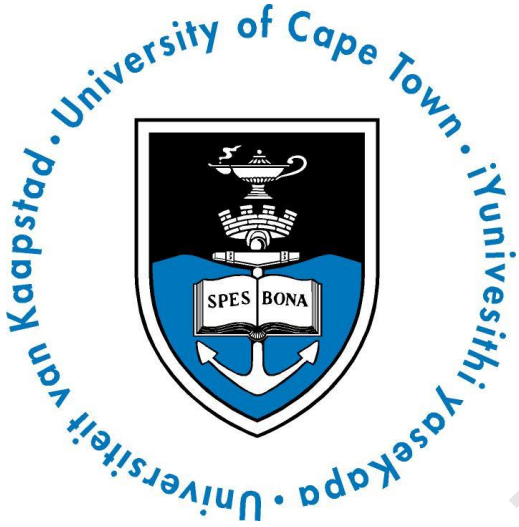

A deep learning algorithm for contour detection in synthetic 2D bi-planar X-ray images of the scapula: towards improved 3D reconstruction of the scapula

Submitted to the University of Cape Town in fulfilment of the academic requirements for the degree of MSc in Biomedical Engineering by dissertation.

Catherine Namayega (NMYCAT001)

Supervisor:

Dr. Tinashe Mutsvangwa

Co-supervisors:

Dr. Bessie Malila and Professor Tania Douglas

August 31, 2020

Declaration

I, **Catherine Namayega**, hereby declare that the work on which this dissertation/thesis is based is my original work (except where acknowledgements indicate otherwise) and that neither the whole work nor any part of it has been, is being, or is to be submitted for another degree in this or any other university. I empower the university to reproduce for research either the whole or any portion of the contents in any manner whatsoever.

Signature: Signed by candidate Date: ……August 31, 2020……

Abstract

Three-dimensional (3D) reconstruction from X-ray images using statistical shape models (SSM) provides a cost-effective way of increasing the diagnostic utility of two-dimensional (2D) X-ray images, especially in low-resource settings. The landmark-constrained model fitting approach is one way to obtain patient-specific models from a statistical model. This approach requires an accurate selection of corresponding features, usually landmarks, from the bi-planar X-ray images. However, X-ray images are 2D representations of 3D anatomy with super-positioned structures, which confounds this approach. The literature shows that detection and use of contours to locate corresponding landmarks within bi-planar X-ray images can address this limitation. The aim of this research project was to train and validate a deep learning algorithm for detection the contour of a scapula in synthetic 2D bi-planar X-ray images.

Synthetic bi-planar X-ray images were obtained from scapula mesh samples with annotated landmarks generated from a validated SSM obtained from the Division of Biomedical Engineering, University of Cape Town. This was followed by the training of two convolutional neural network models as the first objective of the project; the first model was trained to predict the lateral (LAT) scapula image given the anterior-posterior (AP) image. The second model was trained to predict the AP image given the LAT image. The trained models had an average Dice coefficient value of 0.926 and 0.964 for the predicted LAT and AP images, respectively. However, the trained models did not generalise to the segmented real X-ray images of the scapula. The second objective was to perform landmark-constrained model fitting using the corresponding landmarks embedded in the predicted images. To achieve this objective, the 2D landmark locations were transformed into 3D coordinates using the direct linear transformation. The 3D point localization yielded average errors of (0.35, 0.64, 0.72) mm in the X, Y and Z directions, respectively, and a combined coordinate error of 1.16 mm. The reconstructed landmarks were used to reconstruct meshes that had average surface-to-surface distances of 3.22 mm and 1.72 mm for 3 and 6 landmarks, respectively. The third objective was to reconstruct the scapula mesh using matching points on the scapula contour in the bi-planar images. The average surface-to-surface distances of the reconstructed meshes with 8 matching contour points and 6 corresponding landmarks of the same meshes were 1.40 and 1.91 mm, respectively.

In summary, the deep learning models were able to learn the mapping between the bi-planar images of the scapula. Increasing the number of corresponding landmarks from the bi-planar images resulted into better 3D reconstructions. However, obtaining these corresponding landmarks was non-trivial, necessitating the use of matching points selected from the scapulae contours. The results from the latter approach signal a need to explore contour matching methods to obtain more corresponding points in order to improve the scapula 3D reconstruction using landmark-constrained model fitting.

Acknowledgements

My sincere thanks to the African Biomedical Engineering Mobility (ABEM) project, under the Intra-Africa Academic Mobility Scheme of the European Commission's Education, Audio-visual and Culture Executive Agency, and the South African Research Chairs Initiative of the Department of Science and Technology and the National Research Foundation in South Africa (grant no. 98788), for funding this research.

I would also like to express my sincere gratitude to my supervisor Dr Tinashe Mutsvangwa and my co-supervisors Professor Tania Douglas and Dr Bessie Malila for the endless support through their time, dedication, vast knowledge, patience, words of wisdom and motivation during this research study. I could not have imagined having better supervisors and advisers for my master's study.

Besides, my supervisors and co-supervisors, I would like to thank my mentor Dr. Robert Ssekitoleko for his dedication and endless words of wisdom that kept me hopeful though-out this time.

I would also like to thank all my research group mates in Medical Image Inferencing & Distributed Diagnostics (Mi2D2) Group: Ms X Thusini, Ms Y Karanja, Mr C Reyneke, Mr Jean R Fouefack, Dr. A Duniya, Mr T Belay, Mr F Atuhaire, Ms S Maclean, Ms T Dawood, Mr E Kamuhire and Mr K Majola, for the stimulating discussions and assistance.

In addition, I would like to thank my family and friends: Mr P Mulinde, Ms A Nanyonjo, Ms S Nalubwama, Ms M Tusabe, Ms B Arinda, Ms J Nantume, Ms M Mulerwa, Ms M Ninsiima, Ms M Nakyanzi, Ms H Nakayiza, Mr M Wowire, Mr M Kiwanuka, Ms J Nagawa, Mr E Egor, Mr A Enywaku, Mr H Yimam, Dr L Fryda and the ENC Baxter family for the love and support.

Finally, I am grateful to God for the good health and well-being that were necessary to complete this degree.

Table of contents

List of Figures

List of Tables

List of Abbreviations

AP	Anterior-posterior
3D	Three-dimensional
2D	Two-dimensional
CNN	Convolutional neural network
COP	Centre of projection
CT	Computer tomography
DLT	Direct linear transformation
DRR	Digitally reconstructed radiograph
EOS	Digital low-dose bi-plane X-ray imaging system
FCN	Fully connected network
GPA	Generalised Procrustes analysis
HREC	Human Research Ethics Committee
ICC	Intraclass correlation coefficient
LAT	Lateral
LAC	Linear attenuation coefficient
MLP	Multi-layer perception
MSE	Mean square error
PC	Principal component
PCA	Principal component analysis
ReLU	Rectified linear unit
SGD	Stochastic gradient decent
SSIM	Statistical shape and intensity model
SSM	Statistical Shape Model
SVD	Singular Value Decomposition

1 Introduction

Imaging of the skeletal system is mainly carried out using two-dimensional (2D) and three-dimensional (3D) imaging modalities. In resource-limited settings, 2D imaging modalities such as conventional X-ray imaging are the most commonly used means of imaging the skeletal structures of the human body (Muhogora & Pitcher, 2016). This is mainly because 2D modalities are cheaper to acquire, maintain and service compared to the 3D imaging modalities such as computed tomography (CT) and magnetic resonance imaging (MRI) (Mariani *et al.*, 2017; Yu *et al.*, 2016). However, for clinical diagnosis, it is difficult to view the surface of the target anatomy in a single 2D X-ray image due to the super-imposition of structures onto a single image plane (Shah *et al.*, 2014; Yu *et al.*, 2016). Furthermore, the inherent 3D nature of the skeletal anatomy makes an accurate assessment of its 2D representation largely dependent upon an expert's experience and ability to intrinsically assess the image (Laporte *et al.*, 2003; Mitulescu *et al.*, 2001; Mutsvangwa *et al.*, 2017).

In order to improve X-ray image interpretation, 2D bi-planar X-ray images (images lying in two different planes) may be used, although these images still do not provide details compared to the corresponding 3D images (Amirlak *et al.*, 2009; Chimhundu *et al.*, 2016; Melhem *et al.*, 2016; Yang *et al.*, 2016). Three-dimensional modalities such as CT and MRI scanners clearly show all elements of the structure and this improves image interpretation. However, these have other disadvantages apart from cost. Computed tomography exposes patients to a radiation dose over 40 times more than that from conventional X-ray imaging, thus increasing the risk of cancer exposure (Brenner & Hricak, 2010; Franco & Turgeon, 2010; Kim *et al.*, 2016; Linet *et al.*, 2009; Mettler *et al.*, 2008). While MRI has no radiation effect on patients, it presents challenges when imaging patients with metallic implants and does not give a distinct contrast between bone and air (Chan *et al.*, 2013; Chang *et al.*, 1987).

One way to improve medical image diagnosis using conventional X-ray without necessarily increasing radiation exposure and cost of imaging, could be 3D image reconstruction of 2D X-ray images (Diop & Burdin, 2013; Laporte *et al.*, 2003; Mutsvangwa *et al.*, 2017; Yu *et al.*, 2016). The introduction of low dose X-ray systems such as the Lodox Statscan (www.Iodox.com) and EOS X-ray imaging systems (www.eos-imaging.com) promise further reduction of radiation exposure in 3D from 2D reconstruction (Amirlak *et al.*, 2009; Melhem *et al.*, 2016). Other potential benefits include reduced cost of imaging and increased possibility for patient monitoring in pre and post-surgical assessment (Evangelopoulos *et al.*, 2009).

The 3D images can be reconstructed from the 2D bi-planar X-ray images using intensity-based and feature-based algorithms. Feature-based algorithms may leverage user-selected landmarks from the image as the prior information about the geometry. However, due to the overlap of contours of anatomical structures in the X-ray images, identifying corresponding landmark points in bi-planar

images in the region of interest is difficult. The nature of 2D bi-planar X-ray images leaves few clear corresponding features for 3D image reconstruction, leading to higher reconstruction errors. Researchers have proposed several innovations to overcome this challenge (Diop & Burdin, 2013; Leondes, 2003; Markelj *et al.*, 2012; Yu *et al.*, 2016). These include contour detection methods and the use of active contours with shape priors embedded before segmentation. However, most methods still depend on human expert delineation of the structure of interest due to low contrast of the X-ray images. In addition, most of these methods are not generalisable to all body structures and fail due to the presence of noise in the X-ray image. Human expert delineation introduces human bias making the methods subjective, error-prone, and not generalisable. Thus, this project was motivated by the need for alternative, less subjective, methods to detect the contours of structures of interest and locate corresponding landmarks.

Currently, medical imaging research especially contour detection, segmentation and 3D reconstruction are directed towards the use of deep learning approaches. Deep learning is a subclass of machine learning based on learning data representations with algorithms. Deep learning approaches have been shown to provide more consistent segmentation results; eliminating the complete reliance on the subjective judgement of human experts. This is due to the ability of a deep learning model to randomly initialize weights on filters and update itself with each training epoch. However, limited research has been done on contour detection in 2D bi-planar X-ray images. The use of contours as shape priors for extracting landmarks from 2D bi-planar X-ray images, may be a step towards more accurate reconstruction of 3D patient-specific bone geometry. Accurate 3D reconstruction from 2D X-ray images coupled with low dose X-ray imaging systems would reduce imaging costs, time and radiation exposure to the patient compared to imaging in CT (Cernazanu-Glavan & Stefan, 2013; Isin *et al.*, 2016; Long *et al.*, 2015; Middleton & Damper, 2004; Prasoon *et al.*, 2013; Ronneberger *et al.*, 2015; Simonyan & Zisserman, 2014; Wernick *et al.*, 2010; Zhang *et al.*, 2018).

1.1 Aim and objectives

This study aimed to train and validate a convolutional neural network (CNN)-based deep learning algorithm for detection of the contour of a scapula in synthetic 2D bi-planar X-ray images. To achieve the aim, the specific research objectives below were formulated.

1. Train a CNN algorithm to predict the contour of the scapula in the lateral image given the corresponding anterior-posterior image and vice versa.
2. Evaluate the performance of the trained algorithm using a landmark-constrained statistical shape model fitting of the embedded corresponding landmarks on the predicted contours.
3. Extract matching points from the predicted scapula contour in the bi-planar images and reconstruct the scapula mesh using landmark-constrained statistical shape model fitting.

1.2 General overview of the project

This research project aimed to train and validate a deep learning algorithm for contour detection in synthetic bi-planar X-ray images of the scapula. This was achieved in four main steps.

The first step was to generate synthetic data from a statistical shape model (SSM) of the scapula available in the Division of Biomedical Engineering, University of Cape Town. This was achieved through randomly generating mesh scapula samples with automatically annotated landmarks from the model. The generated meshes and landmark points were converted to Hounsfield volumes and projected to create landmarked bi-planar images of each scapula mesh. The projection of Hounsfield volumes was done using a digitally reconstructed radiograph (DRR) renderer develop in the Division of Biomedical Engineering at the University of Cape Town (Reyneke, 2019).

The second step was to train and evaluate a CNN model to predict the scapula contour in the lateral images given the anterior-posterior images and vice versa.

The third step was to evaluate the trained CNN models through landmark-constrained model fitting. This step was achieved by extracting the predicted corresponding scapula landmarks from the CNN predicted images. The extracted 2D landmarks were transformed to 3D coordinates using the direct linear transformation (DLT) (Abdel-Aziz *et al.*, 2015). The obtained 3D localised landmark points were used to constrain the SSM to predict a posterior model. The most likely reconstruction which was the mean of the posterior model was used as the reconstructed mesh and compared to the original mesh sample that was used to generate the DRRs.

The final step was to manually extract matching points from the predicted scapula contour in the bi-planar images. These extracted 2D matching points were transformed using DLT into 3D points and used to constrain the model. The predicted posterior from the constrained model was used to obtain the reconstructed mesh. The reconstructed meshes were compared to the original mesh samples and the results obtained compared to the results from the third step.

1.3 Scope and limitations

This research primarily used synthetic data for the development and validation of the scapula contour detection deep learning algorithm. Synthetic data in this research project refers to data generated from a statistical model and projected into 2D images using a DRR renderer. Digitally reconstructed radiographs are artificial X-ray images for visual simulation of the effect of real X-rays. A renderer is a virtual machine used to generate the DRRs. The use of synthetic data was important to avoid repetitive exposure to ionising radiation that have irreversible effects on the human body and to also aid the proof of concept. The trained models were tested using data of a single matching pair of cadaveric images of

X-ray and CT scans that were collected during a previous study (Wasswa, 2016). Although these were not sufficient to make clinical approximations of the 3D geometry of the scapula, they were suitable to test the concept for future recommendations on real datasets. Thus, to implement these algorithms on real X-ray bi-planar datasets, real images would have to be used in training the CNN models.

1.4 Ethical considerations

Ethics approval was granted by the Human Research Ethics Committee (HREC) of the University of Cape Town (reference number HREC REF: 100/2019). The ethical approval was granted for a single pair of the cadaveric upper-torso scans of both CT and X-ray collected during a previous research study (Wasswa, 2016).

1.5 Dissertation overview

This dissertation is divided into eight chapters; Chapter 2 presents a literature review leading to the identification of research gaps. Chapter 3 introduces the reader to the theory behind the concepts and ideas used in this research. Chapter 4 presents an overview of the research methodology overview, the tools used in the project, and the steps taken to generate datasets required to implement the objectives. Chapter 5 presents the first objective of the research project which was to train and evaluate a deep learning algorithm to predict the contour of the scapula in the bi-planar X-ray images. Chapter 6 presents the evaluation of objective 1 through landmark-constrained model fitting using known corresponding landmarks in the predicted images. Chapter 7 presents the extraction of matching points from scapula bi-planar images. This is followed by reconstruction of the scapula meshes using landmarks selected matching points of the contour. Finally, chapter 8 presents the overall summary of findings, conclusion and recommendation for future work.

2 Literature review

This chapter reviews the literature on three-dimensional (3D) reconstruction of X-ray images, contour detection in medical images and deep learning applications in medical imaging.

2.1 Related work on 3D from 2D image reconstruction

This section examines research on 3D from two-dimensional (2D) image reconstruction using a statistical models and landmarks, contour detection techniques and the application of deep learning approaches in medical imaging.

2.1.1 Landmark-constrained model fitting from X-ray images

A research study by Mutsvangwa *et al.* (2017) described a method for the 3D approximation of scapula bone shape from 2D bi-planar X-ray images using landmark-constrained statistical shape model (SSM) fitting. The method involved developing a virtual calibration frame to map the 2D image coordinates to their corresponding 3D real-world coordinates using X-ray stereo-photogrammetry. The 3D point reconstruction yielded an absolute reconstruction error of 0.19 mm. This was followed by assessing the scapula landmark reproducibility in bi-planar X-rays using inter and intra-observer landmark selection reliability evaluation (Ohl *et al.*, 2010). However, only 3 landmark points were identifiable from the 2D bi-planar images (the inferior angle, acromion, and coracoid). This is mainly due to the nature of the scapula orientation and the existence of the surrounding super-imposed structures.

The 3D scapula was reconstructed from the bi-planar images using the 3 corresponding landmarks. The 3D approximation of a scapula was done using an SSM built from training data made up of 84 computed tomography (CT) scapulae images. The scapula SSM was constrained with the 3D localised points of the selected reproducible 2D landmarks from the bi-planar X-ray images. The predicted posterior model was used to select the most likely instance given the 3 points. The selected instance was validated against a CT-derived ground-truth of the same cadaver resulting in a surface-to-surface average distance of 4.28 mm. Finally, a random instance of the SSM was 3D printed, embedded with 16 steel fiducial markers and imaged. Model reconstructions were performed using 3 and 16 landmarks from the bi-planar X-ray images of the printed instance. The 3D predictions of 3 and 16 landmarks were validated against the CT ground truth. The surface-to-surface average distance was 3.20 mm and 2.46 mm for 3 and 16 landmarks, respectively. This suggested that increasing the number of corresponding landmarks leads to more accurate patient-specific reconstructions.

A limitation of this study was the use of a dry bone scapula rather than a scapula in the presence of other body structures and tissues. The research study suggested that the use of a contour to aid the

location of more corresponding and matching features would produce a better prediction of the 3D patient-specific model.

2.1.2 Edge and contour detection in images

To improve the outcome of X-ray image reconstruction without increasing the number of X-ray images, researchers have reported on the use of different contour detection techniques; implemented manually or automatically. Contour detection techniques have shown great success in improving the 3D reconstruction of X-rays images by providing prior information for landmark selection (Diop & Burdin, 2013; Kaur & Singh, 2016; Laporte *et al.*, 2003; Zhang *et al.*, 2010; Zhang *et al.*, 2011; Zheng *et al.*, 2007). This information reduces the search space for corresponding landmarks.

Kaur and Singh (2016) examined different edge detection techniques available using transformation and filtration methods. The edge detection techniques that were examined are Sobel, Prewitt, Roberts, Canny, and Laplacian Gaussian techniques (Acharjya *et al.*, 2012; Kaur & Singh, 2016; Zhang *et al.*, 2010). Based on the performance of the edge detection techniques using synthetic X-ray images, the authors concluded that the Canny edge detection operator yields the best results for edge detection in images. However, there was a need to perform a benchmarking test using real medical image data and find a unified noise elimination method that does not lead to loss of image details. Acharjya *et al.* (2012) also carried out research on edge detectors and the results of this work ranked the Canny edge detector as a robust operator compared to the rest, although performance maybe vulnerable to noise in the images.

Zhang *et al.* (2010) reconstructed a femur from bi-planar X-ray images using the direct linear transformation (DLT) (Abdel-Aziz *et al.*, 2015). Their approach required extracting the contour of the femur and using the contour to locate corresponding landmarks. The images were first processed to eliminate noise using a median filter. A Canny edge detector was used to define the contour as it was considered the best edge detector of the time. Although reconstruction of the femur was feasible for the shaft, they faced challenges on the irregular parts of the femur such as the condyles, due to use of a regular geometric shape-based reconstruction model. Qualitative assessments of the 3D reconstruction showed that this method would present a cost-effective imaging method for resource-limited settings. However, further quantitative validation and improvement of this method was required to cater for irregular bone shapes.

Zhang *et al.* (2011) proposed an automated contour detection and extraction method for medical CT images of the knee joint. Their approach used different edge detection algorithms including Canny, Laplacian, Sobel, Roberts, and a chain code method for contour extraction. The approach worked well for building 3D geometrical models from 3D medical images, and it was suggested that the method be

tested on other medical images (Zhang *et al.*, 2011). In contrast to Kaur and Singh (2016), and Acharjya *et al.* (2012), Zhang *et al.* (2011) chose the Roberts edge detection over the Canny edge detection. The Laplacian operator was considered to be too general and could only detect obvious contours (Kaur & Singh, 2016; Zhang *et al.*, 2011). Thus, different contour detection algorithms cannot be generalised to work the same way for all images especially when applied to real medical images.

Other researchers have emphasised the use of active contours with shape priors to improve the accuracy of segmentation and registration algorithms. The active contour acts as prior knowledge leading to improved accuracy of the reconstructed patient-specific 3D models (Diop & Burdin, 2013; Kass *et al.*, 1988; Middleton & Damper, 2004; Pereira *et al.*, 2016). Image segmentation methods that use geometric active contours have been proposed to overcome poor contrast between anatomical structures (Auroux *et al.*, 2011; Chabrier *et al.*, 2008; Diop & Burdin, 2013). Diop and Burdin (2013) proposed a method based on the use of active contours with an embedded shape prior to segment a femur from bi-planar images. The researchers incorporated prior information about the form of the target object to improve the segmentation of X-ray images that suffer from poor contrast, and sometimes missing parts, due to the super-imposition of structures. Both quantitative and qualitative results were reported for synthetic and real data, as there is no standard way of measuring the segmentation error (Chabrier *et al.*, 2008). The reconstruction gave a minimum squared error (MSE) of 0.2 mm compared to 2.02 mm using a classical active contour method (Li *et al.*, 2005). Thus, the proposed geometric active contour method performed better than classical methods which do not use shape priors. However, the output was still different from the ground-truth. The authors proposed further testing of the method on other real datasets, in order to improve it for clinical use (Diop & Burdin, 2013). This research showed that adding a shape prior to the segmentation task improves the result because of the reduced search space when embedded prior information is added.

2.1.3 Deep learning in medical imaging

Deep learning is a class of machine learning that deals with training a model to learn data representations using multiple layers of abstraction (LeCun *et al.*, 2015). Deep learning techniques have recently been introduced in medical image processing and analysis and have shown promising results in various applications such as segmentation, registration and image reconstruction (Litjens *et al.*, 2017; Maier *et al.*, 2019; Ronneberger *et al.*, 2015). Deep learning methods are different from optimisation-based techniques that iteratively determine the transformation parameters; creating a need for expert intervention to reduce computation bottlenecks. Once a deep learning model is trained to generalise there is no need for optimisation, thus saving time, computation power and expert resources (Long *et al.*, 2015; Miao *et al.*, 2016; Middleton & Damper, 2004; Wernick *et al.*, 2010).

Miao *et al.* (2016) reported on a method for real-time 3D registration from 2D synthetic X-ray images using convolutional neural network (CNN) regression. The study employed a hierarchical regression strategy to detect contours in the X-ray image. The proposed solution using CNN gave a registration success rate of 92.3% with a mean target registration error of 0.282mm, and a speed of 0.08s. The result of registration using an optimisation-based method (mutual information and gradient correlation) gave a success rate of 92.7% with a mean target registration error of 0.260mm, and a registration speed of 4.71s. These results indicated the potential of using a deep learning approach to 3D image reconstruction from 2D X-ray images.

Other deep learning methods have also been used in medical imaging research studies especially in segmentation of microscopy slide samples, X-ray images and 3D CT scans of the brain, knee cartilage, ribs, liver and prostate (Cernazanu-Glavan & Stefan, 2013; Isin *et al.*, 2016; Prasoon *et al.*, 2013; Ronneberger *et al.*, 2015). These research studies have shown the potential of deep-learning algorithms for medical image processing. However, most deep learning medical image analysis research has been limited by the need for large training datasets required to develop and train deep learning models.

