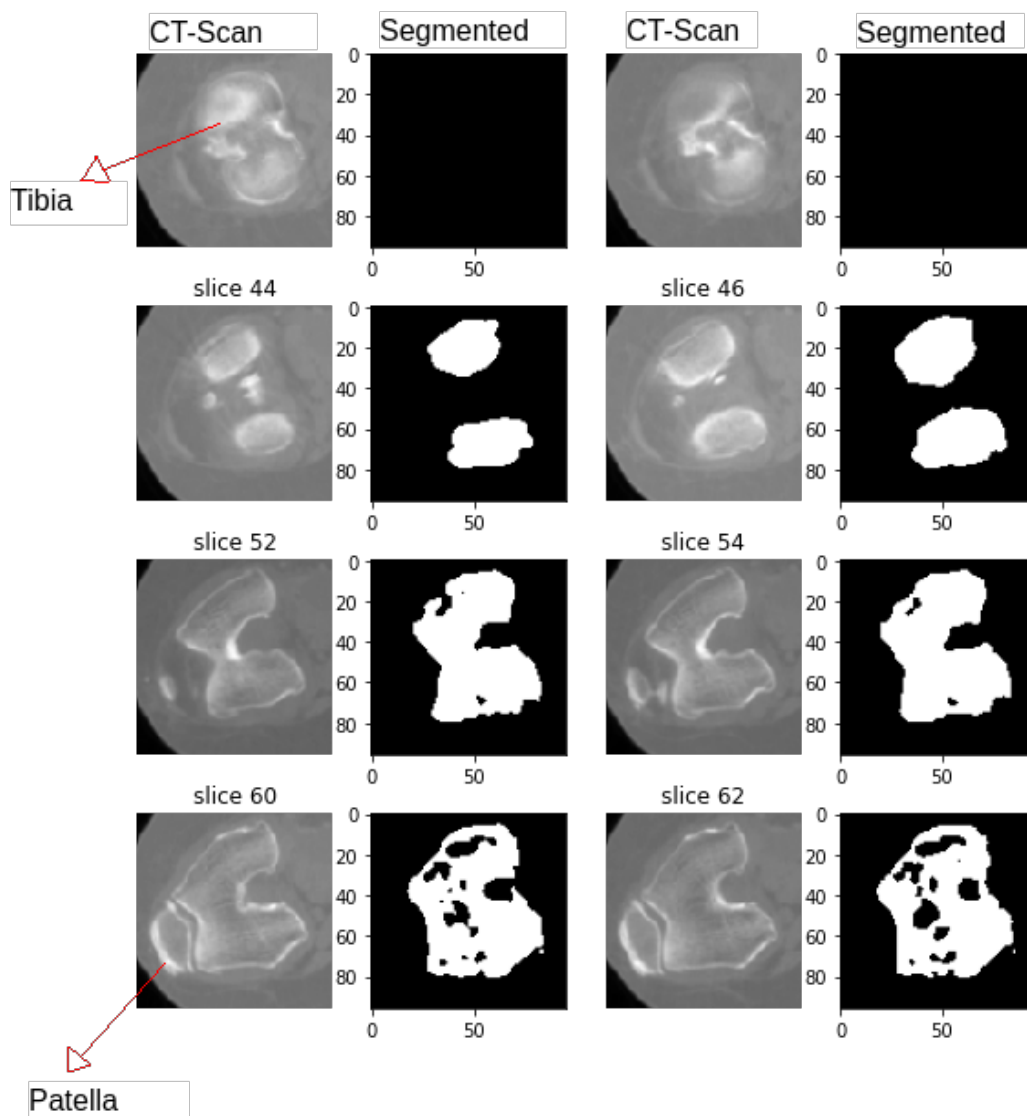


Figure 6.17: Segmentation result of 3D UNet model on axial slice of Distal femur

CHAPTER 7

Conclusion

The objective of this thesis is to design an efficient architecture for segmentation of bone, femur head, and distal femur which should be effective and which should do the needful in an automated way. Usually a CT scan is a 3D image of large size and bones are just a small part of this image. In case of an OA patient, the ortho surgeon is interested in just this small part of the CT scan image. Therefore the first task that we dealt with was cropping the full image and extracting just the part which contained the bones. Once it is done, we drew the landmark of the femur followed by the segmentation of bones. Cropping of the image obviously reduces the size of the image and hence needs less computational resources to carry out the later processes. And the automated system enables the surgeon analyse the patient's condition without any unnecessary human intervention.