2.1.4 Summary of review

Research has been done on 3D image reconstruction from 2D bi-planar X-ray images of the scapula using landmark-constrained model fitting. However, Mutsvangwa *et al.* (2017) found only 3 corresponding reproducible landmarks in the bi-planar images resulting in unacceptable reconstruction errors. The main reason for the limited number of corresponding landmarks was identified as the presence of a uniform grayscale of intensities in X-ray images with inconspicuous features due to super-imposed structures. Some researchers have used shape priors to improve feature detection in bi-planar images (Diop & Burdin, 2013; Li *et al.*, 2005; Middleton & Damper, 2004). However, reconstruction errors still exist, and the procedures are subjective and not easy to generalise to all bone shapes, due to the presence of noise and super-positioned structures onto the image plane. Deep learning approaches have proved to be robust in feature detection especially for computer vision, but the application of these approaches is still limited in the field of medical imaging due to the large datasets required.

3 Review of theoretical concepts

This chapter presents a theoretical background on the technical concepts which were introduced in chapters 1 and 2 and on the methods described in chapter 4.

3.1 Image reconstruction

Three-dimensional image reconstruction from 2D X-ray images is the process of extracting shape or intensity values from a 2D image and transforming the values into a 3D space to generate a 3D image or geometry. The 2D images are acquired from 2D modalities like fluoroscopy, ultrasound and X-ray imaging systems. There are two major image processing sub-tasks used in a 3D reconstruction pipeline; namely, image registration and segmentation. A description of these processes is given below.

3.1.1 Registration

Image registration is a fundamental task in image processing, and it involves aligning two or more images into a common coordinate system. Usually, the datasets are not aligned because of differences in their acquisition, for example use of different sensors, different viewpoints or different time points. Through the alignment process, all the datasets are transformed into the same coordinate system for easy comparison and integration. Registration can either be rigid or non-rigid and may be performed based on image features or image intensity values (Besl & McKay, 1992; Liao *et al.*, 2016).

Rigid registration is the process of moving data into the same coordinate system by the elimination of translation and rotation. It can be based on corresponding landmarks between the datasets to transform and align the datasets into the same coordinate system (Besl & McKay, 1992). Landmarks are meaningful points describing similar features across a dataset of a given population. They are used to find and maintain correspondences within the dataset. However, when applying rigid registration to a dataset of the same population which exhibit variations in shape and size like a medical image dataset of the same structure but different people, it is difficult to cater for the variations in shape thus the use of non-rigid registration.

Non-rigid registration aligns datasets through localised deformations to attain correspondence across image regions (Crum *et al.*, 2004; Gerig *et al.*, 2014; Lüthi *et al.*, 2013; Mayya *et al.*, 2013). This is especially common when aligning datasets with variable shape and size, for example, medical image data. Non-rigid registration defines deformation fields that map a defined reference object to the target object (Crum *et al.*, 2004).

Image matching for registration may use image features or image intensity to establish correspondence between images. The main drawback of feature-based registration algorithms is that the accuracy of the

methods relies on the accurate detection of the features. However, feature detection is a challenging task. The features used describe the shape of the structure in question, but when the landmarks are spread out or misaligned the shape is easily lost (Miao *et al.*, 2016). This makes feature-based methods less accurate compared to intensity-based methods.

Intensity-based registration algorithms use pixel or voxel intensities for correspondence without the use of feature landmarks. The target intensity is interpolated towards the reference to maximise the similarity measure. The similarity measure determines the proximity of the statistical distribution of intensities in the images. Intensity-based registration gives accurate and detailed properties of an image. Also, the delineation of feature landmarks is not required. However, intensity-based methods are not able to cope with large geometric deformations and they often require many similarity measure evaluations. This makes intensity-based registration computationally expensive, time-consuming and results in computational bottlenecks because of the iterative nature of the similarity evaluation (Lam & Lui, 2014; Miao *et al.*, 2016).

3.2 Segmentation

Image segmentation refers to the partitioning of an image into parts with similar attributes/features and properties. The main aim of partitioning an image is to represent the image in a form that is more meaningful and gives a better understanding of the characteristics of the image. Through segmentation, features in an image are located by assigning labels to pixels in the region of interest. The result is pixels with the same label sharing certain characteristics like colour, intensity, or texture, making the image easier to understand. During segmentation, the region of interest is usually mapped out by its contour. A contour is a closed curve joining all the continuous points along a boundary with the same intensity; thus a contour is made up of edges (Kaur & Singh, 2016). In 3D reconstruction from 2D bi-planar X-ray images, it is important to accurately locate corresponding features in the images during segmentation. These corresponding features are used in registration and subsequently reconstruction, therefore the result of any reconstruction depends on the quality of segmentation. Contour detection is one way to approach a segmentation task. There are several methods for contour detection, which are grouped into two categories: search-based and zero-crossing based.

Search-based contour detection methods are discrete differential operators and detect edges by first computing a measure of edge strength. These include Roberts and Sobel operators. The Sobel operator computes an approximation of the gradient of the image intensity function using discrete differences between rows and columns whereas the Roberts cross operator is used to approximate the gradient of an image by computing the sum of the squares of the discrete differences between diagonally adjacent pixels. These operators are both relatively inexpensive in terms of computations (Acharjya *et al.*, 2012; Zhang *et al.*, 2011).

Zero-crossing based edge detection methods search for zero crossings in a second-order derivative expression computed from the image to find the edge. These include the Canny and the Laplacian edge detectors. The Canny edge detector is one of the most popular algorithms for edge detection because of its robustness and accuracy where there is a high entropy (Acharjya *et al.*, 2012; Chabrier *et al.*, 2008; Kaur & Singh, 2016; Zhang *et al.*, 2011). The Canny edge detector is a multi-stage algorithm and detects a wide range of edges in images by looking for the local maxima of the gradient. The Laplacian operator is isotropic and is more appropriate in situations where the edge position is prioritised over its surrounding pixel difference. This operator is usually applied to denoised images since it responds to isolated pixels more than to the edge or line. Although the Canny edge detector has proved to be robust, it fails in some images. Thus, there is no reliable unified way to detect edges in an image.

3.3 Shape and intensity modelling

Statistical shape models (SSM) and statistical shape and intensity models (SSIM) are deformable models learned from a set of labelled examples of a given statistical population (Cootes *et al.*, 1992; Cootes *et al.*, 1995). These models describe the average shape and or intensity distribution within that population. Statistical models are used in 3D reconstruction from 2D images by constraining the model parameters of a shape of interest; for example, landmarks or intensities. These models can be used as shape priors in a segmentation process.

The SSM model describes only the shape and is inherently a set of points describing a surface. Shape can be defined as the geometrical information that remains after all transformation effects (translation, rotation, and scaling) have been eliminated from an object (Dryden & Mardia, 1998; Stegmann & Gomez, 2002). The SSM model is described by the covariance matrix of the training dataset. The mean shape is given by $\bar{x}$ in equation (3.1):

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i \tag{3.1}$$

where N is the number of training datasets and x_i are the coordinates of the number of landmarks describing the shape.

The covariance matrix, S is calculated using equation (3.2):

$$S = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})(x_i - \bar{x})^T \tag{3.2}$$

where T is the transpose of the matrix. The model can be built from either 2D or 3D training datasets. To build the model, a dataset with variations that describe the population of interest is required. There

are two major steps in building the model given the required dataset. The first is to determine the best technique to represent the given dataset. The most common is the use of landmarks. The coordinates of each landmark are used to describe the shape as indicated in equations (3.1) and (3.2) by x_i. The second step is to define landmarks on each object in the training dataset; this can be done manually or automatically by selecting the same anatomical location on the objects consistently throughout the dataset so that a dataset in correspondence is generated. Once correspondence has been established through-out the dataset, the objects are aligned to eliminate all the variations in the dataset that are not due to shape. Alignment can be performed automatically using general Procrustes analysis (GPA)(Dryden & Mardia, 1998). The next step after alignment of the dataset is calculation of the mean, $\bar{x}$ and covariance, S. This is followed by applying principle component analysis (PCA) (Jolliffe, 1993; Stegmann & Gomez, 2002) on the covariance matrix to find the most important modes of variation in the training dataset from the mean shape, $\bar{x}$. A valid new shape, x can be generated from the resultant SSM using the equation (3.3):

$$x = \bar{x} + \sum_{m=1}^{M} b_m \varphi_m \tag{3.3}$$

where b_m is the value that describes the contribution of the first m modes of the shape variation to the shape of the object, φ_m represents the m^{th} eigenvalues and eigenvectors and the sum of b_m and φ_m is the covariance matrix.

An SSIM can be generated using the same steps as the SSM but replacing the shape as the feature being described with both shape and intensity (Cootes & Taylor, 2001). The SSIM gives detailed information about the model intensity. However, it requires high computation power leaving most researchers with no option but to use the SSM or a binarized SSIM. Binarising the SSIM, entails setting a threshold intensity value and replacing the values below it with zeroes and the rest with ones. The main aim of binarising the model is to reduce computational requirements. Another advantage of an SSIM is that it can be used to generate infinite volumetric instances that can be projected into 2D space, generating synthetic X-ray images for each instance while the SSM would generate mesh instance that have to be converted into volumes before rendering them into synthetic X-ray images.

3.4 Model fitting reconstruction

Three-dimensional landmarks from the region of interest are required for 3D reconstruction using landmark-constrained model fitting. The 3D landmarks are often obtained through 3D point localisation methods. One approach for 3D point localization used direct linear transformation (DLT). However, the accuracy of DLT depends on the accuracy of the matching points selected from two different views of the same object. Epipolar geometry is often used to reduce the search space of a corresponding point

and therefore the selection accuracy of matching points. The landmark-constrained method of reconstruction aims to find a likelihood function that maps the model reference landmarks to the reconstructed 3D points (target points).

3.4.1 3D point localisation

The DLT is a frequently used camera calibration method that allows mapping between 3D space coordinates and image coordinates (Abdel-Aziz *et al.*, 2015). The main aim of the DLT algorithm is to calculate the calibration parameters also known as the projective transformation parameters between two coordinate systems (Adams, 1981; Douglas *et al.*, 2004). The calculated transformation parameters were used to calculate the 3D object space coordinates for the given set of 2D image space coordinates. In X-ray stereophotogrammetry, where multiple radiographic images are taken, the DLT enables the determination of the 3D coordinates of an object given the selected 2D coordinates are visible in more than one camera perspective (view) (Adams, 1981; Chimhundu *et al.*, 2014; Douglas *et al.*, 2004; Zhang *et al.*, 2010). The main advantage of the DLT is that one does not require knowledge about the imaging parameters to project the 3D points to 2D or vice versa. However, the DLT is applicable when known 3D object points, their image projections, and the viewing source are collinear. The known object space coordinates are used as the control points which are usually fixed to a calibration frame. Thus, given the control points and image points, the transformation parameters between the object and image space can be calculated. The transformation yields a 3 x 4 calibration matrix made up of 11 parameters describing the relationship between the image and object. However, when using the DLT to determine the transformation parameters from 2D to 3D, a minimum of 6 control points that are not co-planar are required.

Given a pair of 2D bi-planar images as shown Figure 3.1, the first view with points (u, v) and the corresponding second view with points (u^i, v^i) are mapped by unknown transformation parameters L_{ij} and L_{ij}^i. to the 3D object coordinates (X, Y, Z).

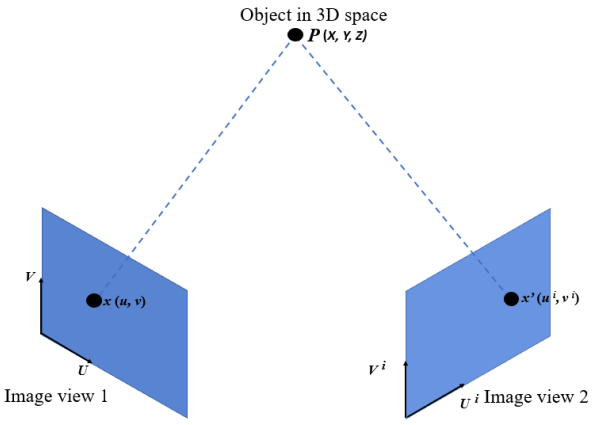

Figure 3.1: 3D localization of a point *P* using 2D projections *x* and *x'* in a two-view system.

Mathematically, Adams (1981) defined the constraints mapping the points, (u, v) in image view 1 and points, (u^i, v^i) in image view 2 to their 3D coordinates using equations (3.4). Points (u, v) and (u^i, v^i) in image view 1 and 2, respectively are given by;

$$u = \frac{(L_{11}X + L_{12}Y + L_{13}Z + L_{14})}{(L_{31}X + L_{32}Y + L_{33}Z + 1)} \qquad v = \frac{(L_{21}X + L_{22}Y + L_{23}Z + L_{24})}{(L_{31}X + L_{32}Y + L_{33}Z + 1)}$$
$$u^i = \frac{(L^i_{11}X + L^i_{12}Y + L^i_{13}Z + L^i_{14})}{(L^i_{31}X - L^i_{32}Y - L^i_{33}Z + 1)} \qquad v^i = \frac{(L^i_{21}X + L^i_{22}Y + L^i_{23}Z + L^i_{24})}{(L^i_{31}X - L^i_{32}Y - L^i_{33}Z + 1)} \tag{3.4}$$

where *X*, *Y*, and *Z* are known 3D reference points, L_{ij} and L^i_{ij}. are the transformation parameters.

For transformation of the image points in the bi-planar images equation (3.4) can be re-written as equations (3.5) - (3.6) and (3.7) - (3.8) for points (u, v) and (u^i, v^i) in the first view and second view, respectively; using cross-multiplication.

$$u = L_{11}X + L_{12}Y + L_{13}Z + L_{14} - L_{31}uX - L_{32}uY - L_{33}uZ \tag{3.5}$$

$$v = L_{21}X + L_{22}Y + L_{23}Z + L_{24} - L_{31}vX - L_{32}vY - L_{33}vZ \tag{3.6}$$

$$u^i = L^i_{11}X + L^i_{12}Y + L^i_{13}Z + L^i_{14} - L^i_{31}u^iX - L^i_{32}u^iY - L^i_{33}u^iZ \tag{3.7}$$

$$v^i = L^i_{21}X + L^i_{22}Y + L^i_{23}Z + L^i_{24} - L^i_{31}v^iX - L^i_{32}v^iY - L^i_{33}v^iZ \tag{3.8}$$

The equations (3.5) and (3.6) can be expressed in the form of a matrix in equation (3.9).

$$\begin{bmatrix} X_1 & Y_1 & Z_1 & 1 & 0 & 0 & 0 & 0 & -u_1X_1 & -u_1Y_1 & -u_1Z_1 \\ 0 & 0 & 0 & 0 & X_1 & Y_1 & Z_1 & 1 & -v_1X_1 & -v_1Y_1 & -v_1Z_1 \end{bmatrix} * \begin{bmatrix} L_{11} \\ L_{12} \\ L_{13} \\ L_{14} \\ L_{21} \\ L_{22} \\ L_{23} \\ L_{24} \\ L_{31} \\ L_{32} \\ L_{33} \end{bmatrix} = \begin{bmatrix} u_1 \\ v_1 \end{bmatrix} \tag{3.9}$$

Given a minimum of 6 control points to any number of points, n in the first image view, Equation (3.10) is obtained. The transformation parameters, L_{ij} for the first view would be calculated by constructing matrix A and B from the known $(X, Y, Z)_n$ and $(u, v)_n$ object and image control points, respectively.

$$\underbrace{\begin{bmatrix} X_1 & Y_1 & Z_1 & 1 & 0 & 0 & 0 & 0 & -u_1X_1 & -u_1Y_1 & -u_1Z_1 \\ X_2 & Y_2 & Z_2 & 1 & 0 & 0 & 0 & 0 & -u_2X_2 & -u_2Y_2 & -u_2Z_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ X_n & Y_n & Z_n & 1 & 0 & 0 & 0 & 0 & -u_nX_n & -u_nY_n & -u_nZ_n \\ 0 & 0 & 0 & 0 & X_1 & Y_1 & Z_1 & 1 & -V_1X_1 & -v_1Y_1 & -v_1Z_1 \\ 0 & 0 & 0 & 0 & X_1 & Y_1 & Z_1 & 1 & -V_2X_2 & -v_2Y_2 & -v_2Z_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & X_1 & Y_1 & Z_1 & 1 & -V_nX_n & -v_nY_n & -v_nZ_n \end{bmatrix}}_{\mathbf{A}} * \underbrace{\begin{bmatrix} L_{11} \\ L_{12} \\ L_{13} \\ L_{14} \\ L_{21} \\ L_{22} \\ L_{23} \\ L_{24} \\ L_{31} \\ L_{32} \\ L_{33} \end{bmatrix}}_{\mathbf{L}} = \underbrace{\begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \\ v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}}_{\mathbf{B}} \tag{3.10}$$

Equation (3.10) simplifies to equation (3.11) where $\boldsymbol{A}$ is the first matrix on the left, $\boldsymbol{L}$ is the second matrix on the left, and B is the matrix on the right.

$$\boldsymbol{A} * \boldsymbol{L} = \boldsymbol{B} \tag{3.11}$$

The mean values of the calibration parameters, $\boldsymbol{L}$ are obtained by the least square minimisation because matrix $\boldsymbol{A}$ is not square. The pseudo-inverse of matrix $\boldsymbol{A}$ in equation (3.10) can be solved using singular value decomposition (SVD) to obtain the calibration parameters of image view 1 given the 3D object and 2D image coordinates (Andrews & Patterson, 1976).

$$\boldsymbol{L} = \boldsymbol{pinv}\,(\boldsymbol{A}) * \boldsymbol{B} \tag{3.12}$$

The points in image view 2 can be expressed in the same ways as equation (3.9) and for n control points the equations (3.10) can be formulated to calculate the calibration parameters of image view 2. After obtaining the calibration parameters the 3D point coordinates of the point, P can be calculated given the 2D point u, v, u^i and v^i. Consider the linear equations (3.5) - (3.6) with points (u, v) and (u^i, v^i) in the image view 1 and 2, respectively. The mapping between the 3D object and 2D image coordinates is established through the matrix, $\boldsymbol{L}$ given in equation (3.13):

$$\begin{bmatrix} u - L_{14} \\ v - L_{24} \\ u^i - L^i_{14} \\ v^i - L^i_{24} \end{bmatrix} = \begin{bmatrix} L_{11} - L_{31}u & L_{12} - L_{32}u & L_{13} - L_{33}u \\ L_{21} - L_{31}v & L_{22} - L_{32}v & L_{23} - L_{33}v \\ L^i_{11} - L^i_{31}u & L^i_{12} - L^i_{32}u & L^i_{13} - L^i_{33}u \\ L^i_{21} - L^i_{31}v & L^i_{22} - L^i_{32}v & L^i_{23} - L^i_{33}v \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \tag{3.13}$$

If the matrix on the left is denoted by $\boldsymbol{UV}$ and the first matrix on the right as $\boldsymbol{C}$, equation (3.13) can be written as:

$$\boldsymbol{UV} = \boldsymbol{C} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \tag{3.14}$$

The points X, Y, Z can be found from equation (3.14) by getting the pseudo-inverse of $\boldsymbol{C}$:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = UV * pinv\,(C) \tag{3.15}$$

3.4.2 Epipolar geometry

Epipolar geometry is employed to avoid the extrapolation of image points outside the calibration volume since the accuracy of stereo-photogrammetry highly depends on selecting of matching corresponding points for the DLT. Epipolar geometry aims to map images in one view to the corresponding points in another view to reduce the search space for corresponding points.

Given an object in a 3D scene viewed from two camera perspectives, the relationship between the two image planes and the object can be established by matching image points in the views to the 3D source. However, to find the two corresponding points in the two perspectives, epipolar geometry is applied to the two planes (Zhang, 2000). Figure 3.2 describes the relationship between the 3D point P, its projection in two planes (x and x') and source of view (O and O').

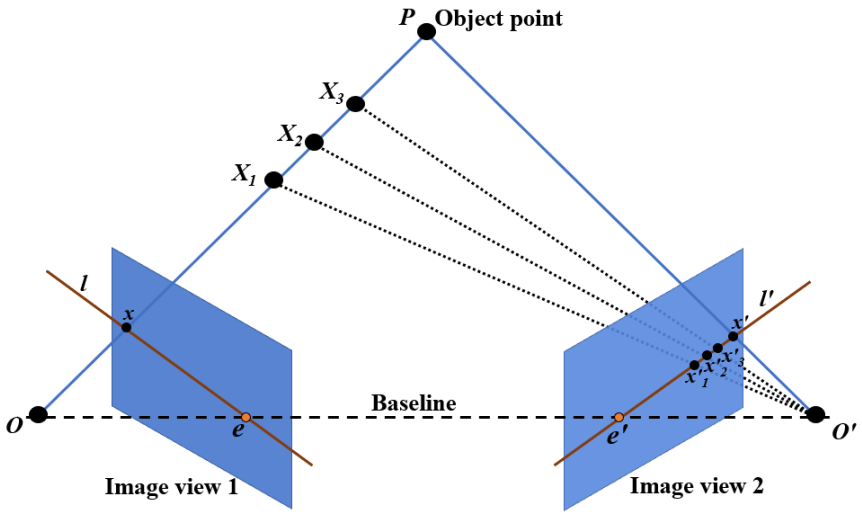

Figure 3.2: Epipolar geometry with two camera system used to locate a point *x* in the corresponding image view. Adapted from (Hartley et al., 2004).

In Figure 3.2 OP, $O'P$, and OO' form the epipolar plane, and:

OO' is the baseline

l and l' are the epipolar lines

e and e' are the epipoles

Given a point, P in 3D space, projected into two image planes viewed with two cameras O and O', the two cameras are joined by the baseline, which intersects the two image planes at points called epipoles (e and e'); the virtual projection centres of the cameras. The virtual point of camera O in the first view is e and that in the second view is e'. There two lines OP and $O'P$ joining the 3D point P and the camera sources, contain the projection of the point P into the image planes x and x' for the first and second view, respectively. The lines OP and $O'P$, joined by the baseline, form the epipolar plane. The epipolar plane intersects the image planes to form the epipolar lines l and l'. The projection of P in the first view, x, and the virtual projection of the camera source e, on to the first image plane, are joined by the epipolar line l. While the corresponding point x' in the second view and the virtual projection of the second camera source, e' are joined by the second epipolar line l'.

The relationship between the 2D projections x and x' and the 3D point P is determined by epipolar geometry. For a calibrated scene, an essential matrix (Helmke *et al.*, 2007) is used; while for an uncalibrated scene, a fundamental matrix (Hartley & Zisserman, 2004) is used to find the relationship between x, x', and P.

For an uncalibrated scene where the camera intrinsic and extrinsic parameters are unknown, the epipolar geometry is achieved using the fundamental matrix, F. The matrix, F is a homogeneous singular matrix of 3 x 3 dimensions. The matrix is used to map the 2D projection point, x in the first image plane view on the epipolar line l to the corresponding epipolar line l' in the second image plane view. Finding the epipolar line, l' helps to reduce the search space in locating the corresponding point in the second image plane view.

Any point, $x_i = (u_i, v_i, 1)$ in image view 1 is related to the corresponding point $x_i^{'} = (u_i^{'}, v_i^{'}, 1)$ in image view 2 using a fundamental matrix through equation (3.16):

$$xFx' = 0$$

$$(u_i, v_i, 1) \ * \ F \ * \begin{pmatrix} u_i' \\ v_i' \\ 1 \end{pmatrix} = 0 \tag{3.16}$$

where F is the 3 x 3 fundamental matrix.

$$F = \begin{bmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{bmatrix}$$

In order to calculate F using the eight-point algorithm (Hartley, 1997), eight or more matching points are required from the image view 1 and 2 to obtain equation (3.17):

$$\begin{bmatrix} u_1u_1' & v_1u_1' & u_1' & u_1v_1' & v_1v_1' & v_1' & u_1 & v_1 & 1 \\ u_2u_2' & v_2u_2' & u_2' & u_2v_2' & v_2v_2' & v_2' & u_2 & v_2 & 1 \\ u_3u_3' & v_3u_3' & u_3' & u_3v_3' & v_3v_3' & v_3' & u_3 & v_3 & 1 \\ u_4u_4' & v_4u_4' & u_4' & u_4v_4' & v_4v_4' & v_4' & u_4 & v_4 & 1 \\ u_5u_5' & v_5u_5' & u_5' & u_5v_5' & v_5v_5' & v_5' & u_5 & v_5 & 1 \\ u_6u_6' & v_6u_6' & u_6' & u_6v_6' & v_6v_6' & v_6' & u_6 & v_6 & 1 \\ u_7u_7' & v_7u_7' & u_7' & u_7v_7' & v_7v_7' & v_7' & u_7 & v_7 & 1 \\ u_8u_8' & v_8u_8' & u_8' & u_8v_8' & v_8v_8' & v_8' & u_8 & v_8 & 1 \end{bmatrix} * \begin{bmatrix} f_{11} \\ f_{12} \\ f_{13} \\ f_{21} \\ f_{22} \\ f_{23} \\ f_{31} \\ f_{32} \\ f_{33} \end{bmatrix} = 0 \quad (3.17)$$

This can be rewritten as equation (3.18):

$$WF = 0 \quad (3.18)$$

where W is an $N\,x\,9$ matrix derived from $N \geq 8$ matching points (u_i, v_i, u_i' and v_i') and F is the fundamental matrix. The matrix F can be computed as a least square solution by SVD, as matrix W is rank deficient. However, matrix F has a rank of 2 and the best approximation is given by adding a constraint, $det(F) = 0$.