In this thesis, we performed labeling of femur, patella, fibula and tibia from 40 CT scan images with full stack of corresponding slices of 2D images. Since these images were provided to us by a local ortho surgeon, we created its in-house ground truth using an open source 3D slicer tool. For bone segmentation we used four techniques of bone segmentation, namely, 2D U-Net, 3D U-Net, Otsu's and Fixed threshold. We noted that though all these techniques segmented the bones successfully, they were deficient in identifying or differentiating the bone types. However, our objective was to identify the bones also. For this we implemented a deep learning based technique, i.e., U-Net. We presented a fully automatic framework based on U-Net for bone segmentation. We presented the effectiveness of our algorithm using U-Net and found it satisfactory.

The proposed framework consists of 2 stages, The first stage of this framework consists of 2D U-Net architecture for landmark prediction of the distal femur and femoral head in the original 3D CT scans. The next stage of the framework is to perform local 3D segmentation around the landmark. The first stage of the framework reduces the size of 3D volume from 512 X 512 X 1000 to 96 X 96 X 96, which reduces the requirement of computational resources and time for segmentation.

We also demonstrated and performed tissue and femur bone segmentation using various methods such as fixed threshold segmentation, U-Net segmentation, and Otsu's based segmentation. Further focusing on the femur bone of CT scans of the knee joint, we had used deep learning method the U-Net as segmentation model. We had also did the tissue segmentation using other segmentation methods based on Grey level feature methods and evaluated with our in house ground truth.

However, this study has limitations with our ground truth, and with a few adjustments results can be improved. First, the data used for training is limited to 25 full leg CT scans, validation is limited to 9 full leg CT scans and then model is tested on 6 full leg Ct scans. Therefore with a larger training dataset, performance of our approach are expected to be improved. This framework can be extended to locate anatomical landmarks and segmentation of other bones or organs' medical images with different modalities in future work.

References

- [1] 3d slicer image computing platform. <https://www.slicer.org/>.
- [2] P. U. Adiga and B. Chaudhuri. An efficient method based on watershed and rule-based merging for segmentation of 3-d histo-pathological images. *Pattern recognition*, 34(7):1449–1458, 2001.
- [3] A. Al-Zaghal, D. P. Yellanki, C. Ayubcha, and A. A. M. Hon. Ct-based tissue segmentation to assess knee joint inflammation and reactive bone formation assessed by 18 18 f-fdg and f-naf pet/ct: Effects of age and bmi. *Hellenic journal of nuclear medicine*, 21(2):102–107, 2018.
- [4] V. Badrinarayanan, A. Kendall, and R. Cipolla. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE transactions on pattern analysis and machine intelligence*, 39(12):2481–2495, 2017.
- [5] A. Bieniek and A. Moga. An efficient watershed algorithm based on connected components. *Pattern recognition*, 33(6):907–916, 2000.
- [6] C. M. Bishop and N. M. Nasrabadi. *Pattern recognition and machine learning*, volume 4. Springer, 2006.
- [7] P. A. Bjornsson, B. Helgason, H. Palsson, S. Sigurdsson, V. Gudnason, and L. M. Ellingsen. Automated femur segmentation from computed tomography images using a deep neural network. In B. S. Gimi and A. Krol, editors, *Medical Imaging 2021: Biomedical Applications in Molecular, Structural, and Functional Imaging*. SPIE, feb 2021.
- [8] G. J. Brostow, J. Shotton, J. Fauqueur, and R. Cipolla. Segmentation and recognition using structure from motion point clouds. In *European conference on computer vision*, pages 44–57. Springer, 2008.
- [9] A. J. Burghardt, G. J. Kazakia, and S. Majumdar. A local adaptive threshold strategy for high resolution peripheral quantitative computed tomography of trabecular bone. *Annals of biomedical engineering*, 35(10):1678–1686, 2007.