3.4.3 Digitally reconstructed radiographs

Digitally reconstructed radiographs (DRRs) are synthetic X-ray images generated to simulate the effect of real X-rays. Synthetic X-ray images are formed when volumetric CT data is projected onto a plane. In a clinical setting, the DRR projection is important to avoid repetitive exposure of patients to ionising radiation thus enabling better surgical planning and the possibility of continuous post-operative diagnosis (Reyneke *et al.*, 2019; Sarkalkan *et al.*, 2014). Digitally reconstructed radiographic projection is commonly done using the ray-casting method (Mu, 2016), which is used for rendering in computer graphics and is based on the Beer-Lambert's law. This law describes the attenuation of X-rays travelling through space (Staub & Murphy, 2013). During ray-casting, the voxel values of the volumetric model encountered by a single ray as it travels between the centre of projection (COP), and the current pixel, are integrated to obtain the value of each pixel of the DRR image. Mathematically, DRR projection is represented by equation (3.19), after all the physical phenomena like scatter, veiling glare and beam hardening are eliminated. The source is modelled as mono-energetic (Staub & Murphy, 2013).

$$I_i = I_o exp\left(-\int_{x_o}^{x_i} c\mu(x)dx\right) \quad (3.19)$$

where x_o is the location of the source or centre of projection (COP) and x_i is the location of the pixel detector; I_o is the initial intensity of the photon beam, I_i is the resultant intensity of the photon beam after travelling through the volume from the COP, O to the pixel detector. The volumetric data linear attenuation contribution (LAC) distribution at point x, is $\mu(x)$. The LAC that corresponds to a given

voxel and ray is determined using the Hounsfield unit equation (3.20). The Hounsfield unit is also known as the CT number.

$$h_x = \frac{10^{-3}\,(\mu_x - \mu_w)}{\mu_w} \tag{3.20}$$

where h_x is a CT number, μ_w is the LAC of water at a specific CT energy, and μ_x is the LAC for the current voxel. There are different projection strategies used to select which ray paths to include in the projection. These include orthographic (parallel rays), perspective (rays emanating from a point), and fan-beam projection. Orthographic projection is the simplest case, and is defined using equation (3.21):

$$DRR_{(x,y)} = I_0 exp\left(-\sum_{z=1}^{Z} \mu_{x,y,z} I_z\right) \tag{3.21}$$

where $DRR_{(x,y)}$ is the function that gives the intensity value for a set of 2D coordinates, (x, y); I_z is the distance travelled by the ray through a single voxel; $\mu_{x,y,z}$ is the LAC for a voxel at coordinates (x, y, z) which is derived from its CT number.

3.5 Deep learning

Currently, there are several deep learning models that are being used, such as convolutional neural network (CNN), recurrent neural networks (RNN), multi-layer neural networks, unsupervised pre-trained networks like auto-encoders, generative adversarial networks (GAN) (Maier *et al.*, 2019).

3.5.1 Neural networks

Neural networks are the main form of deep learning approaches deployed in medical image processing and analysis. Neural networks are inspired by the nature of information processing and communication patterns as shown by neurons in the nervous system (He *et al.*, 2016; LeCun *et al.*, 2015). Similar to the nervous system, in neural networks, the information is transferred through interconnected layers with several branches. Each neuron consists of inputs and their corresponding weights, the sum of each input multiplied by its weights, the set threshold value, and the output which is usually the signal activated by the sum above the threshold. Figure 3.3 is an illustration of the single neural network architecture (perceptron) with several neurons with inputs x_1 up-to x_n and their corresponding weights, w_i. The neuron takes the sum Z of the inputs and their weights and outputs signal, O if the sum Z is higher than the set threshold, T, where Z is a function represented by equation (3.22).

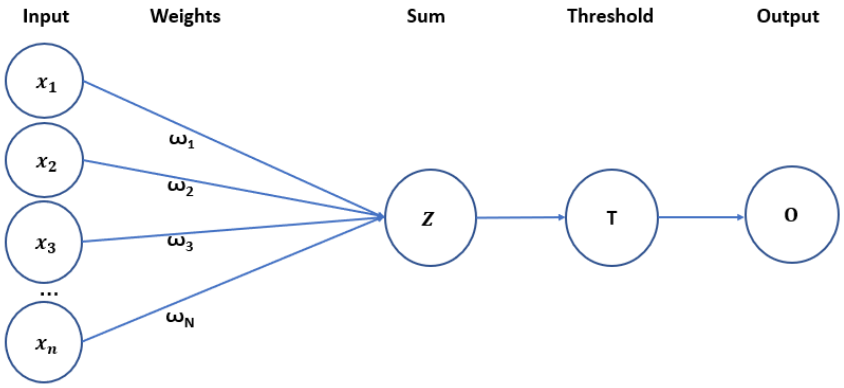

Figure 3.3: Representation of a neural network.

$$Z = f\left(\sum_{i=1}^{n} x_i x w_i\right) \tag{3.22}$$

During training of a neural network to generate a model, the weight of each input is usually initially randomly selected to create a starting point for the training. These parameters are updated until parameters that activate the network to produce the expected output are learned. Since neural networks are trained on datasets with their corresponding labels, the expected output, O is usually known. During the training, the predicted output signal is compared to the label signal and the difference between the signals is used to update weights on each neuron using back propagation to improve the model prediction. The neural network may comprise various layers to form a larger interconnection of layers referred to as a multi-layer perceptron (MLP). The MLP has been used to solve complex elementary logical functions. The MLP fails computationally when deployed on large functions (Hinton, 2007). This resulted in the use of CNNs which model features with translation and spatial invariance (non-linear functions) (Hinton, 2007; Zhang *et al.*, 1990).

3.5.2 Convolution Neural Networks

Convolutional neural networks are neural networks that use cross-correlation instead of general matrix multiplication in at least one of their layers. The CNN has a hierarchical architecture that has proved to be effective in the fields of computer vision and speech recognition because of its non-linear and sub-sampling nature (He et al., 2016). The CNN input goes through a series of processing steps and each step is referred to as a layer. Each layer gets its input from a previous layer creating a forward flow of information. However, a backward error propagation layer is created to get the difference between the input and label to update the weights of the kernels. A backward error propagation layer also known as the loss layer calculates the difference between the output and the label forming a loss function. Equation (3.23) is a simplest form of a loss function:

$$loss = ||label - output||^2 \tag{3.23}$$

Another commonly, applied loss function is the binary cross-entropy loss which is expressed as:

$$H_p(q) = -\frac{1}{N}\sum_{i=1}^{N} y_i \,.\, log(p(y_i)) + (1 - y_i) \,.\, log(1 - p(y_i)) \tag{3.24}$$

where y is the label, $p(y_i)$ is the probability prediction on y for all N points.

The training process runs in both forward and backward directions, for the model to learn the parameters in the input. The input goes through the forward process first then, the prediction is compared to the target using the loss layer to achieve the loss. The loss is used to guide the modification of the kernel weights, (w_i). The update and modification of the parameters can be done using stochastic gradient descent (SGD) as the simplest form of an optimizer with learning rate, η and number of epochs, t.

$$\left(w^i\right)^{(t+1)} = \left(w^i\right)^t - \eta \frac{\delta loss}{\delta (w^i)^t} \tag{3.25}$$

Other optimizers include RMSProp, Adagrad, Adadelta, Adams, Adamax, and Nadam which can easily be implemented in Keras (https://keras.io/optimizers/). The more convolutional layers used in any architecture, the more complex the features that the network can learn (He et al., 2016). This ability for the neural network to learn complex features has influenced many attempts to use the CNNs in medical image analysis applications; for example, 2D and 3D image segmentation, registration and 3D from 2D image reconstruction. In a 3D from 2D image reconstruction procedure, the CNN directly estimates feature transformation parameters from the 2D images which give competitive reconstruction results (Miao *et al.*, 2016). The CNNs consist of mainly 3 layers, namely the convolutional layer, the pooling layer, and the activation function layer.

The CNN can have several convolutional layers and the primary function of the convolutional layer is to extract features from the input image in a small area to create a feature map that is passed onto the pooling layer. Each convolutional layer consists of kernels or filters which are matrices that slide across the input image to identify the unique features in the image. The deeper the convolutional layers, the more abstract the features learnt; the first layers usually extract edges and contours in an image. The training dataset consists of inputs and their corresponding labels. The kernels are selected randomly at the start of the training process and the kernel values are adjusted during training until the kernel values describe the correlation between the input and its label. When the kernels can predict an output that describes the label, data independent of the training set called the test data is fed into the network to determine the reliability of the prediction. The trained model performance highly depends on the quality and nature of the training dataset used.

The pooling layer sub-samples from the convolution layer output by extracting an even smaller area of the image, which creates precision while decreasing computational time. In this layer, there is no parameter (feature) learning but image dimensional reduction into the next layer. The pooling layer accepts inputs of varying sizes and reduces them without losing any information about the image feature. This gives CNN the ability to model complex non-linear and large functions. There are two most commonly used types of pooling operators: the maximum pooling also known as the max pooling, which maps a sub-region of the input and extracts the maximum value. The second pooling operator, the average pooling, also maps the sub-region but extracts the average value.

The activation function layer captures the output of the CNN layers by mapping the resulting values between the set maximum and minimum value, usually 0 to 1 or - l to 1. This is done by performing a truncation on each element of the input without changing its size. There is also no parameter learning in this layer. There are two types of activation functions. The linear activation function is usually a straight line through the origin from - ∞ to + ∞. Represented by:

$$f(x) = x$$

The non-linear activation functions are the most commonly used type of activation functions in real-world applications, because of their ability to generalise with a variety of data. These functions include sigmoid or logistic, tanh or hyperbolic tangent, rectified linear unit (ReLU) activation.

The sigmoid function exists between 0 to +1 thus, used in cases where the model must predict the probability as any output between 0 and 1.

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

Tanh function is also sigmoidal in shape but ranges between - l and + 1 thus the negative inputs are mapped strongly negative and the zero inputs are mapped near zero in the tanh graph.

$$f(x) = tanh(x) = \frac{2}{1 + e^{-2x}} - 1$$

Rectified Linear Unit function also known as the ReLU function, ranges between 0 and + ∞. It is half rectified thus maps to 0 when the input is negative and maps to the input when the input is either above zero or equal to itself.

$$f(x) = \begin{cases} 0, & x < 0 \\ x, & x \geq 0 \end{cases}$$

Several CNN segmentation architectures exist in computer vision. An example of these architectures is the U-net (Ronneberger *et al.*, 2015), which gets its name from the nature of its shape as shown in Figure 3.4. It is a fully convolutional network (FCN) meaning it has no densely connected layers and this makes the network accept various image sizes. Another advantage of the U-net is its ability to augment the data to increase training data.

The U-net is separated into 3 main parts: the contracting/down-sampling path, bottleneck and the expanding/up-sampling path as shown in Figure 3.4.

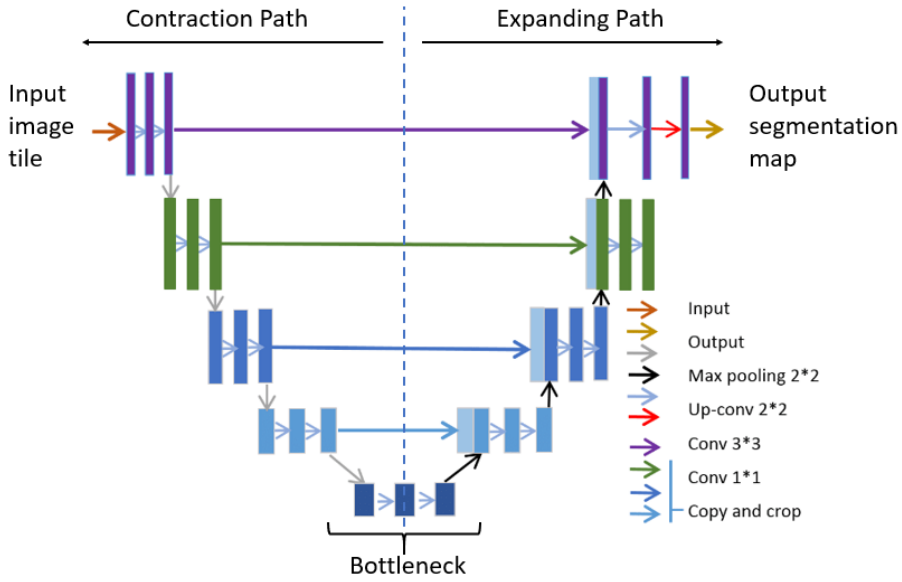

Figure 3.4: U-net architecture adapted from Ronneberger et al. (2015).

According to Ronneberger *et al.* (2015) the contracting path captures coarse contextual information of the input image to enable segmentation. The path consists of 4 blocks each composed of two 3x3 convolutional layers, their activation functions, and one 2x2 maximum (max) pooling layer. In the contracting path, the number of feature maps is doubled at each pooling. The learned course information is transferred to the expanding layer through skip connections at the end of each block. At the end of the whole contraction path, there is a bottleneck. The bottleneck consists of two convolutional layers and receives information from the contacting path and transfers it to the expanding path. Finally, the expanding path that enables precise localisation combined with the coarse information from the contracting path. This path comprises of 4 blocks each composed of a de-convolutional layer with a stride of 2, concatenation with the corresponding cropped feature maps from the contracting path and two 3x3 convolution layers with their activation functions.

The symmetrical nature of the U-net gives the network a larger number of feature maps in the up-sampling path which aids information transfer. In addition, the U-net has skip connections between the contracting and the expanding path that apply a concatenation operator rather than a summation

operator. The concatenation operator provides local information to the global information while up-sampling (Ronneberger *et al.*, 2015).

3.6 Evaluation metrics

This section reviews several 2D and 3D evaluation metrics which are important in image analysis and reconstruction assessment.

3.6.1 Two-dimensional evaluation metrics

The 2D evaluation metrics include the Dice coefficient and 2D landmark distance error.

The Dice Coefficient or the Dice similarity or overlap index measures the similarity between two sets. Its commonly used to evaluate image segmentation tasks by measuring the overlap between the predicted segmented image and its ground-truth (Dice, 1945; Zou *et al.*, 2004). The Dice coefficient value ranges from 0 to 1, where 0 indicates completely different elements present in both sets and 1 indicates perfect similarity in elements of both sets. Given two sets, the Dice coefficient can be calculated using equation (3.26):

$$Dice\ Coefficient = \frac{2TP}{2TP + FP + FN} \tag{3.26}$$

where *TP* stands for true positives, *FP* stands for false positives and *FN* stands for false negatives.

Landmark distance error: The point distance error is calculated by obtaining the difference between points in the ground-truth images and the trained model predicted images (Chimhundu *et al.*, 2014; Douglas *et al.*, 2004). Calculation of the distance error would inform the research on the trained model's ability to learn details in the image, especially embedded landmarks. When this method is applied to 2D points, absolute errors (e_x and e_y) in the x and y and the resultant error e_i between the 2D predicted and the reference 2D points are calculated. Given an image with initially known points (x_i, y_i) and a predicted image with predicted points (x_{pi}, y_{pi}) , absolute errors e_x and e_y can be calculated with equations (3.27) and (3.28).

$$e_x = \frac{1}{n}\sum_{i=1}^{1}|x_i - x_{pi}| \tag{3.27}$$

$$e_y = \frac{1}{n}\sum_{i=1}^{1}|y_i - y_{pi}| \tag{3.28}$$

where n is the number of points used, (x_i) and (y_i) are the known 2D coordinates and (x_{pi}) and (y_{pi}) the predicted 2D coordinates for the i^{th} point. The resultant error, e_i for the i^{th} point is given by:

$$e_i = \sqrt{(x_i - x_{pi})^2 + (y_i - y_{pi})^2} \tag{3.29}$$

For n predicted points, the average resultant error, e is calculated by applying the equation (3.29) above to each point then estimating the average using the equation (3.30) below.

$$e = \frac{1}{n}\sum_{i=1}^{1} |e_i| \tag{3.30}$$

3.6.2 Three-dimensional evaluation metrics

The 3D evaluation metrics applied in this research include the 3D point reconstruction error, the Hausdorff distance, and the average distance.

As with the 2D point distance error described in section 3.6.1, the 3D point reconstruction error can be calculated to evaluate the 3D point localisation process. Douglas *et al.* (2004) and Chimhundu *et al.* (2014) used the control and test point tests to validate the 3D localised points obtained using DLT for X-ray stereophotogrammetry.

Control points are used to ensure the mathematical correctness of the DLT equation. Control points are the points whose 2D and 3D coordinates are known and are used to calculate the transformation parameters. Control point reconstruction involves the use of the 2D points used to calculate the transformation parameters and the calculated transformation parameters to obtain the 3D localised points. The 3D localised points are evaluated by using the known 3D points that were used to obtain the transformation parameters. The absolute reconstruction errors E_x, E_y, and E_z and the resultant reconstruction error, E_i between the 3D localised points X_{ri}, Y_{ri}, and Z_{ri}, and the known 3D points X_i, Y_i, and Z_i are calculated. Given equations (3.31) - (3.35), the reconstruction errors E_x, E_y, E_z, and E_i are calculated by replacing the 2D points with 3D points.

$$E_x = \frac{1}{n}\sum_{i=1}^{1} |X_i - X_{ri}| \tag{3.31}$$

$$E_y = \frac{1}{n}\sum_{i=1}^{1} |Y_i - Y_{ri}| \tag{3.32}$$

$$E_z = \frac{1}{n}\sum_{i=1}^{1} |Z_i - Z_{ri}| \tag{3.33}$$

where X_i, Y_i, and Z_i are the known 3D coordinates and X_{ri}, Y_{ri}, and Z_{ri} are the 3D localised points of the i^{th}control point and n is the number of points used. The resultant reconstruction error, E_i for the i^{th} control point is determined using equation (3.34).

$$E_i = \sqrt{(X_i - X_{ri})^2 + (Y_i - Y_{ri})^2 + (Z_i - Z_{ri})^2} \tag{3.34}$$

For n 3D localised points, the average resultant reconstruction error, E is calculated by applying the equation (3.34) above to each point then estimating the average using equation (3.35).

$$E = \frac{1}{n}\sum_{i=1}^{1} |E_i| \tag{3.35}$$

Equations (3.31) - (3.35) are also applied to test points in order to determine the accuracy of the DLT calibration frame when used to predict unknown 3D points. Test points are the 2D landmarks points that were not used during the calculation of the transformation parameters.

The Hausdorff distance is a measure of similarity between two sets of points. This metric can be used to evaluate 3D surfaces especially in medical image and face reconstructions (Achermann & Bunke, 2000; Dubuisson & Jain, 1994; TakÁCs, 1998). Hausdorff distance finds the closest point between two mesh surfaces and returns the maximum distance between them. The smaller the distance the better the reconstruction. Given two meshes in correspondence, to evaluate their similarity two finite points sets are selected from the meshes say $X = (x_1, \ldots x_p)$ and Y = $(y_1, \ldots, y_p)$, the Hausdorff distance is given by equation (3.36):

$$H(X,Y) = max(\, h(X,Y), h(Y,X)\,) \tag{3.36}$$

where *h (X, Y)* is the directed Hausdorff distance given by equation (3.37):

$$h(X,Y) = max_{x \in X} min_{y \in Y} ||x - y|| \tag{3.37}$$

Equation (3.38) is the modified Hausdorff distance as reported by Takacs (1998).

$$H(X,Y) = \frac{1}{N_p}\sum_{x \in X} min_{y \in Y} ||x - y|| \tag{3.38}$$

where N_p is the total number of points in set X.

The average distance gives the mean of the shortest distance between two mesh surface points. It is computed by getting the shortest distance between each point from one mesh surface to the other and finding the average, X over all the points as shown in equation (3.39). The smaller the average value, the more similar the mesh surfaces are to each other.

$$X \coloneqq E\left[\left||x-y|\right|\right] = \frac{1}{\lambda(X)^2}\int_x\int_x \left||x-y|\right| d\,\lambda(x) d\,\lambda(y) \tag{3.39}$$

where $\mathrm{E}\|x-y\|]$ is the Euclidean distance (Burgstaller & Pillichshammer, 2009) between the corresponding points on the meshes and λ is a metric measure assigned to subsets of n-dimensional Euclidean space (n = 1, 2 or 3) known as the Lebesgue measure (Góra & Boyarsky, 1988).

4 Methods, tools and data

This research project aimed to address the absence of a unified automated methodology for detecting the contour of the structure of interest (scapula) in synthetic bi-planar X-ray images. This was to ease the identification of landmarks required for accurate three-dimensional (3D) reconstruction. This chapter gives an overview of the implemented methodology, describes the hardware and software tools and, explains the generation process of the data used in the project.

4.1 Methodology overview

Training data were generated as a prerequisite to the implementation of the proposed research methodology. Data generation entailed dataset acquisition and processing from a scapula statistical shape model (SSM). This was followed by the implementation of objective 1, namely scapula contour mapping in synthetic bi-planar X-ray images using a convolutional neural network (CNN) deep learning algorithm. Objective 1 was implemented in three phases namely: 1) preparation of training data; 2) training a selected CNN model architecture to predict the scapula contour in the lateral (LAT) view given the anterior-posterior (AP) view and vice versa; and 3) evaluation of the trained models using bi-planar X-ray images of a cadaveric upper-torso. Objective 2 encompassed the reconstruction of 3D mesh models using landmark-constrained model fitting. This objective was divided into two phases, namely 3D point localisation of the landmarks embedded in the predicted binary images and model fitting using the transformed landmarks. The final objective was the extraction of matching points from the contour of bi-planar images to perform landmark-constrained model fitting. Figure 4.1 shows an overview of the methods adopted to achieve the objectives.

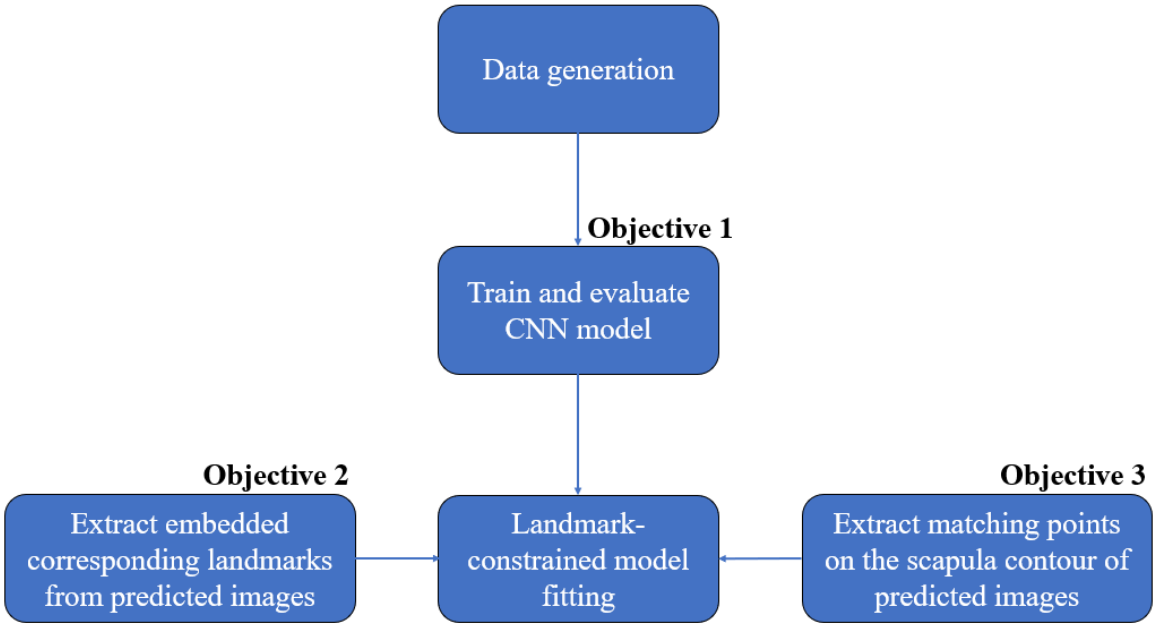

Figure 4.1: Overview of the research methods.