- [10] J. T. Bushberg and J. M. Boone. *The essential physics of medical imaging*. Lippincott Williams & Wilkins, 2011.
- [11] T. M. Buzug. Computed tomography: from photon statistics to modern cone-beam ct. *Soc Nuclear Med*, 2009.
- [12] J. Calder, A. M. Tahmasebi, and A.-R. Mansouri. A variational approach to bone segmentation in ct images. In *Medical Imaging 2011: Image Processing*, volume 7962, page 79620B. International Society for Optics and Photonics, 2011.
- [13] K. P. Chandar and T. Satyasavithri. Segmentation and 3d visualization of pelvic bone from ct scan images. In *2016 IEEE 6th International Conference on Advanced Computing (IACC)*, pages 430–433. IEEE, 2016.
- [14] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille. Semantic image segmentation with deep convolutional nets and fully connected crfs. *arXiv preprint arXiv:1412.7062*, 2014.
- [15] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger. 3d u-net: learning dense volumetric segmentation from sparse annotation. In *International conference on medical image computing and computer-assisted intervention*, pages 424–432. Springer, 2016.
- [16] D. Cireşan, A. Giusti, L. Gambardella, and J. Schmidhuber. Deep neural networks segment neuronal membranes in electron microscopy images. *Advances in neural information processing systems*, 25, 2012.
- [17] W. Dai, X. Liang, H. Zhang, E. Xing, and J. Doyle. Structure correcting adversarial network for chest x-rays organ segmentation, June 30 2020. US Patent 10,699,412.
- [18] Y. Deng, L. Wang, C. Zhao, S. Tang, X. Cheng, H.-W. Deng, and W. Zhou. A deep learning-based approach to automatic proximal femur segmentation in quantitative ct images. *Medical & biological engineering & computing*, pages 1–13, 2022.
- [19] T. D. DenOtter and J. Schubert. Hounsfield unit. 2019.
- [20] X. Descombes. Chapter 8 - markov models and mcmc algorithms in image processing. In R. Chellappa and S. Theodoridis, editors, *Academic Press Library in Signal Processing, Volume 6*, pages 305–344. Academic Press, 2018.

- [21] V. Duay, N. Houhou, and J.-P. Thiran. Atlas-based segmentation of medical images locally constrained by level sets. In *IEEE International Conference on Image Processing 2005*, volume 2, pages II–1286. IEEE, 2005.
- [22] A. Fedorov, J. Johnson, E. Damaraju, A. Ozerin, V. Calhoun, and S. Plis. End-to-end learning of brain tissue segmentation from imperfect labeling. In *2017 International Joint Conference on Neural Networks (IJCNN)*, pages 3785–3792. IEEE, 2017.
- [23] Y. Fu, S. Liu, H. H. Li, and D. Yang. Automatic and hierarchical segmentation of the human skeleton in ct images. *Physics in Medicine & Biology*, 62(7):2812, 2017.
- [24] R. C. Gonzalez and R. E. Woods. *Digital Image Processing*. Addison-Wesley Longman Publishing Co., Inc., USA, 2nd edition, 2001.
- [25] I. Goodfellow, Y. Bengio, and A. Courville. *Deep Learning*. MIT Press, 2016. <http://www.deeplearningbook.org>.
- [26] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial nets. *Advances in neural information processing systems*, 27, 2014.
- [27] V. Grau, A. Mewes, M. Alcaniz, R. Kikinis, and S. K. Warfield. Improved watershed transform for medical image segmentation using prior information. *IEEE transactions on medical imaging*, 23(4):447–458, 2004.
- [28] 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*, pages 770–778, 2016.
- [29] K. Held, E. R. Kops, B. J. Krause, W. M. Wells, R. Kikinis, and H.-W. Muller-Gartner. Markov random field segmentation of brain mr images. *IEEE transactions on medical imaging*, 16(6):878–886, 1997.
- [30] S. Hong, H. Noh, and B. Han. Decoupled deep neural network for semi-supervised semantic segmentation. *Advances in neural information processing systems*, 28, 2015.
- [31] H. Hu. Multi-slice helical ct: scan and reconstruction. *Medical physics*, 26(1):5–18, 1999.

- [32] Y. Huo, J. G. Terry, J. Wang, S. Nair, T. A. Lasko, B. I. Freedman, J. J. Carr, and B. A. Landman. Fully automatic liver attenuation estimation combining CNN segmentation and morphological operations. *Medical Physics*, 46(8):3508–3519, jul 2019.
- [33] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell. Caffe: Convolutional architecture for fast feature embedding. In *Proceedings of the 22nd ACM international conference on Multimedia*, pages 675–678, 2014.
- [34] A. Jose, S.Ravi, and M.Sambath. Brain tumor segmentation using k-meansclustering and fuzzy c-means algorithmsand its area calculation. *International Journal of Innovative Research in Computer and Communication Engineering*, 2:3496–3501, 2014.
- [35] Y. Kang, K. Engelke, and W. A. Kalender. A new accurate and precise 3-d segmentation method for skeletal structures in volumetric ct data. *IEEE transactions on medical imaging*, 22(5):586–598, 2003.
- [36] R. Kikinis, S. D. Pieper, and K. G. Vosburgh. *3D Slicer: A Platform for Subject-Specific Image Analysis, Visualization, and Clinical Support*, pages 277–289. Springer New York, New York, NY, 2014.
- [37] T. Klinder, C. Lorenz, J. v. Berg, S. P. Dries, T. Bülow, and J. Ostermann. Automated model-based rib cage segmentation and labeling in ct images. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 195–202. Springer, 2007.
- [38] S. Kohl, D. Bonekamp, H.-P. Schlemmer, K. Yaqubi, M. Hohenfellner, B. Hadaschik, J.-P. Radtke, and K. Maier-Hein. Adversarial networks for the detection of aggressive prostate cancer. *arXiv preprint arXiv:1702.08014*, 2017.
- [39] M. Krčah, G. Székely, and R. Blanc. Fully automatic and fast segmentation of the femur bone from 3d-ct images with no shape prior. In *2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pages 2087–2090, 2011.
- [40] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. W. M. van der Laak, B. van Ginneken, and C. I. Sánchez. A survey on deep learning in medical image analysis. *Med. Image Anal.*, 42:60–88, Dec. 2017.

- [41] S. Liu. Image segmentation technology of the ostu method for image materials based on binary pso algorithm. In D. Jin and S. Lin, editors, *Advances in Computer Science, Intelligent System and Environment*, pages 415–419, Berlin, Heidelberg, 2011. Springer Berlin Heidelberg.
- [42] X. Liu, C. Han, H. Wang, J. Wu, Y. Cui, X. Zhang, and X. Wang. Fully automated pelvic bone segmentation in multiparametric mri using a 3d convolutional neural network. *Insights into imaging*, 12(1):1–13, 2021.
- [43] J. Long, E. Shelhamer, and T. Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440, 2015.
- [44] P. Luc, C. Couprie, S. Chintala, and J. Verbeek. Semantic segmentation using adversarial networks. *arXiv preprint arXiv:1611.08408*, 2016.
- [45] S. Madhukumar and N. Santhiyakumari. Evaluation of k-means and fuzzy c-means segmentation on mr images of brain. *The Egyptian Journal of Radiology and Nuclear medicine*, 46:475–479, 2015.
- [46] I. Manousakas, P. Undrill, G. Cameron, and T. Redpath. Split-and-merge segmentation of magnetic resonance medical images: performance evaluation and extension to three dimensions. *Computers and Biomedical Research*, 31(6):393–412, 1998.
- [47] F. Milletari, S.-A. Ahmadi, C. Kroll, A. Plate, V. Rozanski, J. Maiostre, J. Levin, O. Dietrich, B. Ertl-Wagner, K. Bötzel, et al. Hough-cnn: deep learning for segmentation of deep brain regions in mri and ultrasound. *Computer Vision and Image Understanding*, 164:92–102, 2017.
- [48] 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)*, pages 565–571. IEEE, 2016.
- [49] J. Minnema, M. van Eijnatten, W. Kouw, F. Diblen, A. Mendrik, and J. Wolff. CT image segmentation of bone for medical additive manufacturing using a convolutional neural network. *Comput. Biol. Med.*, 103:130–139, Dec. 2018.
- [50] P. Moeskops, M. Veta, M. W. Lafarge, K. A. Eppenhof, and J. P. Pluim. Adversarial training and dilated convolutions for brain mri segmentation. In *Deep learning in medical image analysis and multimodal learning for clinical decision support*, pages 56–64. Springer, 2017.