4.2 Hardware and software tools

The project was implemented on a 64-bit Proline workstation with an Intel(R) Core (TM) i7-4790 CPU @ 3.60 GHz, with 32.0GB of RAM and an NVIDIA GeForce GTX 960 graphics card.

The software packages included: Amira (www.fei.com/software/amira-avizo/), Scalismo (www.github.com/unibas-gravis/scalismo), Anaconda (https://www.anaconda.com) and MATLAB.

Amira is used for 3D and 4D visualization and was developed by Thermo Fisher Scientific. Amira version 5.4.3 was used to generate, visualize and inspect volumetric objects and meshes.

Scalismo is an open-source software for modelling statistical shapes and was developed by the University of Basel, Switzerland. Scalismo uses the Scala programming language and is hosted in IntelliJ IDEA Community Edition (https://www.jetbrains.com/idea/). The Scalismo software was used for generation and visualization of landmarks, meshes, volumetric objects, and digitally reconstructed radiographs (DRRs). Scalismo was also used in the validation of the results from the trained CNN model. The Scala version used in this project was 2.11.7. The DRR renderer obtained from Reyneke (2019) required additional drivers from OpenCL v1.2 and the Lightweight Java Gaming Library (LWJGL) v2.9.0.

Anaconda is an open-source software that uses Python and R programming languages for scientific computing, machine learning, data analysis and visualisation with simplified package installation and use. Packages included: Scikit-learn (https://scikit-learn.org/stable), TensorFlow (https://www.tensorflow.org), NumPy (http://www.numpy.org), Pandas (https://pandas.pydata.org), Matplotlib (https://matplotlib.org), OpenCV (https://opencv.org/) and Keras (https://keras.io/). All the experiments in Python were launched in Jupyter which is a package in the Anaconda distribution.

MATLAB (R2014a) is commercial software developed by MathWorks for multi-paradigm numerical computing. The MATLAB computer vision toolbox was used for the location of matching points using epipolar geometry.

4.3 Data generation

Datasets generated in this research project were obtained from an SSM of the scapula. The generated data included the model reference, mesh samples, and their annotated landmarks. The dataset also included a bounding box to aid the calibration of the projection space for the volumetric objects and in the calculation of the transformation parameters for 3D point localisation. Other dataset included volumetric objects and synthetic bi-planar X-ray images of the scapula. Figure 4.2 shows the systematic flow of datasets acquisition.

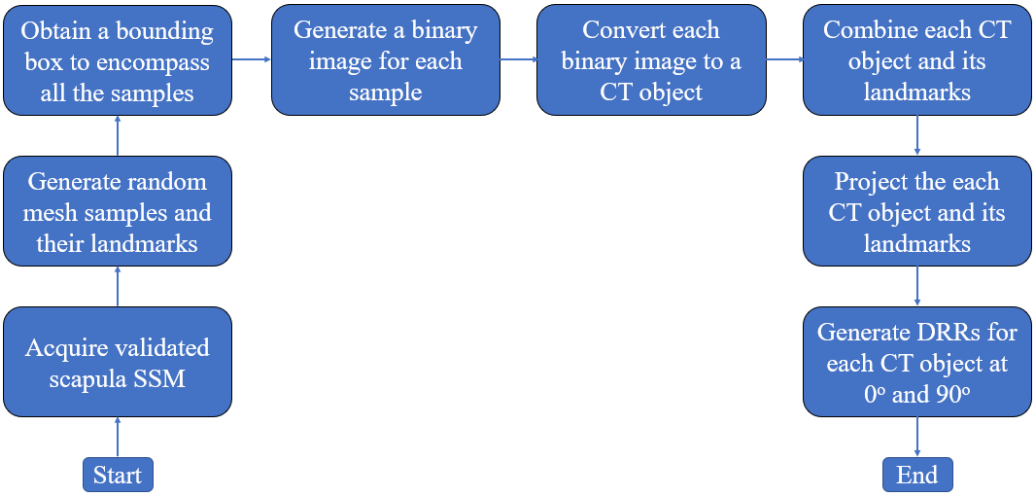

Figure 4.2: Steps taken to generate the datasets required to implement the methods.

The elements of dataset generation are described below.

4.3.1 Scapula statistical shape model

A scapula SSM was acquired from the Division of Biomedical Engineering at the University of Cape Town. The model had been built and validated from a dataset of 76 computed tomography (CT) scans. The model was obtained as a Hierarchical Data Format (.h5 file) and was initially used to generate random scapula mesh samples for the deep learning model training and subsequently used to reconstruct and validate the reconstructed 3D meshes. The existing SSM was used because of its ability to generate an infinite amount of "legal" but different mesh instances in correspondence. These characteristics of the statistical model made it an appropriate source of synthetic data.

4.3.2 Scapula mesh generation and landmark reliability

The mesh samples from the SSM were initially processed into volumetric objects that were used to generate synthetic bi-planar X-ray images using the digitally reconstructed radiograph (DRR) renderer. A subset of the mesh samples would later be used as the ground-truth meshes to evaluate the reconstructed meshes.

To obtain the mesh samples and their landmarks, the mean (reference) mesh sample was obtained using the *getreference* function in Scalismo. The reference mesh was displayed in the Scalismo user interface and landmarked with a subset of 6 out of the 16 reproducible scapula landmarks. The 16 landmarks are the scapula bone reproducible landmarks selected according to the results of studies done by Borotikar *et al.* (2015) and Ohl *et al.* (2010). The subset 6, of the 16 landmarks used in this study were located on the most varying parts of the scapula bone as indicated in Fouefack (2018). These 6 selected landmarks also included the 3 landmarks (inferior angle, coracoid and acromion) obtained by Mutsvangwa *et al.* (2017) that were the only identifiable landmarks in both the AP and LAT X-ray image perspectives

required for stereo-photogrammetry. The selected 6 landmarks included the inferior angle (A), infra glenoid rim (B), coracoid (C), acromion (D), superior angle (E) and base of the scapula (F) as shown in Figure 4.3. This subset was used as fiducial markers in the mesh instances which would later be used as prior information on the landmark locations for 3D approximation of the scapula using landmark-constrained SSM fitting.

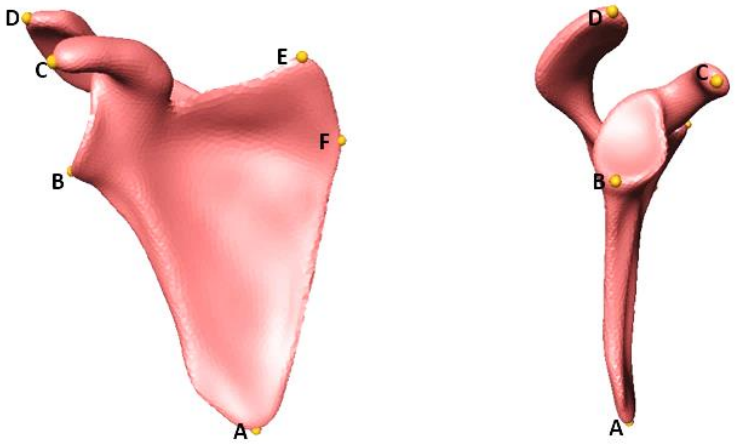

Figure 4.3: Selected reproducible landmarks on the scapula reference mesh inferior angle (A), infra glenoid rim (B), coracoid (C), acromion (D), superior angle (E) and base of the scapula (F) used in this research project (Borotikar et al., 2015; Ohl et al., 2010).

Reliability of landmark selection

The scapula landmarking guide developed by Borotikar *et al.* (2015) was used to obtain the 6 reference mesh landmarks. The landmarking process was repeated three times by three observers with at least 24 hours between each selection by an observer. The landmarks were selected in the same order each time. The selected sets of landmarks were tested for intra- and inter-observer selection precision and inter-observer reliability. These tests were to investigate consistency and reproducibility of landmark points among different observers during the landmark selection process (Borotikar *et al.*, 2015; Ohl *et al.*, 2010).

In order to quantify each set of landmark measurement errors, the intra and inter-observer precision were obtained. The intra and inter-landmark precision were given as the distance between the mean position and the observed position of the landmark (Pérez-Pérez *et al.*, 1990; Victor *et al.*, 2009). Intra-observer precision per observer was obtained by getting the mean position of the three attempts by an observer and then calculating distance, d_i of the observed position to the mean position. Given 3 landmark points Q_1, Q_2, and Q_3 each with coordinates in the *X*, *Y*, and *Z* direction. The mean landmark position, $\bar{Q}$ is given by equation (4.1) and equation (4.2). After obtaining each distance d_i from the mean position, the mean value of d_i and standard deviation (SD) were calculated for each landmark as the measure of overall intra-observer precision per landmark.

$$\bar{Q} = \frac{(Q_1 + Q_2 + Q_3)}{3} \rightarrow where \begin{cases} \bar{X} = \dfrac{X_1 + X_2 + X_3}{3} \\ \bar{Y} = \dfrac{Y_1 + Y_2 + Y_3}{3} \\ \bar{Z} = \dfrac{Z_1 + Z_2 + Z_3}{3} \end{cases} \tag{4.1}$$

$$d_i = \|\bar{Q} - Q_i\| = \sqrt{(\bar{X} - X_i)^2 + (\bar{Y} - Y_i)^2 + (\bar{Z} - Z_i)^2} \tag{4.2}$$

The inter-landmark precision was obtained by calculating the average of the mean landmark positions of the three observers (global mean), $\bar{\bar{Q}}$ and then the distance, d_i of the observed position from each observer mean landmarks, $\bar{Q}$. The inter-landmark precision was calculated using equation (4.3) and equation (4.4).

$$\bar{\bar{Q}} = \frac{(\bar{Q}_1 + \bar{Q}_2 + \bar{Q}_3)}{3} \rightarrow where \begin{cases} \bar{\bar{X}} = \dfrac{\bar{X}_1 + \bar{X}_2 + \bar{X}_3}{3} \\ \bar{\bar{Y}} = \dfrac{\bar{Y}_1 + \bar{Y}_2 + \bar{Y}_3}{3} \\ \bar{\bar{Z}} = \dfrac{\bar{Z}_1 + \bar{Z}_2 + \bar{Z}_3}{3} \end{cases} \tag{4.3}$$

$$d_i = \|\bar{\bar{Q}} - \bar{Q}_i\| = \sqrt{(\bar{\bar{X}} - \bar{X}_i)^2 + (\bar{\bar{Y}} - \bar{Y}_i)^2 + (\bar{\bar{Z}} - \bar{Z}_i)^2} \tag{4.4}$$

The mean and SD were obtained using the calculated distances d_i for each landmark measured by each observer as a measure of the overall inter-observer precision. The intraclass correlation coefficient (ICC) was obtained using a two analysis of variance (ANOVA) considering the choice of the observer as a two-way mixed-effects model based on absolute agreement on the mean of multiple measurements following the steps described in the ICC selection and reporting guide for reliable research (Koo & Li, 2016). The inter-observer reliability test was done in IMB SPSS version 25. This was done by comparing each average landmark position across all observers. The obtained values were compared to the precision levels shown in Table 4.1 which were defined by Mutsvangwa *et al.* (2011) in the assessment of stereo-photogrammetrically derived landmarks.

Table 4.1: Precision levels and the defined error range

Precision levels	**Error values (mm)**
Highly precise	$[0 \ to < 1]$
Precise	$[1 \ to < 1.5]$
Moderately precise	$[1.5 \ to \ < 2]$
Imprecise	$[\geq 2]$

Generation of mesh samples

After obtaining the global mean landmarks of the reference, these landmarks were visually inspected to ensure that they all lie on the mesh surface. The next step was to randomly sample the model to obtain mesh samples and their annotated landmarks. The reference mesh landmark points were used to find the corresponding landmarks for each mesh sample using the *findClosestPoint* function in Scalismo.

The model was randomly sampled for 1500 scapula meshes to obtain a training dataset that is able to train a U-net deep learning model as shown in a study by Shvets *et al.* (2018). However, the larger the training dataset the better the model would learn to generalise to unseen data (Cernazanu-Glavan & Holban, 2013; Hesamian *et al.*, 2019; Livne *et al.*, 2019; Ronneberger *et al.*, 2015). The reference landmarks were automatically annotated in the correct anatomical position on each sample mesh. Random sampling was applied to all the principal components of the model using a Scalismo in a built method called *sample*. The *sample* function was applied to the model to randomly generate instances and 3D triangular meshes were returned. Each mesh instance was saved as stereolithography (.stl) format with its annotated landmarks saved as JavaScript object notation (.json) format. The generated mesh samples had 15000 vertices with 29996 triangles.

Amira visualisation of mesh surfaces

All the obtained mesh instances and the annotated landmarks were inspected in Amira. During the inspection of the meshes, the researcher realised that 5% of the meshes had faulty triangles. The surface editor in Amira was used to initially perform two tests on the mesh surface; aspect ratio and intersection ratio. The aspect ratio of the mesh surface indicated the length of the longest edge of the triangle to the shortest edge. On the other hand, the intersection ratio indicated the number of intersecting triangles present on the mesh. According to the Amira's user guide, the mesh surface is considered to have a good quality when there are no triangle intersections and the aspect ratio should be below 20. Ideally, the best aspect ratio should be less than or equal to 4 (Berg *et al.*, 2008).

The mesh surfaces with intersecting triangles in the obtained dataset were corrected using an inbuilt Amira function called *fix intersections* on the surface edit menu. All the mesh surfaces with an aspect ratio above 20, were fixed using the prepare *tetragen* function located on the surface edit menu. However, manual editing was carried out for the mesh surfaces whose aspect ratio or triangle intersection persisted after applying the two inbuilt functions. After obtaining zero triangle intersections and an aspect ratio below 20 for any selected mesh, the mesh surface was remeshed using the *compute remesh surface* function.

The surface was remeshed in two steps: the first step was to fix contours on the whole surface. The desired size was adjusted to 50% and error threshold smoothness set to 0.6 which were applied to the mesh surface. The second step involved contracting boundary edges only around contours. The desired

size was made 100% and error threshold smoothness set to 0.6 and then applied to the mesh surface. This was followed by saving each remeshed surface in .stl format. These inspected meshes would be used to obtain the volumetric objects required to generate the DRRs.

4.3.3 Results of landmark reliability assessment

The intra-observer and inter-observer precision values are shown in Figure 4.4 and Figure 4.5, respectively. The graphs show the variability in the landmark selection process as the maximum distance, mean distance value and SD represented per landmark. The mean intra-observer precision error was approximately 0.91 mm with a range of 0.65 mm to 1.56 mm. Most of the landmarks had a precision error of less than 1mm, except for the acromion (landmark D) and base of the scapula (landmark F) which had precision errors of 1.14 mm and 1.56 mm, respectively.

The mean inter-observer precision error was about 1.19 mm with a range of 0.47 mm to 2.41 mm. The inferior angle (landmark A), infra glenoid rim (landmark B), and superior angle (landmark E) landmarks had precision errors of less than 1mm. The coracoid (landmark C), acromion (landmark D), and base of the scapula (landmark F) landmarks had a precision error above 1mm. Landmark F was the least reliable landmark to select with an inter-observer precision of 2.4 mm. The ICC obtained for the inter-observer reliability was ranked at 1.00 at a *p*-value of 0.05 and a 95% confidence interval for all the landmarks.

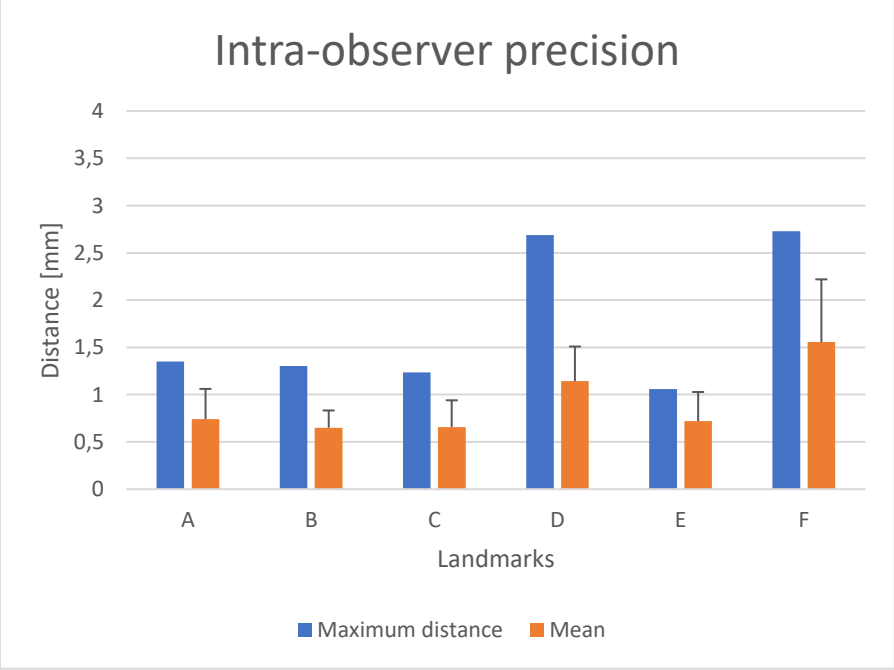

Figure 4.4: Overall intra-observer precision in the selection of 6 landmarks on the scapula reference mesh represented by the maximum distance, mean and SD per landmark.

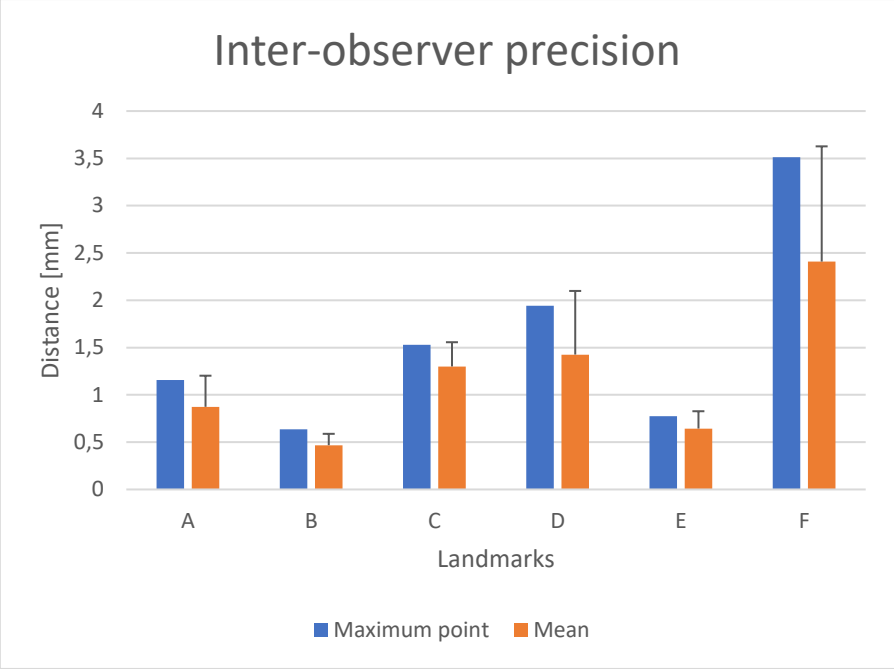

Figure 4.5: Overall inter-observer precision in the selection of 6 landmarks on the scapula reference mesh represented by the maximum distance, mean and SD.

Further details on the landmark selection precision per coordinate are shown in Table 4.2, which shows the intra and inter-observer mean and SD of the distance of the observed position to the mean position per coordinate.

Table 4.2: Intra- and inter-observer distances from the mean to the observed positions for each landmark

Landmark	Intra-observer deviations in mm Mean (SD)				Inter-observer deviations in mm Mean (SD)			
	X	***Y***	***Z***	**Combined coordinate error**	***X***	***Y***	***Z***	**Combined coordinate error**
A	0.16 (0.07)	0.28 (0.14)	0.29 (0.32)	0.43	0.15 (0.08)	0.58 (0.21)	0.61 (0.31)	0.85
B	0.34 (0.13)	0.40 (0.18)	0.09 (0.10)	0.53	0.38 (0.16)	0.20 (0.13)	0.06 (0.04)	0.43
C	0.31 (0.15)	0.49 (0.22)	0.12 (0.12)	0.59	0.97 (0.54)	0.61 (0.26)	0.25 (0.13)	1.17
D	0.65 (0.26)	0.69 (0.31)	0.12 (0.14)	0.96	0.41 (0.20)	1.19 (0.82)	0.37 (0.23)	1.31
E	0.12 (0.06)	0.11 (0.05)	0.32 (0.36)	0.36	0.30 (0.11)	0.32 (0.12)	0.45 (0.17)	0.63
F	1.41 (0.72)	0.30 (0.16)	0.20 (0.22)	1.46	2.32 (1.29)	0.28 (0.12)	0.31 (0.12)	2.36
Global mean (SD)	**0.50 (0.23)**	**0.38 (0.18)**	**0.19 (0.21)**	**0.72**	**0.76 (0.40)**	**0.53 (0.28)**	**0.34 (0.17)**	**1.13**

The intra-observer mean deviation still shows that most landmarks were selected with precision errors below 1 mm in the X, Y and Z coordinates, except for landmark F which had an error of 1.41 mm in the X direction. The inter-observer mean deviation shows five landmarks were selected with precision errors below 1mm within the X, Y and Z direction, except for landmark D, which had a high error of 1.19 mm in the Y coordinate and landmark F, which had an error of 2.32 mm the X coordinate.

The global mean errors and the SD in X, Y, Z and combined coordinate error E were 0.50 (0.23), 0.38 (0.18), 0.19 (0.21) and 0.72 mm for the intra-observer deviation and 0.76 (0.40), 0.53 (0.28), 0.34 (0.17) and 1.13 mm for the inter-observer deviation, respectively.

According to the precision levels in Table 4.1, the variations observed in the intra and inter-observer distances to the observed mean position of each landmark were within the precise range except for landmark F. This landmark was precisely selected according to intra-observer precision but was imprecisely selected across different observers. Most of the combined coordinate landmarks were selected with high precision and excellent reliability according to defined precision levels in Table 4.1 and the obtained the ICC reliability value.

4.3.4 Scapula volumetric objects

The scapula volumetric objects are also referred to as CT volumes and are obtained from the scapula mesh samples. The process of obtaining volumetric objects was accomplished using the DRR renderer (Reyneke, 2019). Each scapula mesh sample was loaded into Scalismo and converted into binary images using the *toBinaryImage* function in Scalismo. The dimension of the 3D binary images was 207 mm x 89 mm x 135 mm in X, Y, and Z, respectively. The binary 3D images had voxel spacing of (1 x 1 x 1) mm. All the binary 3D images were bound by the same bounding box so that they would have the same sizes when projected by the renderer. Each binary 3D image was saved as Neuroimaging Informatics Technology Initiative (NIfTI) format. The binary 3D images were converted into Hounsfield voxel volumes using the *binaryToHU* function in DRR renderer. The Hounsfield voxel volumes were also of dimension 207 mm x 89 mm x 135 mm in X, Y, and Z, respectively. The voxel spacing was maintained at (1 x 1 x 1) mm and the intensity range was -1024 and 2979. These volumes were saved as .NIfTI files for later processing into DRRs. These Hounsfield voxel volumes are referred to as CT volumes or volumetric objects in this report.

4.3.5 Binary synthetic bi-planar X-ray images

Binary synthetic bi-planar X-ray images are henceforth referred to as 2D bi-planar images. The 2D bi-planar images were used to train the CNN model to learn the mapping of the scapula contour in the LAT image view given the AP view and vice versa. The synthetic X-ray images were generated by projecting the volumetric objects obtained in section 4.3.4 along with their annotated landmarks using

the DRR. To combine each volumetric object and its landmarks, the landmarks were replaced by virtual fiducial markers using the *binaryToHU* function in the DRR renderer. This function converted each landmark into a virtual fiducial marker of a unique Hounsfield colour value. The virtual fiducial markers were merged with the volumetric objects using *markVolume* function in the DRR renderer. Each volumetric object and the virtual fiducial markers (landmarks) were rendered within the same defined bounding box using the DRR *projector* function. The *projector* function projected the volumetric objects in three different projection types; orthogonal, perspective and fan-beam projection. For this project, the orthogonal projection was applied to the volumetric objects to generate the DRRs. The volumetric objects of the scapula were rendered at 0^0 and 90^0 to mimic the commonly imaged perspectives of the scapula in the clinical settings. Thus, the terms anterior-posterior (AP) view and lateral (LAT) view apply in this project to the 0^0 and 90^0 views, respectively. The peak kilovoltage (kVp) was set to a high value (above 5.0) to decrease the contrast of the DRRs to obtain silhouettes of the scapula in each view.