- [51] P. Moeskops, J. M. Wolterink, B. H. van der Velden, K. G. Gilhuijs, T. Leiner, M. A. Viergever, and I. Išgum. Deep learning for multi-task medical image segmentation in multiple modalities. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 478–486. Springer, 2016.
- [52] H. Noh, S. Hong, and B. Han. Learning deconvolution network for semantic segmentation. In *Proceedings of the IEEE international conference on computer vision*, pages 1520–1528, 2015.
- [53] C. P. Pal, P. Singh, S. Chaturvedi, K. K. Pruthi, and A. Vij. Epidemiology of knee osteoarthritis in india and related factors. *Indian J. Orthop.*, 50(5):518–522, Sept. 2016.
- [54] G. Papandreou, L.-C. Chen, K. P. Murphy, and A. L. Yuille. Weakly-and semi-supervised learning of a deep convolutional network for semantic image segmentation. In *Proceedings of the IEEE international conference on computer vision*, pages 1742–1750, 2015.
- [55] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32, 2019.
- [56] Y. Pauchard, T. Fitze, D. Browarnik, A. Eskandari, I. Pauchard, W. Enns-Bray, H. Pálsson, S. Sigurdsson, S. J. Ferguson, T. B. Harris, V. Gudnason, and B. Helgason. Interactive graph-cut segmentation for fast creation of finite element models from clinical ct data for hip fracture prediction. *Comput. Methods Biomech. Biomed. Engin.*, 19(16):1693–1703, Dec. 2016.
- [57] C. Rieng. F1 score = dice coefficient. <https://chenriang.me/f1-equal-dice-coefficient.html>, Jun 2020.
- [58] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pages 234–241, Cham, 2015. Springer International Publishing.
- [59] C. Rorden. An open source tool for landmark identification. <https://people.cas.sc.edu/rorden/micro/>.

- [60] M. Shao, S. Han, A. Carass, X. Li, A. M. Blitz, J. Shin, J. L. Prince, and L. M. Ellingsen. Brain ventricle parcellation using a deep neural network: Application to patients with ventriculomegaly. *NeuroImage: Clinical*, 23:101871, 2019.
- [61] N. Sharma and L. M. Aggarwal. Automated medical image segmentation techniques. *J. Med. Phys.*, 35(1):3–14, Jan. 2010.
- [62] A. A. Taha and A. Hanbury. Metrics for evaluating 3d medical image segmentation: analysis, selection, and tool. *BMC medical imaging*, 15(1):1–28, 2015.
- [63] P. A. Thakur. Brain tumor segmentation using k-means clustering algorithm. *International Journal for Research in Applied Science and Engineering Technology*, 10(2):1–8, 2022.
- [64] T. P. Trappenberg. *Fundamentals of machine learning*. Oxford University Press, 2019.
- [65] P. Ubolkosold and P. Tiawongsombat. A study of otsu’s local optimal thresholding algorithm. 04 2014.
- [66] L. Vincent and P. Soille. Watersheds in digital spaces: an efficient algorithm based on immersion simulations. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 13(06):583–598, 1991.
- [67] M.-S. Yang, Y. J. Hu, K. C.-R. Lin, and C. C.-L. Lin. Segmentation techniques for tissue differentiation in mri of ophthalmology using fuzzy clustering algorithms. *Magnetic resonance imaging*, 20 2:173–9, 2002.
- [68] F. Yokota, T. Okada, M. Takao, N. Sugano, Y. Tada, and Y. Sato. Automated segmentation of the femur and pelvis from 3d ct data of diseased hip using hierarchical statistical shape model of joint structure. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 811–818. Springer, 2009.
- [69] J. Zhang, C.-H. Yan, C.-K. Chui, and S.-H. Ong. Fast segmentation of bone in ct images using 3d adaptive thresholding. *Computers in biology and medicine*, 40(2):231–236, 2010.