The original orientation of the SSM at 0^0, 0^0 and 0^0 for the roll, pitch and yaw angles resulted into a model that was not in an erect anatomical position thus it had to be adjusted to obtain the erect position of the scapula samples in the body. The AP image of the scapula volumetric object was obtained by adjusting the roll, pitch and yaw angles of the object to 90.0^0, -90.0^0 and 0.0^0, respectively. This was followed by projection of the LAT image of the scapula volumetric object where the roll, pitch and yaw were 90.0^0, 0.0^0 and 0.0^0, respectively. The rendered images were saved in a portable network graphic (PNG) format with 207 x 207 pixels in size and the landmarks that were selected on the 3D mesh could be identified on these images. A total of 1500 image pairs were generated to enable training of a deep learning model.

The landmarks in the rendered images were detected using the *detectLandmarks* function in the DRR renderer and saved to. json files. These landmarks were the 2D landmark positions for the 3D landmarks that were annotated on each scapula mesh sample. These 3D landmarks were rendered along with the volumetric objects to enable validation of the trained U-net model performance both qualitatively and quantitatively. Although the images rendered were binary in nature, the super-imposed landmarks were rendered with colours as unique identifiers to ease the selection of corresponding points. Figure 4.6 shows the rendered bi-planar images of the volumetric objects for the AP and the LAT view respectively; with the 2D landmarks of the 3D landmarks that were initially selected on the mesh.

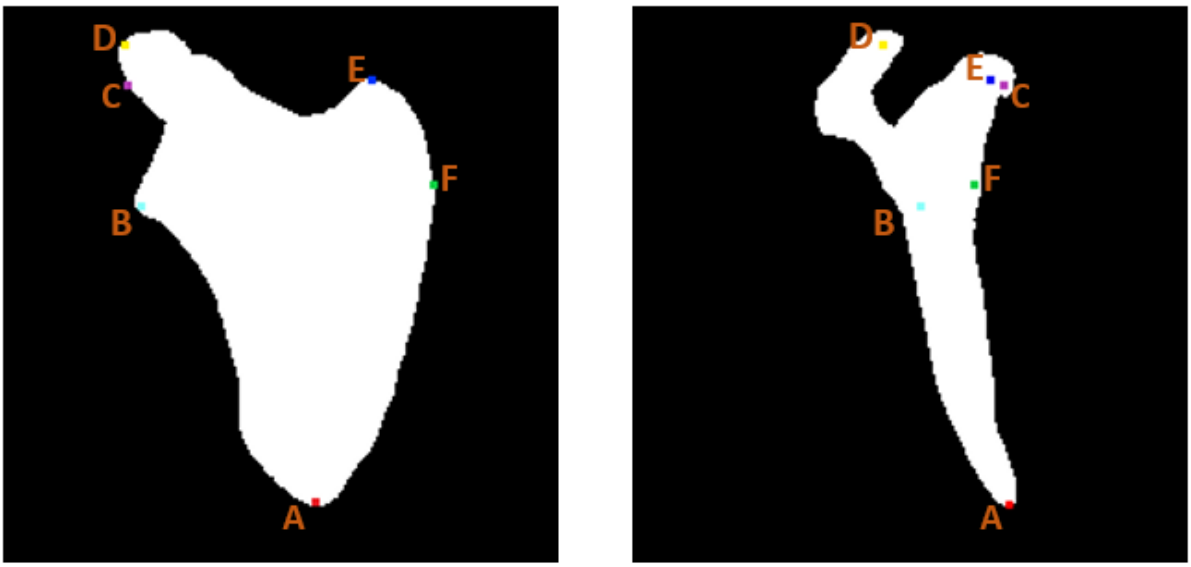

Figure 4.6: Rendered bi-planar images of the volumetric objects for the anterior-posterior and the lateral view respectively, with the landmarks (magnified for better viewing).

4.3.6 Bounding box

A bounding box is an imaginary box that encompasses all elements of an object of interest. In this research, the objects encompassed by the bounding box were the scapula mesh samples and the volumetric objects. The bounding box was used in two different ways. The first bounding box was used as a reference boundary to generate a virtual calibration frame for DRR projection. The second bounding box was to generate a virtual calibration frame for the acquisition of transformation parameters for 3D point localisation using direct linear transformation (DLT) transformation. A scapula describing the biggest domain of the SSM was used to obtain the bounding boxes using the steps shown in Figure 4.7. This was to ensure that all the DRRs are projected with in the same frame to avoid extrapolation during 3D point localisation.

The biggest domain was obtained by adjusting the first mode (principal component (PC1)) of the SSM model was adjusted to +3 standard deviation (SD) which represented the largest size variation of all the mesh samples in the dataset. This was followed by adjusting the second mode (PC2) to -3 (SD) which represented the longest height of the scapula. The resultant sample was saved as a mesh instance and later converted to a synthetic CT volume using the *toBinaryImage* function in Scalismo.

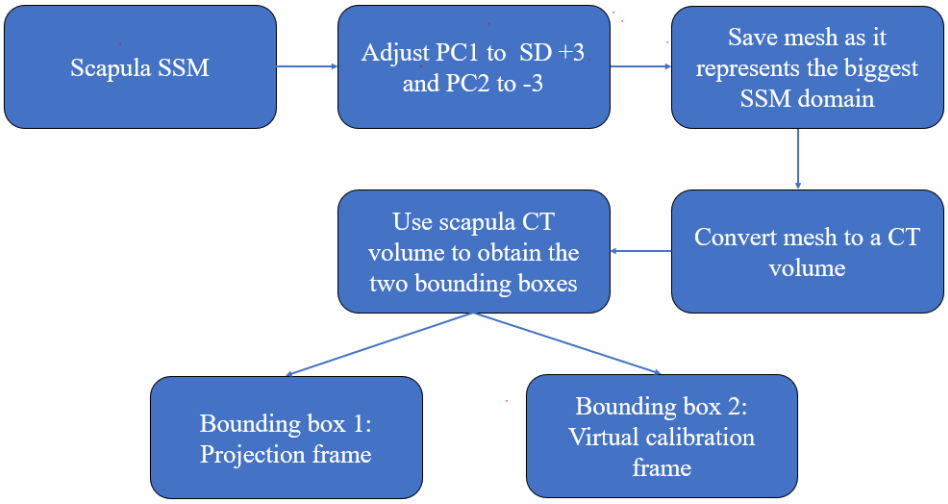

Figure 4.7: Steps taken to obtain the scapula SSM biggest domain.

Figure 4.8 shows the steps taken to obtain the two bounding boxes after obtaining the scapula CT volume.

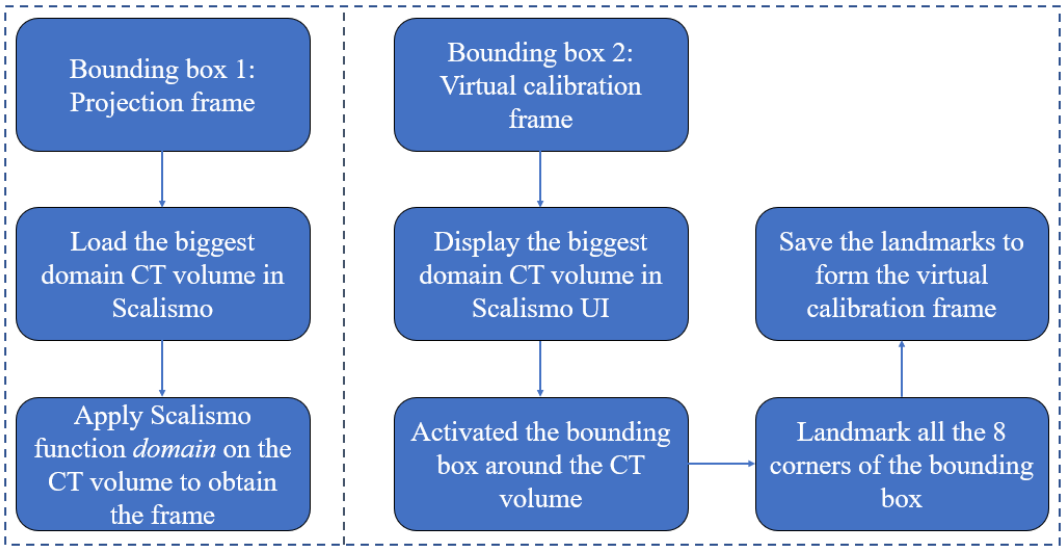

Figure 4.8: Steps taken to obtain the scapula CT volume bounding boxes.

To obtain the first bounding box, the generated CT volume was used to generate a bounding box using a Scalismo in-built function *domain*. This bounding box was used as a projection frame for each mesh sample during its conversion into a volumetric object, followed by projection into 2D bi-planar images.

To obtain the second bounding box which was a virtual calibration frame used to calculate the projection transformation parameters that are required to transform 2D landmark points to their 3D representation. The generated CT volume for the biggest scapula mesh instance was displayed in the Scalismo user interface. This was followed by selecting the scene menu and checking the display bounding box button for its visualisation. The displayed bounding box around the scapula was landmarked for all the 8

corners of the cuboid. The selected landmarks were saved as .json files and rendered in the same way as the scapula volumetric objects to find the 2D positions of the 3D landmarks in the two image views.

After the generation of the datasets, the next step was to implement the proposed methodology.

5 Contour detection of the scapula in bi-planar X-ray images

This chapter describes how objective 1 was achieved. It is divided into two main sections. The first section describes experiments on the synthetic data generation in section 4.3 (see previous chapter) and data preparation in section 5.1. Experiment one of the first section involved training the U-net model to predict the scapula contour in the lateral (LAT) image given the scapula contour in the anterior-posterior (AP) image (AP to LAT model). Experiment two of section one was training the U-net model to predict the contour of the scapula in the AP image given the scapula contour in the LAT image (LAT to AP model). The second main section of this chapter describes testing on real image data. This involved obtaining of bi-planar X-ray images of the upper-torso of a cadaver, isolating the scapula by manual segmentation and using the segmented scapula images to test the performance of the trained U-net models on real data.

5.1 Preparation of the generated dataset for training the U-net models

Data preparation was an essential step to organise the data into a structure suitable to train the U-net models. In this phase, the rendered binary 2D bi-planar images of the scapula generated in section 4.3.5 were resized to 256 x 256 pixels and converted into grayscale images using OpenCV. The data was resized and converted to grayscale to fit the design of U-net input layer. The processed dataset was made up of 1500 binary bi-planar images. The images were divided into 80% training set and 20% test set (Newman, 2005). The training set was further split into training and validation sets in a ratio of 80% and 20%, respectively. The validation and test sets are also known as the evaluation sets in this research project.

5.2 U-net model training

The U-net CNN model described in section 3.5.2 was trained to learn the mapping of the scapula contour in the LAT image view given the AP image view and vice versa. Two experiments were conducted as shown in Figure 5.1.

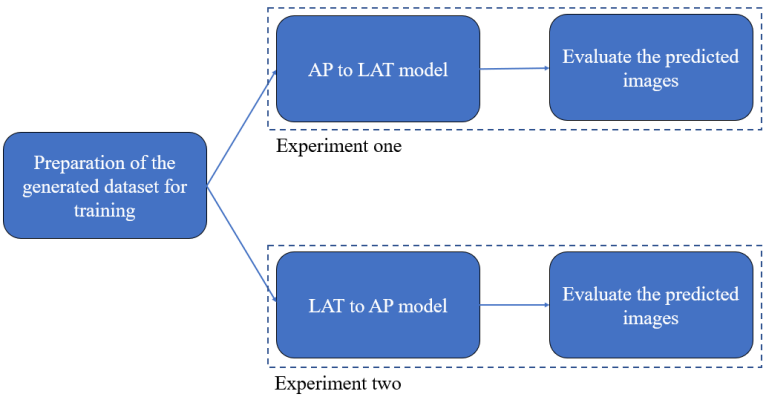

Figure 5.1: Overview of model training experiments.

Table 5.1 shows the dataset used in objective one for model training and evaluation

Table 5.1: Dataset used for Training and evaluating the models

Model	Training and testing images			Evaluation images	
	Training Dataset (80%) - 1200 image pairs		Test set (20%) - 300 image pairs	Dice coefficient	Landmark error
	Training Dataset (80%)	Validation set (20%)			
LAT to AP	960	240	300	300 predicted AP images	30 randomly selected predicted AP images
AP to LAT	960	240	300	300 predicted LAT images	30 randomly selected predicted LAT images

5.2.1 Experiment one: AP to LAT model

The AP to LAT U-net model was trained to predict the contour of the scapula in the LAT image given the scapula contour in the AP image. The training set had inputs and labels totalling 1200 image pairs. The binary AP images were the input training set and the corresponding binary LAT images were the labels in the training set. The test set was made up of 300 (20% of the total training data) binary AP images only. The LAT images corresponding AP test images were used at a later stage as the ground-truth images to evaluate the predicted result from the trained U-net model. The U-net model was implemented using the Keras with a TensorFlow backend along with numerous dependencies such as NumPy, OpenCV, and SciPy in Python 3. The architecture of the model had an input size of 256 x 256 and each layer had a rectified linear unit (ReLU) activation function, except for the output layer which had a sigmoid activation function. The output layer also had an output size of 256 x 256. The size of the output and input was maintained by padding each convolutional layer. This model used a binary cross-entropy loss function, sigmoid activation function and an Adam optimizer with a learning rate of 0.0004. The model was trained using binary AP images to predict the binary LAT images with a batch size of one image for 50 epochs.

The trained model was saved and used to predict the scapula contour in LAT images given the scapula contour in the AP test images.

5.2.2 Experiment two: LAT to AP model

The model training process for experiment one of this chapter was repeated using the same training-evaluation dataset ratio, with the LAT images as the input and the corresponding AP images as the labels for the training process. The model in this experiment was trained to predict the scapula contour in the AP image given the scapula LAT image. The architecture of this model was set up in the same

way as the one for experiment one. The model was also trained with a batch size of one image for 50 epochs. The trained model was saved and used to predict the scapula contour in AP images given the scapula contour in the test LAT images.

5.2.3 Evaluation of predicted contours

The scapula contour images predicted by the U-net trained model were evaluated against their corresponding ground-truth images for both experiments. Two evaluation metrics were used, namely the Dice coefficient and landmark distance error.

Dice coefficient was applied to measure the similarity between the two images (predicted and ground-truth) by comparing pixels in the images to ascertain the ratio of the true-positive (TP) pixels, false-positive (FP) pixels, and false-negative (FN) pixels as shown by equation (3.26).

Landmark distance error was used to evaluate the position of the landmarks in the predicted images relative to the landmarks in the ground-truth images. The landmark distance error method described in section 3.6.1 was used. To evaluate the relative position of the fiducial markers in the predicted images, a sub-set of 30 images were randomly selected from the predicted images. The landmark points in the selected predicted images were required. The landmark point positions in the images predicted by U-net model were extracted using a landmarking software with an interactive graphical user interface (GUI), Landmark-clicker, based on Scalismo (https://github.com/unibas-gravis/landmarks-clicker). The software was used to select the landmark points in the images and return their pixel coordinate position. The landmarks in the predicted images were selected and saved to a .json file. These landmarks were compared to the landmarks of the ground-truth images and the distance error calculated using equations (3.27) - (3.30) for the 6 landmarks.

5.2.4 Results: AP to LAT model

The AP to LAT model training was evaluated using a binary cross-entropy metric. This metric resulted in training and validation accuracy of 99.5% and 98.4%, respectively. Figure 5.2 indicates the trend of the training and validation accuracy per epoch.

The training took approximately 307 milliseconds per step, 295 seconds per epoch and 246 minutes for 50 epochs which amounted to approximately 4.1 hours for the training.

The testing took 72 milliseconds per step. The average Dice coefficient of the 300 predicted images was 0.926 with a SD of 0.024 and a range of 0.959 to 0.776.

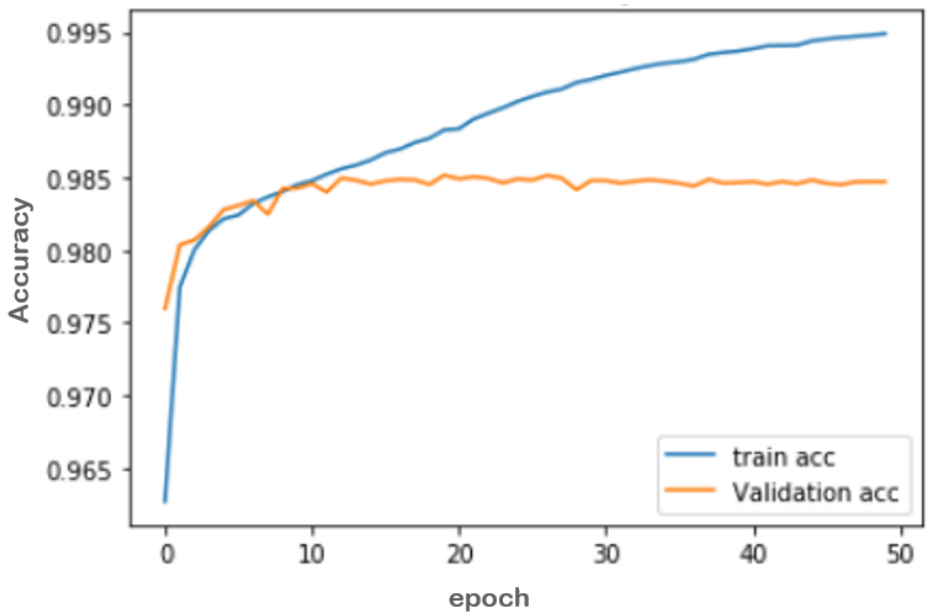

Figure 5.2: Training and validation accuracy per epoch.

Figure 5.3 shows the best and worst Dice coefficient (Dice) values of the predicted scapula contour LAT images and the corresponding ground-truth image on the left of the prediction.

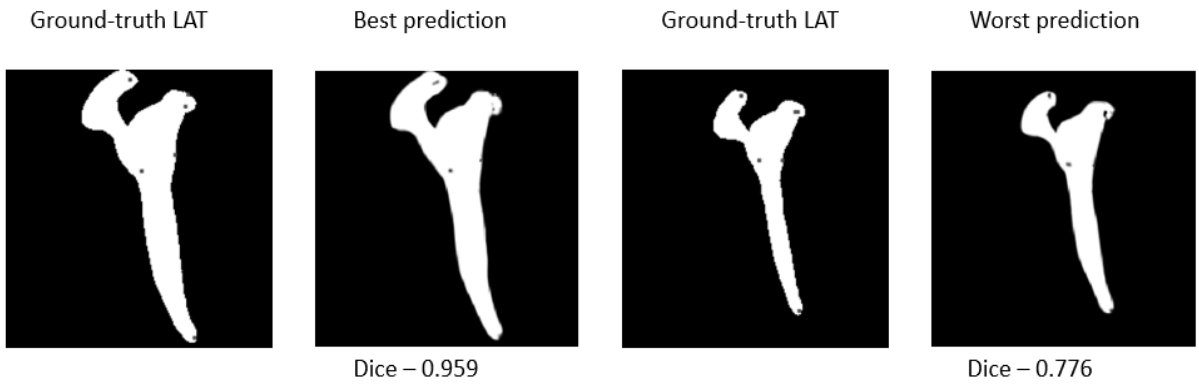

Figure 5.3: The ground-truth images of the best and worst predicted lateral images.

The global mean landmark error for the e_x, e_y and, combined coordinate resultant error, e were 1.64, 1.37 and 2.31 pixels, respectively. The results shown in Table 5.2 include the mean distance errors, e_x and e_y and their SD obtained per landmark coordinate, the combined coordinate resultant error e_i and the mean resultant error, e.

The training was stopped at 50 epochs as the validation accuracy had flattened. However, according to the overall trend, after 15 epochs the additional training epochs had little effect on the accuracy. The results suggest that the U-net model can learn the mapping of the LAT image given the AP image.

Table 5.2: Landmark errors for predicted lateral images from the AP to LAT model in pixels.

Predicted LAT landmarks	Error, e_x (SD) $[px]$	Error, e_y (SD) $[px]$	Resultant error, e_i $[px]$
A	1.76 (0.99)	1.06 (0.79)	2.22
B	1.38 (0.80)	1.04 (0.66)	1.87
C	2.02 (0.75)	1.43 (1.18)	2.67
D	1.47 (0.91)	1.31 (0.80)	2.15
E	1.81 (0.86)	1.63 (0.79)	2.56
F	1.43 (0.76)	1.75 (0.75)	2.38
Global mean (SD)	**1.64 (0.84)**	**1.37 (0.83)**	**2.31**

5.2.5 LAT to LAT model

In this section the U-net model was used to predict the scapula from the full grayscale LAT image and the results compared to the AP to LAT model predictions in 5.2.4. To train the LAT to LAT model the scapula volumetric objects obtained in section 4.3.4 were projected according to the steps taken in section 4.3.5 to obtain the binary synthetic bi-planar images. However, peak kilovoltage (kVp) was set to a low value (2.0) to increase the contrast of the digitally reconstructed radiograph (DRR) in order to obtain full-grayscale intensities of the scapula in each view. Figure 5.4 shows scapula images of the full grayscale and the binary LAT projections.

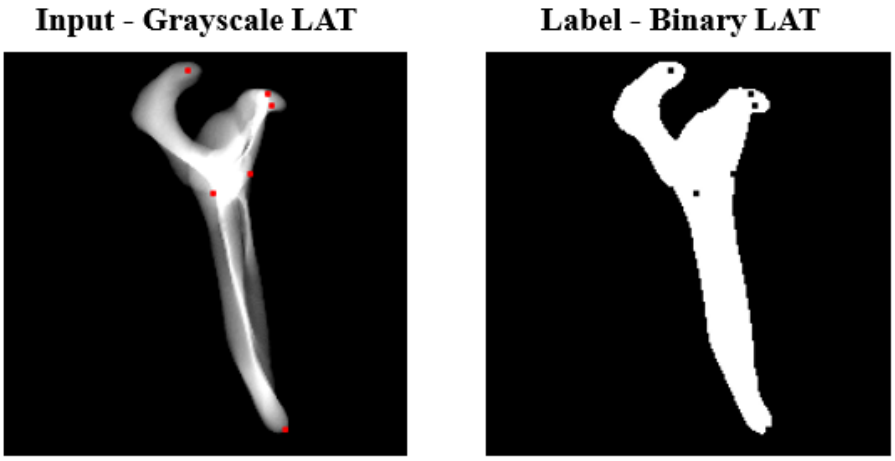

Figure 5.4: Generated training dataset for training the U-net segmentation model.

The generated data were processed following the steps taken in section 5.1. The processed images combined with the corresponding images from the dataset generated in section 4.3.5 were used to create the LAT to LAT model training dataset. The full-intensity LAT images were paired with the corresponding binary scapula LAT images generated from the same mesh sample. The training set had inputs and their labels totalling up 1200 image pairs (80% of the total dataset). The full-grayscale intensity LAT images were the input training set and corresponding binary LAT images were the labels in the training set. The validation set consisted of 20% of the training set. Lastly, the test set (20% of

the total training set) was made up of 300 LAT full-grayscale intensity images only. The binary LAT images corresponding the full-grayscale intensity test images were later used as the ground-truth images to evaluate the predicted result from the trained U-net model. The LAT to LAT model was implemented in the same way as the model implemented in section 5.2.1 and section 5.2.2 except that the model in this section was trained to predict binary LAT images given full grayscale LAT images. The training process had a batch size of one image for 5 epochs to obtain the best results.

The trained model was saved and used to predict the scapula contour in LAT images given the LAT full-grayscale intensity test images.

5.2.6 Results: LAT to LAT model

The trained model was evaluated using a binary cross-entropy function. This metric resulted into training and validation accuracy of 99.9% and 99.7%, respectively. The training took approximately 304 milliseconds per step, 346 seconds per epoch and 28.8 minutes for 5 epochs which amounted to approximately 0.48 hours for the training.

The scapula contour in the LAT images predicted by the U-net trained model were evaluated against their corresponding ground-truth LAT images by using two evaluation metrics used to evaluate the models in section 5.2.3. The average Dice coefficient of 300 predicted images was 0.982 with a SD of 0.003 and a range of 0.987 and 0.969. The average landmark distance error e_x, e_y and e were 0.50, 0.48 and 0.78 pixels. The detailed results shown in the Table 5.3 include the mean distance errors, e_x and e_y and their SD obtained per coordinate, the combined coordinate resultant error e_i and the average resultant error, e.

Table 5.3: Landmark errors for predicted lateral images from the LAT to LAT model in pixels.

Segmented LAT landmarks	Error, e_x (SD) $[px]$	Error, e_y (SD) $[px]$	Resultant, e_i $[px]$
A	0.55 (0.38)	0.41 (0.28)	0.75
B	0.34 (0.25)	0.27 (0.18)	0.48
C	0.38 (0.33)	0.47 (0.43)	0.66
D	0.32 (0.32)	0.43 (0.37)	0.59
E	0.71 (0.92)	0.68 (0.76)	1.14
F	0.70 (0.56)	0.60 (0.39)	1.02
Global mean (SD)	**0.50 (0.46)**	**0.48 (0.40)**	**0.78**

However, a comparison of the LAT to LAT model results in section 5.2.6 to the AP to LAT model in section 5.2.4, U-net model performs the scapula prediction task with higher accuracy given the same dataset used. When this model was trained to predict the LAT scapula images from the LAT full-grayscale intensity images, the average Dice coefficient of 300 predicted images was 0.982 which is higher compared to 0.926 obtained from the AP to LAT model. Besides, the average landmark errors

for this model were 0.50, 0.48, and 0.78 pixels which are very small compared to 1.64, 1.37, and 2.31 pixels for the e_x, e_y, and resultant error, e, respectively. These results show that the AP to LAT model would require fine-tuning of parameters to achieve a higher accuracy for a mapping task especially due to the presence of pixels that are hard to classify. The difficulty in pixel classification is due to the label (LAT) and area of interest in the input (AP) images not having an intersection over union of 1 when super-imposed on to each other as it for most image prediction tasks.

5.2.7 Results: LAT to AP model

The LAT to AP model training was also evaluated using a binary cross-entropy metric. This metric resulted in training and validation accuracy of 99.4% and 98.1%, respectively. Figure 5.5 shows the trend of the training and validation accuracy per epoch.

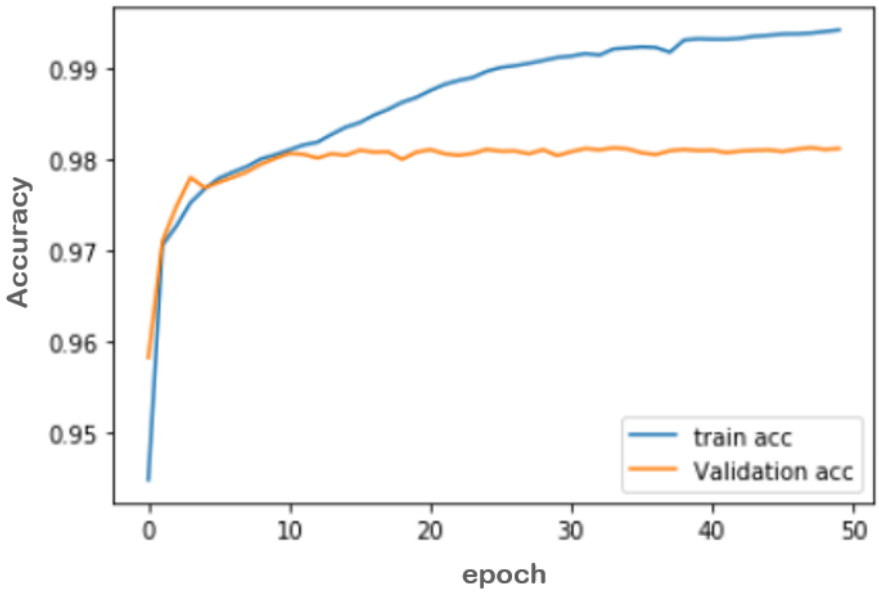

Figure 5.5: Training and validation accuracy per epoch.

The training took approximately 308 milliseconds per step, 296 seconds per epoch and 247 minutes for 50 epochs which amounted to approximately 4.1 hours for the training. The testing took 72 milliseconds per step. The average Dice coefficient of the 300 predicted images was 0.964 with a SD of 0.01 and a range of 0.981 and 0.922. Figure 5.6 shows the best and worst Dice coefficient (Dice) values of the predicted scapula contour AP images and with the corresponding ground-truth images to the left of each prediction.

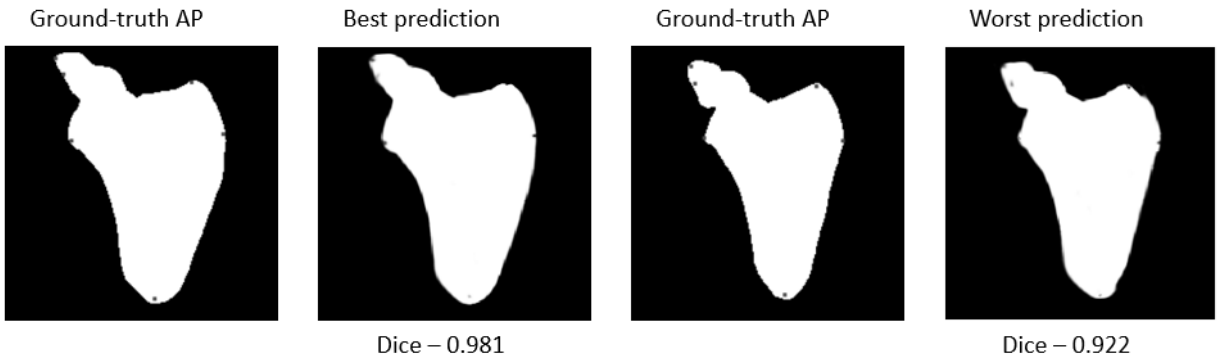

Figure 5.6: The best and worst predicted AP images and their ground-truth images.

The average landmark error for the e_x, e_y and combined coordinate resultant error, e was 1.69, 1.24 and 2.25 pixels, respectively. The results shown in Table 5.4 include the mean distance errors, e_x and e_y, SD per landmark coordinate, the combined coordinate resultant error e_i and the average resultant error, e. All results were calculated using equations (3.27) - (3.30) in section 3.6.1 of the theoretical review.

Table 5.4: Landmark errors for predicted anterior-posterior images from the LAT to AP model in pixels.

Predicted AP landmarks	Error, e_x (SD) $[px]$	Error, e_y (SD) $[px]$	Resultant error, e_i $[px]$
A	1.69 (0.79)	1.49 (0.92)	2.38
B	1.57 (0.84)	0.96 (0.62)	1.94
C	1.81 (0.86)	1.95 (1.02)	2.78
D	1.43 (0.89)	1.05 (0.66)	1.93
E	1.98 (1.05)	0.84 (0.71)	2.28
F	1.68 (0.93)	1.18 (0.86)	2.20
Global mean (SD)	**1.69 (0.89)**	**1.24 (0.80)**	**2.25**

The training was stopped at 50 epochs as the validation accuracy had flattened. However, the training should have also been stopped at 15 epochs started since increasing the epochs had little effect on the training results.

5.2.8 AP to AP model

The procedure in section 5.2.5 was repeated with full-grayscale AP images as the inputs and the corresponding binary AP images as the labels in the training set. Figure 5.7 shows scapula images of the full grayscale and the binary AP projections.

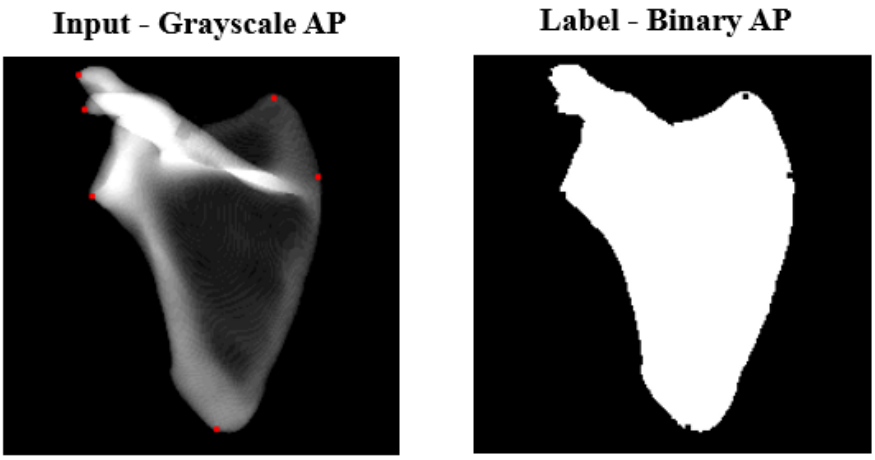

Figure 5.7: Generated training data for training the U-net segmentation model.

The trained model was saved and used to predict the scapula contour in AP images given the scapula contour in the full grayscale intensity test images.

5.2.9 Results: AP to AP model

The trained model had a binary cross-entropy accuracy of 0.998 and 0.996 for training and validation accuracy, respectively. The training took approximately 305 milliseconds per step, 346 seconds per epoch and 28.8 minutes for 5 epochs which amounted to approximately 0.48 hours for the training.

The scapula contour AP images predicted by the U-net trained model were evaluated against their corresponding ground-truth AP images by using two evaluation metrics used to evaluate the models in section 5.2.3. The average Dice coefficient of 300 predicted images was 0.988 with a SD of 0.002 and a range of 0.991 and 0.972. The average landmark distance errors e_x, e_y and e were 0.59, 0.61 and 0.93 pixels (px). The results shown in Table 5.5 include the mean distance errors, e_x and e_y and the SD obtained per coordinate, the combined coordinate resultant error, e_i and the average resultant error, e.

Table 5.5: Landmark errors for predicted anterior-posterior images from the AP to AP model in pixels.

Segmented AP landmarks	Error, e_x (SD) $[px]$	Error, e_y (SD) $[px]$	Resultant, e_i $[px]$
A	0.81 (0.60)	0.79 (0.42)	1.19
B	0.42 (0.40)	0.48 (0.39)	0.72
C	0.85 (0.58)	0.52 (0.42)	1.06
D	0.46 (0.39)	0.62 (0.34)	0.84
E	0.53 (0.45)	0.38 (0.32)	0.72
F	0.47 (0.43)	0.45 (0.32)	0.72
Global mean (SD)	**0.59 (0.48)**	**0.54 (0.37)**	**0.87**

Comparing the results of AP to AP model to the results of the LAT to AP model shown in section 5.2.7; results indicate that the U-net model performs the AP prediction task in the AP to AP model with a

higher accuracy given the same dataset. The AP to AP model predictions had an average Dice coefficient of 0.988 which is higher than 0.964 obtained for the LAT to AP model. Furthermore, the average landmark errors for the AP to AP model were 0.59, 0.54 and 0.87 pixels which more accurate when compared to 1.69, 1.24 and 2.25 pixels for the LAT to AP model for the e_x, e_y, and resultant error, e, respectively. These results show that although the U-net model can perform the mapping between the binary images, the model would require fine-tuning to obtain higher accuracy for a mapping task.

5.3 Testing trained U-net models with real data

This section shows the performance of the trained U-net models on real cadaveric X-ray data of the upper-torso. Ethical approval was acquired from the University of Cape Town research ethics committee to use the bi-planar X-ray images of the upper-torso of a cadaver collected as part of previous research (Wasswa, 2016). This image pair was used to test the ability of the trained U-net model to generalize on real data. However, before testing the trained U-net models, the image pair was manually segmented to obtain the scapula silhouette in both the AP and LAT views, because the U-net models had been trained on binary images of an isolated scapula. The segmented AP image was used to predict the LAT image and vice versa.

5.3.1 Results

The model predictions gave a Dice coefficient of 0.248 and 0.006 for the predicted AP and LAT images, respectively. Figure 5.8 shows the AP and LAT segmentations and the corresponding predicted images, respectively.

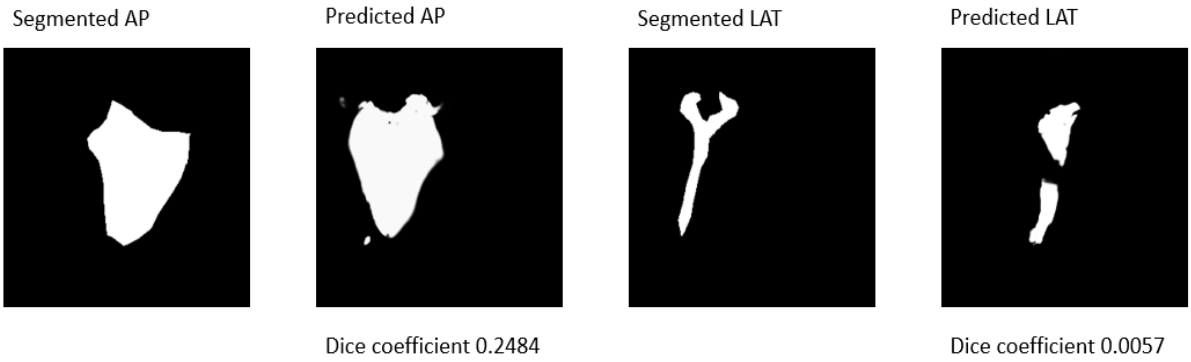

Figure 5.8: Scapula segmented AP, predicted AP, segmented LAT and the predicted LAT, and their Dice coefficient values.

As shown in Figure 5.8 the model did not perform as expected on the segmented images. Several reasons could explain the inability of the trained models to generalize to the segmented scapula images. One reason might be poor manual segmentation of the bi-planar X-ray images, due to poor visual identification of the scapula in the X-ray image pair. Another reason for the poor results may be that

the models had been trained on images with smooth scapula contours as shown in Figure 4.6, while the manually segmented contours were not smooth. Finally, although the position of the scapula in the image should not affect the results, the difference in relative orientation and shape of the scapula itself compared to the scapula in the images used to train the model could be another reason for the poor results obtained. The real bi-planar X-ray images were obtained at 0^0 and 75^0 for the AP and LAT views, respectively, while the synthetic data were projected at 0^0 and 90^0 for the AP and LAT views, respectively.

5.4 Conclusion

Although the Dice coefficient of the LAT to LAT and AP to AP models were higher than that of the models trained to learn the mapping between AP and LAT images of the scapula, all the models had high Dice coefficient values above 0.92. These results are similar to those found in previous work done using U-net, for example, Shvets *et al.* (2018) trained the U-net model for angiodysplasia detection and localization with 1200 images and obtained a Dice coefficient value of 0.831 (Shvets *et al.*, 2018). Livne *et al.* (2019) achieved a Dice coefficient value of approximately 0.891 with training on 81,000 image patches for segmentation of the vessels in cerebrovascular disease affected patients (Livne *et al.*, 2019). These results show the robustness of the U-net model to perform accurately given different image prediction tasks.

The U-net model performs image prediction tasks with intersection over union between the input and the labelled area of interest of 1 with higher accuracy which is the main reason it has been implemented for different medical image segmentation tasks (Andersson *et al.*, 2019; Livne *et al.*, 2019; Ronneberger *et al.*, 2015; Shvets *et al.*, 2018). However, the results obtained also show that the model can also learn the AP to LAT mapping and vice versa. The LAT to AP model slightly outperformed the AP to LAT model but both models did not generalise to real data; this might be attributable to poor manual segmentation of the scapula in real images, differences in contour smoothness between real and synthetic images, and differences in image orientation between real and synthetic images.

6 Three-dimensional reconstruction of bi-planar X-ray images using embedded corresponding landmarks

This chapter describes the use of landmark-constrained statistical shape model (SSM) fitting to reconstruct scapulae to 3D. The landmarks used are the set of corresponding landmarks on the predicted contours from the previous chapter. Specifically, the chapter highlights two experiments. Experiment one focuses on three-dimensional (3D) landmark localisation using the direct linear transformation (DLT). Experiment two focuses on 3D scapula mesh reconstruction by constraining the shape model using the reconstructed 3D landmarks.

6.1 Experiment one: 3D projective transformation

The DLT was used for 3D landmark localisation of the detected 2D landmarks from the images predicted by the U-net model whose landmarks were evaluated in Table 5.2 and Table 5.4.

6.1.1 Calculation of transformation parameters using the calibration frame

The transformation parameters required to transform the 2D landmarks obtained from the images predicted using the U-net model, were calculated. The transformation parameters were obtained through use of a virtual calibration frame. The virtual calibration frame was designed using the second bounding box and its 2D projections as described in section 4.3.6.

Given the DLT equations (3.4) that define the mapping of 2D points to 3D space in section 3.4.1, the transformation parameters for the first image view were obtained by simplifying u and v in equation (3.4) to equations (3.5) and (3.6), respectively. Equations (3.5) and (3.6) are expressed in matrix form to obtain equation (3.9). Given eight control points (points with known 3D and 2D points of the bounding box), the first image view calibration parameters (L_{ij}) were obtained using equation (3.12). This procedure was repeated to obtain the calibration parameters (L_{ij}^{i}) of the second view.

Given equation (6.1) the pseudo-inverse of the calculated transformation parameters from equation (3.10) was used to transform the 2D corresponding landmark points in the bi-planar images to 3D points. The 2D points were measured in pixels and the corresponding 3D points measured in millimetres (mm).

$$\begin{bmatrix} u - L_{14} \\ v - L_{24} \\ u^{i} - L_{14}^{i} \\ v^{i} - L_{24}^{i} \end{bmatrix} = \begin{bmatrix} L_{11} - L_{31}u & L_{12} - L_{32}u & L_{13} - L_{33}u \\ L_{21} - L_{31}v & L_{22} - L_{32}v & L_{23} - L_{33}v \\ L_{11}^{i} - L_{31}^{i}u & L_{12}^{i} - L_{32}^{i}u & L_{13}^{i} - L_{33}^{i}u \\ L_{21}^{i} - L_{31}^{i}v & L_{22}^{i} - L_{32}^{i}v & L_{23}^{i} - L_{33}^{i}v \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \tag{6.1}$$

If the matrix on the left is denoted as UV and the first matrix on the right as C, equation (6.1) can be written as:

$$UV = C \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \tag{6.2}$$

The points X, Y, Z can be found from equation (6.2) by obtaining the pseudo-inverse of C:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = UV * pinv\,(C) \tag{6.3}$$

The calculated transformation parameters were evaluated using control and test point reconstruction as shown in previous studies (Chimhundu *et al.*, 2014; Douglas *et al.*, 2004; Wasswa, 2016).

Control points were used to test the mathematical correctness of the formula used. In this case, the projected 2D landmark points of the bounding box and the calculated transformation parameters were used to reconstruct the 3D landmark points (X_{ri}, Y_{ri}, and Z_{ri}). The reconstructed 3D landmark points were compared to the known 3D landmarks points (X_i, Y_i, and Z_i) of the bounding box that were used to calculate the transformation parameters.

After ascertaining the mathematical correctness of the method used to obtain the transformation parameters, which were used to obtain the unknown 3D coordinates (X_{ri}, Y_{ri}, and Z_{ri}) of the 2D landmarks in the predicted images, in test point reconstruction to evaluate the calculated transformation parameters used.

Given the transformation parameters (C) calculated using the bounding box, the extracted 2D landmarks from predicted lateral (LAT) image (u', v') in section 5.2.3 and section 5.2.4 and their corresponding anterior-posterior (AP) test image landmark points (u, v), the 3D landmark points were calculated using equation (6.3). The reconstructed 3D landmark points were compared to the ground-truth 3D landmark points that were initially annotated on each mesh sample. Reconstruction errors E_x, E_y, E_z, and E were calculated using equations (3.31) - (3.35) to find the landmark reconstruction error. This procedure was repeated for the 2D points of the AP predicted images extracted in section 5.2.7 and the corresponding LAT test images to find their 3D coordinate points.

6.1.2 Results: 3D projective transformation

The average reconstruction errors for eight control points were 0.35 mm, 0.64 mm, 0.72 mm and 1.16 mm for E_x, E_y, E_z, and E, respectively. Table 6.1 shows the detailed control point reconstruction errors per reconstructed point.

Table 6.1: Control point reconstruction error for the bounding box landmarks

Bounding landmarks	Error, E_x $[mm]$	Error, E_y $[mm]$	Error, E_z $[mm]$	Resultant error, E_i $[mm]$
1	0.11	0.05	0.61	0.62
2	0.37	0.03	0.61	0.71
3	0.83	0.05	0.46	0.95
4	0.56	0.06	0.46	0.72
5	0.16	1.20	0.98	1.56
6	0.44	1.24	0.96	1.63
7	0.04	1.28	0.83	1.53
8	0.30	1.25	0.84	1.54
Global mean (SD)	**0.35 (0.24)**	**0.64 (0.60)**	**0.72 (0.20)**	**1.16 (0.42)**

The control points and the resultant 3D localized points of the bounding box are shown in Table 6.2.

Table 6.2: Selected 3D control points on the bounding box and their corresponding reconstructed points in mm

Selected points on bounding box	X_i	Y_i	Z_i	X_{ri}	Y_{ri}	Z_{ri}
1	-216.44	-151.11	-1833.06	-216.21	-152.58	-1833.39
2	-216.60	-151.82	-1698.06	-216.81	-150.43	-1698.19
3	-8.60	-150.38	-1832.93	-8.63	-149.19	-1832.70
4	-9.85	-142.35	-1699.00	-9.84	-143.54	-1698.72
5	-216.55	-61.00	-1697.87	-216.38	-61.21	-1697.72
6	-215.58	-61.00	-1832.42	-215.77	-60.71	-1832.08
7	-7.83	-61.00	-1698.86	-7.83	-61.02	-1699.16
8	-6.28	-61.00	-1833.42	-6.27	-61.00	-1833.64

The reconstruction of the 2D test points extracted from the predicted LAT images in section 5.2.4 and the corresponding AP test images resulted in average reconstruction errors of 0.83 mm, 1.40 mm, 0.67 mm and 1.99 mm for E_x, E_y, E_z, and E, respectively. In addition, the reconstruction of the 2D points extracted from the predicted AP images in section 5.2.7 and the corresponding LAT test images resulted in average reconstruction errors of 0.93 mm, 1.26 mm, 0.89 mm and 1.96 mm for E_x, E_y, E_z, and E, respectively. The details of the errors and SD per landmark E_x, E_y, E_z and E for the AP to LAT and LAT to AP models are shown in Table 6.3 and Table 6.4, respectively.

Table 6.3 shows the reconstruction error obtained after 3D point localisation of the test points extracted from the predicted LAT images.

Table 6.3: 3D localized landmarks extracted from the 30 LAT predicted images and the corresponding AP test images in mm

Landmarks	Error, E_x	Error, E_y	Error, E_z	Resultant error, E_i
A	1.20 (0.51)	1.48 (0.60)	0.56 (0.38)	2.11
B	0.54 (0.39)	1.62 (1.31)	0.75 (0.49)	2.03
C	1.05 (0.57)	1.38 (1.07)	0.35 (0.24)	1.94
D	0.90 (0.56)	1.13 (0.65)	0.97 (0.60)	1.98
E	0.51 (0.44)	1.74 (1.26)	0.44 (0.30)	2.03
F	0.81 (0.59)	1.08 (0.64)	0.98 (0.51)	1.86
Global mean (SD)	**0.83 (0.51)**	**1.40 (0.92)**	**0.67 (0.42)**	**1.99**

Table 6.4 shows the reconstruction error obtained after 3D point localisation of the test points extracted from the predicted AP images.

Table 6.4: 3D localized landmarks extracted from the 30 AP predicted images and the corresponding LAT test images in mm

Landmarks	Error, E_x	Error, E_y	Error, E_z	Resultant error, E_i
A	1.09 (0.51)	1.60 (0.41)	0.75 (0.47)	2.16
B	0.51 (0.35)	1.62 (0.28)	0.92 (0.55)	2.02
C	0.98 (0.48)	1.44 (0.45)	0.98 (0.49)	2.09
D	1.06 (0.44)	0.93 (0.47)	0.94 (0.56)	1.85
E	0.81 (0.55)	1.29 (0.62)	1.02 (0.48)	2.01
F	1.14 (0.49)	0.67 (0.34)	0.76 (0.38)	1.62
Global mean (SD)	**0.93 (0.42)**	**1.26 (0.43)**	**0.89 (0.49)**	**1.96**

The average control point reconstruction errors obtained from the eight control points of 0.35 mm, 0.64 mm and, 0.72 mm are acceptable when compared to 0.37, 0.25, 0.42 mm reconstruction errors in the X, Y, and Z directions obtained by Douglas *et al.* (2004) for a 16-control point system.

The average test point reconstruction error obtained from the predicted LAT images and the corresponding AP test images of 0.83 mm, 1.40 mm and, 0.67 mm and those obtained from the predicted AP images and the corresponding LAT test images of 0.93 mm, 1.26 mm and, 0.89 mm in the X, Y, and Z direction are also comparable to the 0.34 mm, 0.29 mm and 0.39 mm nine test points reconstructed using a 16 control point system at 90^0 separation angle in the X, Y, and Z directions, respectively. The difference in error especially in the Y and Z coordinates is attributed to the orthogonal image projection which makes it difficult to estimate the depth of the structure of interest.

The next step was to use the 3D localized landmark points to predict the most likely mesh reconstruction.

6.2 Experiment two: 3D model approximation

This phase involved 3D scapula mesh reconstruction using the localized 3D landmark points in section 6.1 and then validation of the reconstructed meshes.

After calculating the 3D object points using the transformation parameters of the bounding box, the next step was to reconstruct meshes using the reference landmarks of the validated scapula SSM and the localized 3D landmarks. The 60 sets (30 sets from Table 6.3 and 30 sets from Table 6.4) of localized 3D landmark in experiment one of this chapter were used. Each set of the localized 3D landmark points used to constrain the SSM.

To constrain the SSM, a single set of localized landmarks (target points) in section 6.1.2 and the model reference landmarks obtained in section 4.3.2 were loaded into Scalismo. The best transformation of the reference points to the target points was calculated using the *rigid3DLandmarkRegistration* function in Scalismo. This function translates the reference landmarks into the coordinate system of the target landmarks by rigid transformation. After rigid transformation, the iterative closest point (ICP) method (Besl & McKay, 1992) was applied to find the nearest point from the reference points to the target points on the model using *findClosestPoint* function in Scalismo. The resultant points were the best transform of the reference to the target points, resulting in a posterior distribution (posterior model). The posterior model is the most likely distribution that describes the target mesh. This process is referred to as model constraining. After constraining the model, the mean of the predicted posterior model was saved as the most likely mesh reconstruction.

This process was repeated with 3 landmarks and 6 landmarks for all the localized 3D points in section 6.1.2. The 6 landmarks were those described in section 4.3.2, while the 3 landmarks were the points indicated in previous work as the only visible corresponding points on the scapula in both the AP and LAT bi-planar X-ray images of the upper-torso of a cadaver (Mutsvangwa *et al.*, 2017). After obtaining all the meshes, the next step was to evaluate the mesh reconstructions against the ground-truth meshes for each pair of bi-planar images generated in section 4.3.2.

6.2.1 Evaluation of the reconstructed scapula mesh

The metrics including Hausdorff distance, modified Hausdorff distance and average distance for 3D mesh evaluation described in section 3.6.2 were used to evaluate the reconstructed meshes against their ground-truth meshes. Hausdorff distance finds the closest point between two mesh surfaces and returns the maximum distance between them. This measures how far the closest point in one set is located to another point in the other set. Modified Hausdorff distance finds the average of the maximum distances between the closest point in one set to another point in the other set. The average distance gives the mean of the shortest distance between two mesh surface points. For all these evaluation metrics, a better

reconstruction is indicated by a smaller the distance between the ground-truth and the reconstructed mesh.

The metrics are also in-built parts of the mesh evaluation functions in Scalismo. The evaluation process was as follows. The ground-truth mesh and reconstructed mesh were loaded in Scalismo and the mesh evaluation metrics applied to the two mesh surfaces for comparison. This process was repeated for all the reconstructed meshes.

6.2.2 Results

Figure 6.1 shows the ground-truth mesh sample referred to as mesh 1, mesh 2 reconstructed using 6 corresponding landmarks, mesh 3 reconstructed using 3 corresponding landmark points and mesh 2 aligned to mesh 1, mesh 3 aligned to mesh 1, and lastly the meshes 2 and 3 aligned to mesh 1.

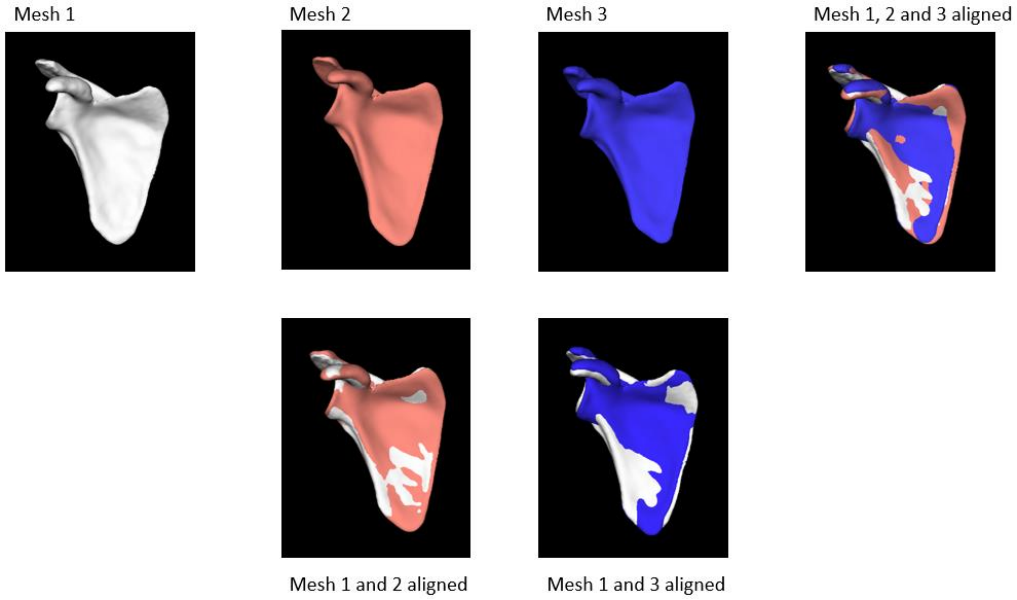

Figure 6.1: Ground-truth mesh sample (mesh 1), mesh 2 reconstructed from 6 corresponding landmarks, mesh 3 reconstructed from 3 corresponding landmarks, and meshes 2 and 3 aligned to mesh 1.

The results shown in the Table 6.5 are average results of the 30 reconstructed meshes for the LAT and AP predicted images with 3 and 6 landmarks, respectively.

Table 6.5: Surface-to-surface distance errors (mm) for reconstructed meshes

Reconstructed 30 mesh instances	Number of landmarks	Hausdorff distance	Average Hausdorff distance	Average distance
Predicted AP	**3**	17.29	3.51	3.23
	6	8.87	1.94	1.79
Predicted LAT	**3**	14.78	3.05	3.21
	6	8.30	1.82	1.64

The results shown in Table 6.6 are the average results of all the 60 reconstructed meshes comprising of the 30 reconstructed meshes from the predicted LAT images of AP to LAT model and 30 reconstructed meshes from the predicted AP images of LAT to AP model obtained using 3 and 6 landmarks, respectively.

Table 6.6: Surface-to-surface distance errors (mm) for reconstructed meshes

Reconstructed mesh instances	Hausdorff distance	Average Hausdorff distance	Average distance
3 landmarks	16.04	3.28	3.22
6 landmarks	8.59	1.88	1.72

The surface-to-surface distances shown in Table 6.5 and Table 6.6 indicate that the reconstructed meshes were different from the ground-truth. The reconstructed scapula meshes with 3 landmarks had a higher average surface-to-surface distance of 3.22 mm compared to the 1.72 mm of the meshes reconstructed with 6 landmarks. In comparison, Mutsvangwa *et al.* (2017) obtained average surface-to-surface errors of 3.20 mm and 2.46 mm for 3 and 16 landmarks, respectively, for a reconstructed mesh within the training dataset of the SSM (Mutsvangwa *et al.*, 2017).

6.3 Conclusion

The results from both studies show a positive relationship between the number of corresponding landmarks used for 3D reconstruction and the average surface-to-surface distance between the reference and the target. However, it is difficult to locate corresponding landmarks even for this case without interference from super-positioned structures on the scapula. Thus, the need to investigate the use of matching landmarks on the contour of the scapula to obtain more accurate patient-specific 3D reconstruction of bi-planar X-rays using the landmark-constrained model fitting method.

7 Three-dimensional reconstruction of bi-planar X-ray images using matching points from the scapula contour

This chapter describes steps carried out to reconstruct the scapula meshes using matching points selected from the scapula contour in bi-planar images. Figure 7.1 shows the steps taken to reconstruct the bi-planar X-ray images using matching points from the scapula contour.

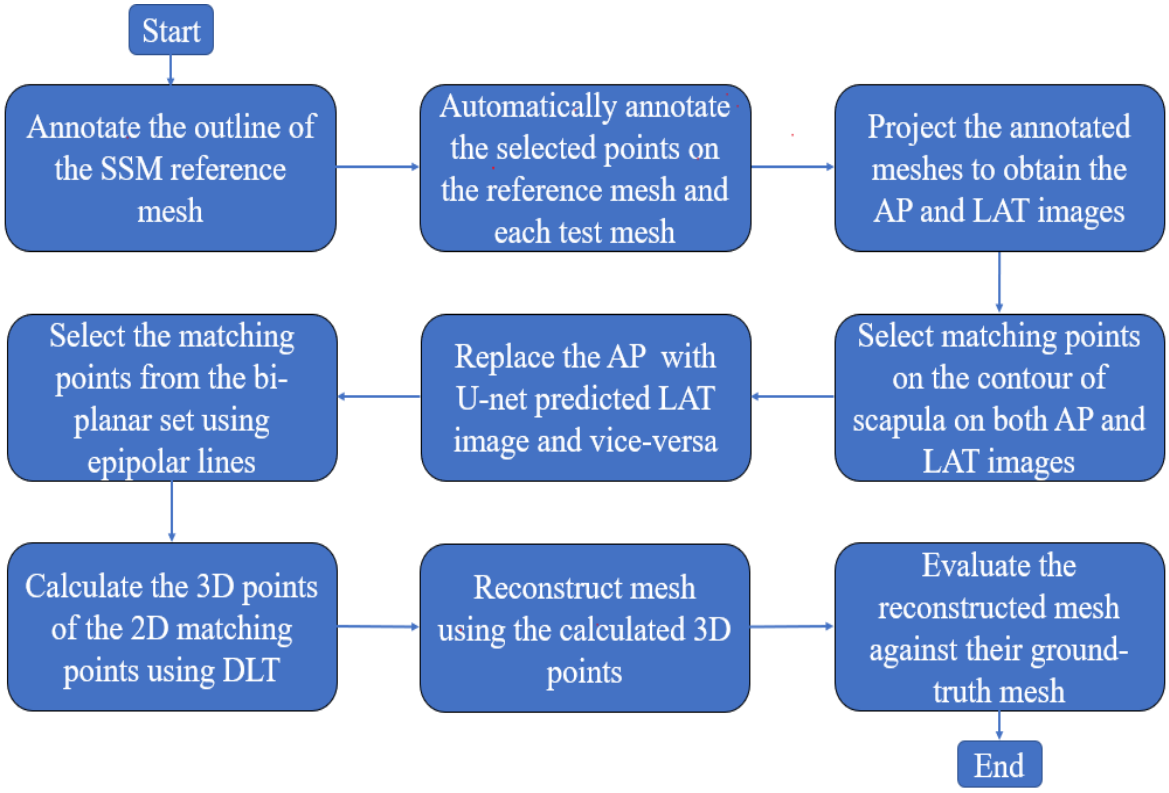

Figure 7.1: Steps taken for 3D reconstruction of the bi-planar X-ray images using matching points from the scapula contour.

7.1 Selection of 2D points from the scapula contour

The scapula statistical shape model (SSM) reference mesh was manually annotated with landmarks along the entire scapula outline. These landmarks included the all the 16 reproducible scapula landmarks in section 4.3.2 (Borotikar *et al.*, 2015; Ohl *et al.*, 2010). The mesh was converted into a volumetric object and projected with its landmarks using the digitally reconstructed radiograph (DRR) renderer (Reyneke, 2019). Figure 7.2 shows the SSM reference mesh manually annotated with points on its outline and the corresponding AP and LAT projections.

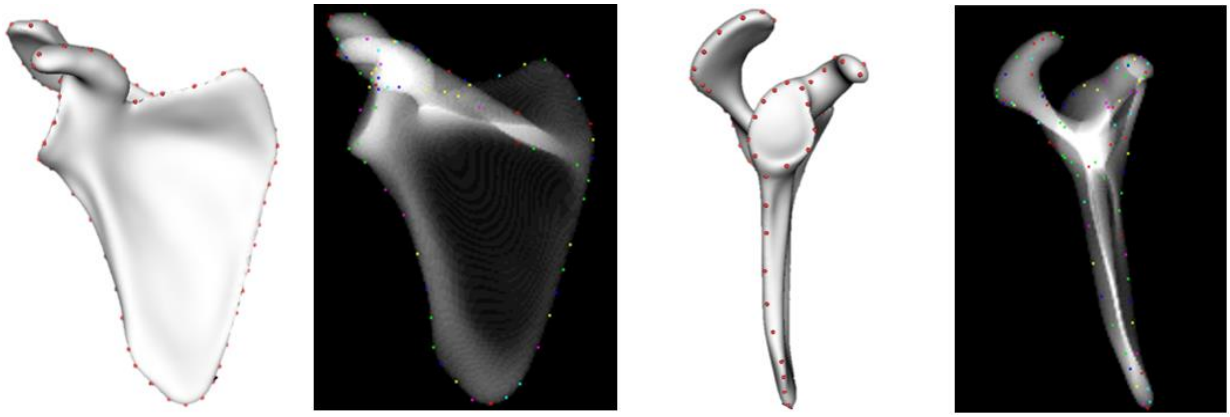

Figure 7.2: SSM reference mesh manually annotated with points on its outline and the corresponding AP and LAT projections.

The projected bi-planar images shown in Figure 7.2 were used to aid the selection of 8 points that are located on the contour of both anterior-posterior (AP) and lateral (LAT) images. Eight points were selected to obtain more points than the 6 corresponding landmarks used in section 6 for three-dimensional scapula reconstruction. After obtaining matching point on the contour of both the AP and LAT images, five meshes were randomly selected from the ground-truth meshes used in section 6.2.1. The meshes were automatically annotated with the same points selected on the reference mesh using the *findClosestPoint* function in Scalismo. This was followed by projection of the annotated meshes to obtain the AP and LAT images similar to those shown in Figure 7.2. The projected bi-planar images were used to aid the selection of 8 matching points that are located on the contour of both AP and LAT images.

However, to avoid mismatch of landmarks each mesh was projected with only the 8 matching points on the contour of both images. Epipolar lines were automatically generated from the selected points on the contour of the AP image to locate the corresponding points in the LAT image. The epipolar line that passes through a point in one image given the selected matching point in the other image is determined using a fundamental matrix for an uncalibrated scene. Given two images with corresponding points x and x' these points are found on epipolar lines (l and l') and they are joined by an epipolar plane to their corresponding 3D point in space. This relation between matching points in stereo images is defined using epipolar geometry. Epipolar geometry helped reduce the search space for the corresponding landmarks to only one direction. The fundamental matrix can be calculated given at least eight matching points from stereo images using the 8-point algorithm shown in equation (3.17).

However, to calculate the fundamental matrix using the 8-point algorithm, at least 8 points are required to obtain a good estimate of the fundamental matrix. A MATLAB graphical user interface (GUI) was used to manually select 8 matching points from the stereo images. To estimate the fundamental matrix the least squares solution is obtained using singular value decomposition (SVD) of matrix W in the

equation (3.17) which relates the matching points. After the estimation of the fundamental matrix, any point selected on the left image would generate an epipolar line on the left image using equation (7.1).

$$l = F\,x' \qquad l' = F^T x \tag{7.1}$$

where l is the epipolar line associated with point x' and l' is the epipolar line associated with point x, F is the fundamental matrix and F^T is the transpose of the fundamental matrix.

The generated epipolar line is the line joining the clicked point in the left image and its epipole in the right image. Figure 7.3 shows epipolar lines in the AP image going through the LAT image.

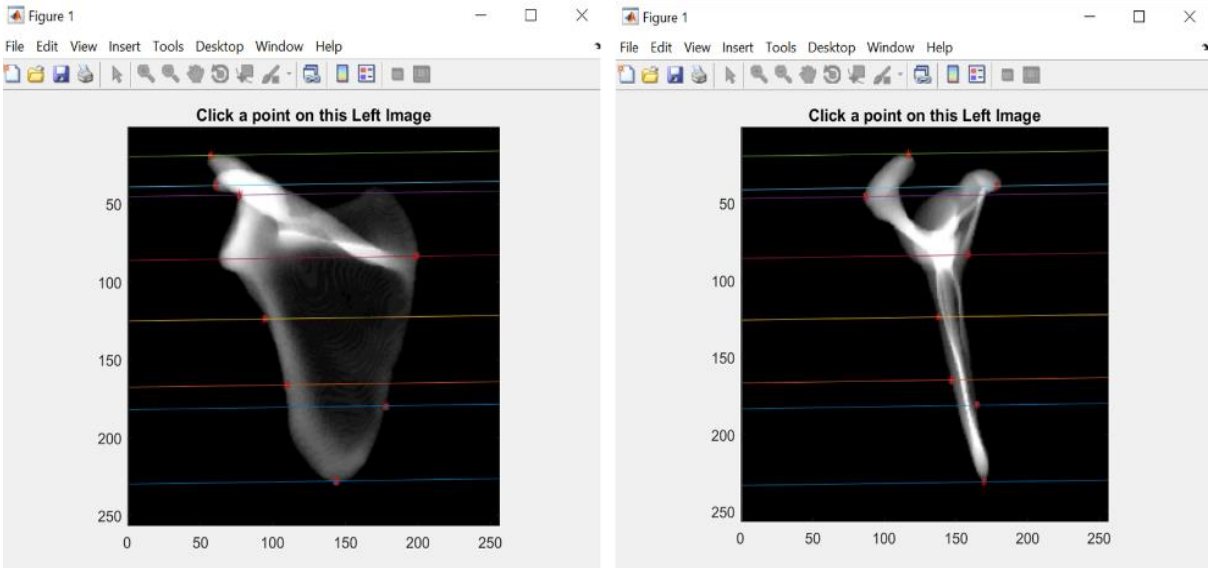

Figure 7.3: Epipolar lines in the AP image going through the LAT image.

This process was repeated replacing the projected LAT (right) image in Figure 7.3 with the image predicted by U-net model. A Sobel filter was applied to the predicted image to extract the scapula contour from the image. This was followed by the location of the points that are found on the contour using epipolar lines. Figure 7.4 shows the projected AP image with selected points found on the outline of the scapula used to aid the location of the matching points on the LAT image using epipolar lines and vice versa.

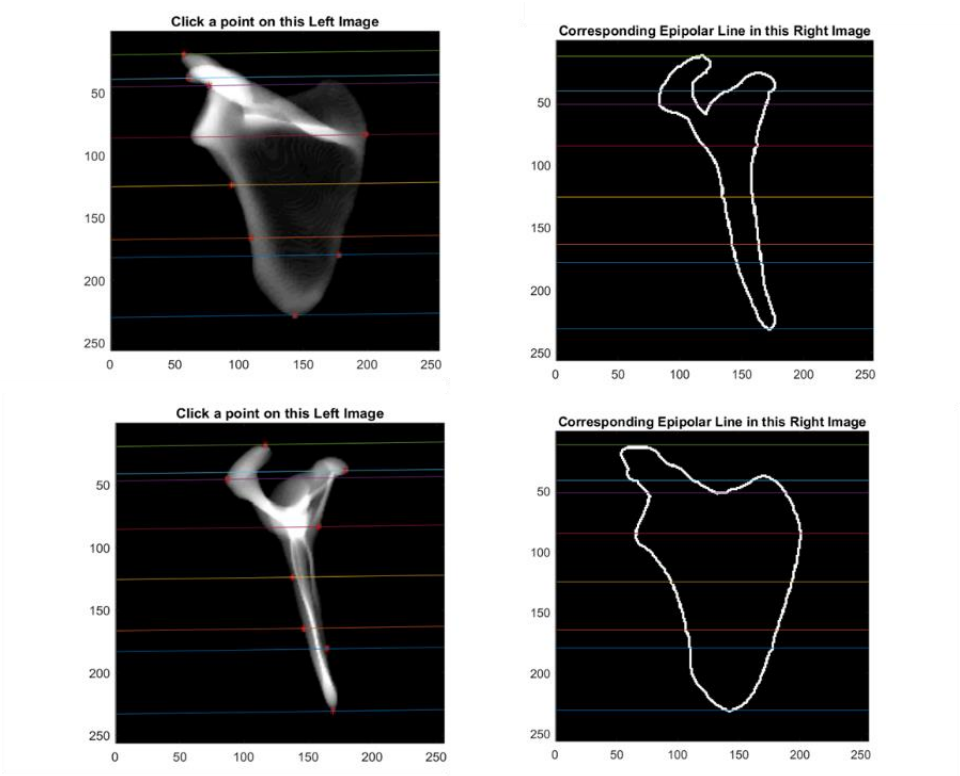

Figure 7.4: Location of corresponding points that are found on the contour in both AP and LAT images using epipolar lines.

The selected points that were located on the contour of the projected images in Figure 7.4 were used to locate the corresponding points in the predicted images to obtain the matching points on the contour. The points on the contour that coincided with the epipolar lines were selected as the corresponding points between the image pair. The extracted 2D points were saved as JavaScript Object Notation (.json) files. This process was repeated for 10 randomly selected meshes and their predicted bi-planar images. The 10 meshes were a subset of the 30 randomly selected dataset that was used for landmark error evaluation in Table 5.2 and Table 5.4 and also landmark localisation and mesh reconstruction in chapter 6.

7.2 Scapula mesh reconstruction

The 2D points selected from the contour in section 7.1 were transformed into 3D points using the transformation parameters obtained in section 6.1.1. The reconstructed 3D points were used to reconstruct scapula meshes using functions in Scalismo. The first step to obtaining the reconstructed mesh was to perform a rigid transformation of the SSM reference mesh landmarks that correspond to the target mesh landmarks using the *rigid3DLandmarkRegistration* function. This process transformed the reference landmarks into the same coordinate system as the 3D localised target points. This was followed by applying the *findClosestPoint* function to find correspondence between the transformed reference landmarks and the target points. This process of mapping the reference landmarks to the target

landmarks resulted in the prediction of a posterior model. The mean of the predicted posterior model was saved as the reconstructed mesh.

Finally, the reconstructed scapula meshes from each set of reconstructed 3D sets of points from the bi-planar image contours were compared to their corresponding ground-truth scapula meshes. The mesh metrics in section 3.6.2 were used for evaluation.

7.2.1 Results

Figure 7.5 shows the ground-truth mesh sample (mesh 1), mesh reconstructed using 6 landmarks (mesh 2), mesh reconstructed using 8 contour matching points (mesh 3), mesh 2 aligned to mesh 1, mesh 3 aligned to mesh 1, and lastly meshes 2 and 3 aligned to mesh 1.

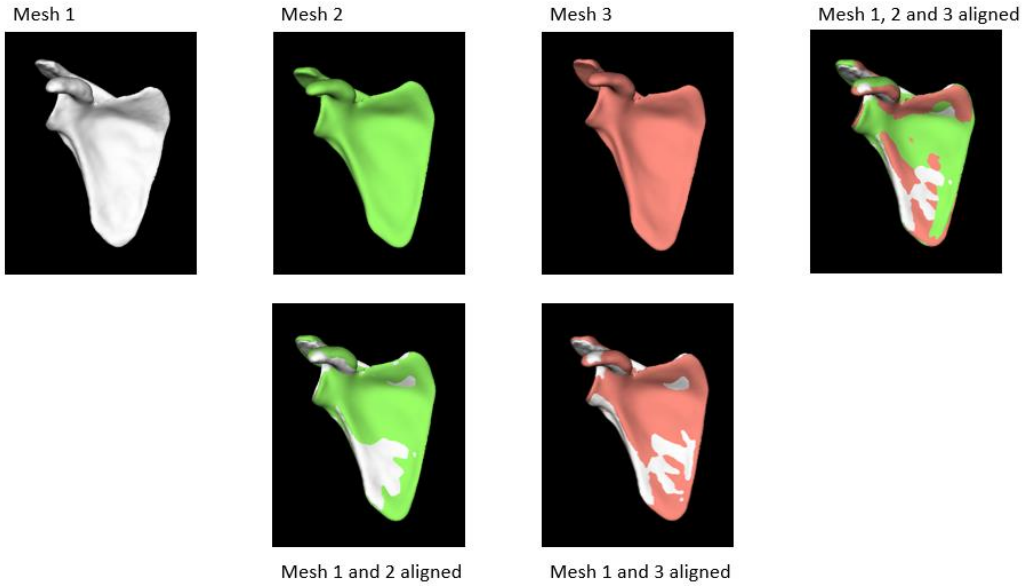

Figure 7.5: Ground-truth mesh sample (mesh 1), mesh 2 reconstructed from 6 corresponding landmarks, mesh 3 reconstructed from 8 matching contour points, and meshes 2 and 3 aligned to mesh 1.

Table 7.1 shows the average surface-to-surface distances for 10 reconstructed meshes with 8 matching contour points and the 6 corresponding scapula landmarks. Detailed results for each mesh are found in Table A.1 and Table A.2 in Appendix.

Table 7.1: Surface-to-surface distance errors (mm)

Reconstructed mesh	Hausdorff distance	Average Hausdorff distance	Average distance
8 matching contour points	6.23	1.42	1.40
6 corresponding scapula landmarks	8.54	1.96	1.91

Table 7.1 shows 8 matching points from the contour give an average surface-to-surface distance of 1.40 mm which is lower compared to the average surface-to-surface distance of 1.91 mm of the meshes reconstructed from 6 corresponding landmarks. The results support the research by Mutsvangwa *et al.* (2017) that an increase in the number of corresponding points from the bi-planar images results in a better reconstruction.

7.3 Conclusion

The results in this section show that an increase in the number of selected landmarks gives a better 3D reconstruction. However, location of corresponding landmarks for a complex bone is difficult thus the use of matching points on the contour of the scapula in the bi-planar images to aid reconstruction of more accurate meshes. In this study, the use of 8 contour matching points for the mesh reconstruction gave better results than the use of 6 corresponding points. However, one needed to use epipolar lines to reduce the search space of locating the contour matching points. Future work should explore the use of contour matching methods to automatically obtain matching points for a more accurate patient-specific 3D reconstruction of bi-planar X-rays using the landmark-constrained model fitting method.

8 Conclusion

This study aimed to train and validate a convolutional neural network (CNN)-based deep learning algorithm could detect the contour of a scapula in synthetic two-dimensional (2D) bi-planar X-ray images.

8.1 Summary of the findings

8.1.1 U-net model training

The first model (AP to LAT) which was trained to predict the lateral (LAT) image view given the anterior-posterior (AP) image view achieved an average Dice coefficient value of 0.926 for 300 predicted LAT images. The second model (LAT to AP) achieved an average Dice coefficient of 0.964 for 300 predicted AP images given the LAT images. These results were comparable to Dice coefficient values of 0.831 and 0.891 obtained by Shvets *et al.* (2018) and Livne *et al.* (2019), respectively on image predictions tasks using a U-net model. Thus, according to these results the U-net model can learn the mapping between the binary bi-planar images of the scapula. However, these models would require fine tuning and further training with manually segmented bi-planar scapula X-ray images to generalise to real data.

8.1.2 Landmark-constrained model fitting using known corresponding points

To perform landmark-constrained model fitting, 3D landmark localisation was initially performed to transform the predicted 2D corresponding landmarks to 3D points. Landmark localisation errors for the control points (0.5 mm, 0.64 mm and, 0.72 mm) were comparable to the control points (0.37 mm, 0.25 mm, 0.42 mm) in a previous study by Douglas *et al.* (2004). The test point reconstruction error for the AP predicted images (0.83 mm, 1.40 mm and, 0.67 mm) and LAT predicted images (0.93 mm, 1.26 mm and, 0.89 mm) were also comparable the test point reconstruction error (0.34mm, 0.29mm and 0.39mm) by Douglas *et al.* (2004). This was followed by reconstruction of the scapula meshes using the obtained localised 3D landmarks.

The average surface-to-surface distance obtained for the 60 landmark-constrained mesh reconstructions using the localised corresponding points, were 3.22 mm and 1.72 mm for 3 and 6 landmarks, respectively. These results are similar to the average surface-to-surface errors of 3.20 mm and 2.46 mm for 3 and 16 landmarks, respectively, which were obtained by Mutsvangwa *et al.* (2017) in a previous study. However, only 3 corresponding reproducible landmarks could be identified from the scapula bi-planar X-ray images of a cadaver, leading to a higher reconstruction error (Mutsvangwa *et al.*, 2017). Thus, the need to identify more corresponding landmarks for 3D mesh reconstruction motivated the last

experiment of this study which was reconstruction of the scapula mesh using matching points that lie on the contours in both images.

8.1.3 Landmark-constrained model fitting using matching points on the contour of the bi-planar scapula images

The average surface-to-surface distance obtained from the reconstruction of the scapula mesh with 8 matching contour points (1.40 mm) was better than the results obtained when 6 corresponding scapula landmarks (1.91mm) were used for reconstruction. These results further confirm that increase in the number of corresponding landmarks in bi-planar images for 3D reconstruction results into an improved 3D reconstruction (Mutsvangwa *et al.*, 2017). However, for this study the 8 matching points were only located with prior knowledge of their location in one of the images, followed by use of epipolar lines to locate the matching point in the corresponding image. This approach can be expected to be a more challenging process in a clinical scenario where a clinician must contend with interference from surrounding structures. Thus, contour matching methods which have been used in some previous studies (Xiao *et al.*, 2016; Zhang *et al.*, 2013) maybe worth exploring to automatically locate matching points from the X-ray bi-planar images in order to obtain more corresponding points to improve the 3D reconstruction.

8.2 Limitations and recommendations for future work

During the data generation step of this research study in section 4.3, the scapula SSM that was used generated some mesh samples with faulty triangles. This resulted in the need to correct the defective meshes to maintain the number of training samples. The process of correcting the faulty triangles on the mesh surfaces was tedious and unique for each surface - a time-consuming process. Re-defining of the model reference mesh is recommended as the better alternative in order to avoid the laborious process of correcting each mesh.

In addition, phantoms may be included in future work. A phantom can be used as a real-world 3D reference model to enable acquisition of multiple 2D X-ray images within a calibrated 3D reference space. Phantoms reduce the need for repeated exposure of human subjects to ionising radiation (Claus, 2006; Groenewald & Groenewald, 2016, 2019; Ng & Yeong, 2014).

In section 5, the U-net model was used in the prediction of the LAT given the AP images and vice versa, giving results that were comparable to the results in literature. However, it is recommended that future work investigate the use of generative adversarial networks (GANs) as they have shown to give competitive results for image synthesis in different planar views (Angsarawanee & Kijsirikul, 2019; Kim *et al.*, 2017).

In section 5, 6, and 7 binary images were used to develop this proof of concept because it was easier to generate and use such data for deep learning purposes. However, the use of these binary images to train the U-net models made it difficult for the trained models to generalise to real data. In addition, these binary images resulted in featureless objects during the selection of landmarks on the contour. Hence the use of prior information from the corresponding full intensity grayscale images in section 7 to aid in the identification of matching points on the contour of both the AP and LAT images. This made the process of matching points selection specific for each mesh projection. Considering these limitations of using the binary images future work should consider additional model analysis using real data.

To locate the points that lie on the contour in the full intensity grayscale image, the user had to project the mesh sample with many landmarks to identify the exact landmarks that lie on the contour in both image views; a point that lies on the contour in one view does not necessarily lie on the contour in the corresponding view. This is mainly due to the large angle (90^0) difference between the bi-planar images. Although a large separation angle between the bi-planar images results in a better accuracy during 3D point localisation, it makes the location of corresponding landmarks hard. In addition, the presence of the variation in the scapula shape per sample from the SSM meant that most landmarks would fall a few pixels from the contour and would not be considered as points on the contour.

The results of this study suggest that the use of contour matching for 3D reconstruction could improve the reconstruction results without the need for corresponding reproducible landmarks. Contour matching was also proposed in previous studies to aid accurate definition of the location and orientation of the vertebrae for 3D reconstruction of the spine from bi-planar radiographs (Xiao *et al.*, 2016; Zhang *et al.*, 2013).

8.3 Overall conclusion and contribution of the project

The U-net model can learn the mapping of the AP images to LAT images and vice versa. However, the models did not learn the embedded landmarks with high precision compared to the U-net models trained to perform the LAT to LAT prediction and AP to AP prediction. Furthermore, the trained models could not generalise to real data.

This project produced a proof of concept that U-net deep learning algorithm can learn the mapping between bi-planar images. Further training of the U-net models with manually segmented images could be one way to improve the results for a clinical scenario where segmentation of both bi-planar images is required. The ability to implement this algorithm in a clinical setting would reduce the time required to obtain the corresponding annotated view of a second bi-planar image given an annotated view of the first image. In addition, if one image view can be used to obtain the corresponding view, the process of

3D from 2D reconstruction using bi-planar X-ray images would be done using a single image thus reducing ionising radiation to patients.

Another contribution of this research project is that it shows how to obtain and use matching points from the contour of the scapula in bi-planar images to obtain an improved 3D reconstructed scapula without necessarily using known corresponding reproducible landmarks.

References

Abdel-Aziz, Y. I., Karara, H. M., & Hauck, M. (2015). Direct Linear Transformation from Comparator Coordinates into Object Space Coordinates in Close-Range Photogrammetry. *Photogrammetric Engineering & Remote Sensing, 81*(2), 103-107. doi:10.14358/PERS.81.2.103

Acharjya, P. P., Das, R., & Ghoshal, D. (2012). Study and comparison of different edge detectors for image segmentation. *Global Journal of Computer Science and Technology, 12*(13), 29-32.

Achermann, B., & Bunke, H. (2000). Classifying range images of human faces with Hausdorff distance. Paper presented at the *Proceedings 15th International Conference on Pattern Recognition. ICPR-2000*, Barcelona, Spain, Vol. 2, 809-813

Adams, L. P. (1981). X-ray stereo photogrammetry locating the precise, three-dimensional position of image points. *Med Biol Eng Comput, 19*(5), 569-578. doi:10.1007/BF02442771

Amirlak, B., Zakhary, B., Weichman, K., Ahluwalia, H., Forse, A. R., & Gaines, R. D. (2009). Novel use of Lodox Statscan in a level one trauma center. *Ulus Travma Acil Cerrahi Derg, 15*(6), 521-528.

Andersson, J., Ahlström, H., & Kullberg, J. (2019). Separation of water and fat signal in whole-body gradient echo scans using convolutional neural networks. *Magnetic Resonance in Medicine, 82*(3), 1177-1186. doi:10.1002/mrm.27786

Andrews, H., & Patterson, C. (1976). Singular Value Decompositions and digital image processing. *IEEE Transactions on Acoustics, Speech, and Signal Processing, 24*(1), 26-53. doi:10.1109/TASSP.1976.1162766

Angsarawanee, T., & Kijsirikul, B. (2019). Generating images with desired properties using the DiscoGAN model enhanced with repeated property construction. Paper presented at the *Proceedings of the International Conference on Advanced Information Science and System*, 1-9.

Auroux, D., Cohen, L. D., & Masmoudi, M. (2011). Contour detection and completion for inpainting and segmentation based on topological gradient and fast marching algorithms. *International Journal of Biomedical Imaging, 2011*, 592924. doi:10.1155/2011/592924

Berg, M., Cheong, O., Kreveld, M., & Overmars, M. (2008). Computational Geometry: Algorithms and Applications. (Third ed., pp. 386). Berlin Heidelberg: Springer-Verlag.

Besl, P. J., & McKay, N. D. (1992). A method for registration of 3-D shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 14*(2), 239-256. doi:10.1109/34.121791

Borotikar, B., Ghorbel, E., Lempereur, M., Mutsvangwa, T., & Burdin, V. (2015). Evaluation of an anatomically augmented Statistical Shape Model of the scapula: Clinical validation and reliability of landmark selection. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging Visualization*. doi:10.13140/RG.2.1.4409.2242

Brenner, D. J., & Hricak, H. (2010). Radiation exposure from medical imaging: time to regulate? . *Jama, 304*(2), 2. doi:doi:10.1001/jama.2010.973

Burgstaller, B., & Pillichshammer, F. (2009). The average distance between two points. *Bulletin of the Australian Mathematical Society, 80*(03), 353 - 359. doi:10.1017/S0004972709000707

Cernazanu-Glavan, C., & Holban, S. (2013). Segmentation of bone structure in X-ray images using convolutional neural network. *Advances in Electrical and Computer Engineering, 13*(1), 87-94.

Cernazanu-Glavan, C., & Stefan, H. (2013). Segmentation of Bone Structure in X-ray Images using Convolutional Neural Network. *Advances in Electrical and Computer Engineering, 13*(1), 87-94. doi:10.4316/AECE.2013.01015

Chabrier, S., Laurent, H., Rosenberger, C., & Emile, B. (2008). Comparative Study of Contour Detection Evaluation Criteria Based on Dissimilarity Measures. *EURASIP Journal on Image and Video Processing, 2008*(693053), 1-13. doi:10.1155/2008/693053

Chan, S.-L. S., Jeffree, R., L. , Fay, M., Crozier, S., Yang, Z., Gal, Y., & Thomas, P. (2013). Automated Classification of Bone and Air Volumes for Hybrid PET-MRI Brain Imaging. Paper presented at the *2013 International Conference on Digital Image Computing: Techniques and Applications (DICTA)*, Hobart, TAS, Australia, 1-8.

Chang, A. E., Matory, Y. L., Dwyer, A. J., Hill, S. C., Girton, M. E., Steinberg, S. M., . . . Doppman, J. L. (1987). Magnetic resonance imaging versus computed tomography in the evaluation of soft tissue tumors of the extremities. *Ann Surg, 205*(4), 340-348. doi:10.1097/00000658-198704000-00002

Chimhundu, C., Sivarasu, S., Steiner, S., Smit, J., & Douglas, T. S. (2016). Femoral neck anteversion measurement using linear slot scanning radiography. *Medical Engineering & Physics, 38*(2), 187-191. doi:10.1016/j.medengphy.2015.11.017

Chimhundu, C., Smit, J., Sivarasu, S., & Douglas, T. S. (2014). Interlandmark Measurements From Lodox Statscan Images1. *Journal of Medical Devices, 8*(3), 030908. doi:10.1115/1.4027102

Claus, B. E. (2006). Geometry calibration phantom design for 3D imaging. Paper presented at the *Medical Imaging 2006: Physics of Medical Imaging*, Vol. 6142, 61422E.

Cootes, T. F., & Taylor, C. J. (2001). Statistical models of appearance for medical image analysis and computer vision Paper presented at the *Medical Imaging: Image Processing*, Vol. 4322, 236-248.

Cootes, T. F., Taylor, C. J., Cooper, D. H., & Graham, J. (1992). Training models of shape from sets of examples. In B. R. Hogg D. (Ed.), *BMVC92* (pp. 9-18). Springer, London: Springer.

Cootes, T. F., Taylor, C. J., Cooper, D. H., & Graham, J. (1995). Active Shape Models-Their Training and Application. *Computer Vision and Image Understanding, 61*(1), 38-59. doi:10.1006/cviu.1995.1004

Crum, W. R., Hartkens, T., & Hill, D. (2004). Non-rigid image registration: theory and practice. *The British journal of radiology, 77*(suppl_2), S140-S153. doi:10.1259/bjr/25329214

Dice, L. R. (1945). Measures of the Amount of Ecologic Association Between Species. *Ecology, 26*(3), 297–302. doi:doi:10.2307/1932409

Diop, E. H. S., & Burdin, V. (2013). Bi-planar image segmentation based on variational geometrical active contours with shape priors. *Medical Image Analysis, 17*(2), 165-181. doi:10.1016/j.media.2012.09.006

Douglas, T. S., Vaughan, C. L., & Wynne, S. M. (2004). Three-dimensional point localisation in low-dose X-ray images using stereo-photogrammetry. *Medical and Biological Engineering and Computing, 42*(1), 37-43. doi:10.1007/BF02351009

Dryden, I. L., & Mardia, K. V. (1998). *Statistical Shape Analysis*. Chichester: Wiley.

Dubuisson, M., & Jain, A. K. (1994). A modified Hausdorff distance for object matching. Paper presented at the *Proceedings of 12th International Conference on Pattern Recognition*, Jerusalem, Israel, Israel, Vol. 1, 566-568.

Evangelopoulos, D. S., Deyle, S., Zimmermann, H., & Exadaktylos, A. K. (2009). Personal experience with whole-body, low-dosage, digital X-ray scanning (LODOX-Statscan) in trauma. *Scandinavian Journal of Trauma, Resuscitation and Emergency Medicine, 17*, 41-41. doi:10.1186/1757-7241-17-41

Fouefack, J.-R. (2018). *Geometric morphometrics for 3D dense surface correspondence: population comparisons of shoulder bone morphology.* (MSc (Biomedical Engineering)), University of Cape Town. Retrieved from http://hdl.handle.net/11427/30024

Franco, E. L., & Turgeon, G.-A. (2010). Radiodiagnostic imaging in pregnancy and the risk of childhood malignancy: Raising the bar. *PLoS medicine, 7*(9), e1000338. doi:10.1371/journal.pmed.1000338.

Gerig, T., Shahim, K., Reyes, M., Vetter, T., & Lüthi, M. (2014). Spatially varying registration using gaussian processes. Paper presented at the *International Conference on Medical Image Computing and Computer-Assisted Intervention - MICCAI 2014*, Cham, Vol. 8674, 413-420.

Góra, P., & Boyarsky, A. (1988). Why computers like lebesgue measure. *Computers & Mathematics with Applications, 16*(4), 321-329. doi:10.1016/0898-1221(88)90148-4

Groenewald, A., & Groenewald, W. A. (2016). Development of a universal medical X-ray imaging phantom prototype. *Journal of applied clinical medical physics, 17*(6), 356-365. doi:10.1120/jacmp.v17i6.6356

Groenewald, A., & Groenewald, W. A. (2019). A universal phantom suitable for quality assurance on X-ray imaging modalities. *Acta Radiologica, 60*(11), 1523-1531. doi:10.1177/0284185119831685

Hartley, R., & Zisserman, A. (2004). *Epipolar geometry and the fundamental matrix* (2 ed.). Cambridge: Cambridge University Press.

Hartley, R. I. (1997). In defense of the eight-point algorithm. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 19*(6), 580-593. doi:10.1109/34.601246

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. Paper presented at the *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, Vol. 7, 770-778.

Helmke, U., H{\"u}per, K., Lee, P. Y., Moore, J. B., & (2007). Essential Matrix Estimation Using Gauss-Newton Iterations on a Manifold. *International Journal of Computer Vision, 74*(2), 117-136. doi:10.1007/s11263-006-0005-0

Hesamian, M. H., Jia, W., He, X., & Kennedy, P. (2019). Deep learning techniques for medical image segmentation: Achievements and challenges. *Journal of digital imaging, 32*(4), 582-596. doi:10.1007/s10278-019-00227-x

Hinton, G. E. (2007). Learning multiple layers of representation. *Trends Cogn Sci, 11*(10), 428-434. doi:10.1016/j.tics.2007.09.004

Isin, A., Direkoglu, C., & Sah, M. (2016). Review of MRI-based brain tumor image segmentation using deep learning methods. *Procedia Computer Science, 102*(C), 317-324. doi:10.1016/j.procs.2016.09.407

Jolliffe, I. T. (1993). Principal component analysis: a beginner's guide—II. Pitfalls, myths and extensions. *Weather, 48*(8), 246-253. doi:10.1002/j.1477-8696.1993.tb05899.x

Kass, M., Witkin, A., & Terxopoulos, D. (1988). Snakes - Active Contour Models. *lnternational Journal of Computer Vision, 1*(4), 321-331. doi:10.1007/BF001335

Kaur, H., & Singh, I. (2016). The Study Edge Detection of Medical Images using transformation techniques and filtration methods. *International Journal of Computer Applications 146*(12), 39-42. doi:10.5120/ijca2016910960

Kim, T., Cha, M., Kim, H., Lee, J. K., & Kim, J. (2017). Learning to discover cross-domain relations with generative adversarial networks. Paper presented at the *Proceedings of the 34th International Conference on Machine Learning-Volume 70*, Vol. 70, 1857-1865.

Kim, Y. Y., Shin, H. J., Kim, M.-J., & Lee, M.-J. (2016). Comparison of effective radiation doses from X-ray, CT, and PET/CT in pediatric patients with neuroblastoma using a dose monitoring program. *Diagnostic and Interventional Radiology, 22*(4), 390. doi:10.5152/dir.2015.15221.

Koo, T. K., & Li, M. Y. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine, 15*(2), 155-163. doi:10.1016/j.jcm.2016.02.012

Lam, K. C., & Lui, L. M. (2014). Landmark-and intensity-based registration with large deformations via quasi-conformal maps. *SIAM Journal on Imaging Sciences, 7*(4), 2364-2392. doi:10.1137/130943406

Laporte, S., Skalli, W., De Guise, J. A., Lavaste, F., & Mitton, D. (2003). A biplanar reconstruction method based on 2D and 3D contours: application to the distal femur. *Comput Methods Biomech Biomed Engin, 6*(1), 1-6. doi:10.1080/1025584031000065956

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521*(7553), 436-444. doi:10.1038/nature14539

Leondes, C. T. (2003). Algorithms for the Recovery of the 3-D Shape of Anatomical Structures from Single X-Ray Images. In C. T. Leondes (Ed.), *Computational Methods in Biophysics, Biomaterials, Biotechnology and Medical Systems: Algorithm Development, Mathematical Analysis, and Diagnostics* (Vol. 1, pp. 93-126). Boston, MA: Springer US.

Li, C., Xu, C., Gui, C., & Fox, M. D. (2005). Level set evolution without re-initialization: a new variational formulation. Paper presented at the *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, Vol. 1, 430-436.

Liao, R., Miao, S., Tournemire, P. d., Grbic, S., Kamen, A., Mansi, T., & Comaniciu, D. (2016). An Artificial Agent for Robust Image Registration. *ArXiv, abs/1611.10336*.

Linet, M. S., Kim, K. P., & Rajaraman, P. (2009). Children's exposure to diagnostic medical radiation and cancer risk: epidemiologic and dosimetric considerations. *Pediatric Radiology, 39* (Suppl 1), S4-26. doi:10.1007/s00247-008-1026-3

Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., . . . Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. *Medical Image Analysis, 42*(2017), 60-88. doi:10.1016/j.media.2017.07.005

Livne, M., Rieger, J., Aydin, O. U., Taha, A. A., Akay, E. M., Kossen, T., . . . Frey, D. (2019). A U-net deep learning framework for high performance vessel segmentation in patients with cerebrovascular disease. *Frontiers in Neuroscience, 13*, 97. doi:doi.org/10.3389/fnins.2019.00097

Long, J., Shelhamer, E., & Darrell, T. (2015). Fully Convolutional Networks for Semantic Segmentation. Paper presented at the *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, USA, Vol. 3431-3440.

Lüthi, M., Jud, C., & Vetter, T. (2013). A unified approach to shape model fitting and non-rigid registration. Paper presented at the *International Workshop on Machine Learning in Medical Imaging (MLMI 2013)*, Cham, Vol. 8184, 66-73.

Maier, A., Syben, C., Lasser, T., & Riess, C. (2019). A gentle introduction to deep learning in medical image processing. *Zeitschrift für Medizinische Physik, 29*(2), 86-101. doi:10.1016/j.zemedi.2018.12.003

Mariani, G., Kasznia-Brown, J., Paez, D., Mikhail, M. N., H. Salama, D., Bhatla, N., . . . Kashyap, R. (2017). Improving women's health in low-income and middle-income countries. Part II: the needs of diagnostic imaging. *Nuclear Medicine Communications, 38*(12), 1024-1028. doi:10.1097/MNM.0000000000000752

Markelj, P., Tomaževič, D., Likar, B., & Pernuš, F. (2012). A review of 3D/2D registration methods for image-guided interventions. *Medical Image Analysis, 16*(3), 642-661. doi:10.1016/j.media.2010.03.005

Mayya, M., Poltaretskyi, S., Hamitouche, C., & Chaoui, J. (2013). Scapula statistical shape model construction based on watershed segmentation and elastic registration. Paper presented at the *2013 IEEE 10th International Symposium on Biomedical Imaging*, San Francisco, CA, USA, 101-104.

Melhem, E., Assi, A., El Rachkidi, R., & Ghanem, I. (2016). EOS((R)) biplanar X-ray imaging: concept, developments, benefits, and limitations. *J Child Orthop, 10*(1), 1-14. doi:10.1007/s11832-016-0713-0

Mettler, F. A., Jr., Huda, W., Yoshizumi, T. T., & Mahesh, M. (2008). Effective doses in radiology and diagnostic nuclear medicine: a catalog. *Radiology, 248*(1), 254-263. doi:10.1148/radiol.2481071451

Miao, S., Wang, Z. J., Zheng, Y., & Liao, R. (2016). A CNN Regression Approach for Real-Time 2D/3D Registration. *IEEE Transactions on Medical Imaging, 35*(5), 1352—1363. doi:10.1109/TMI.2016.2521800

Middleton, I., & Damper, R. I. (2004). Segmentation of magnetic resonance images using a combination of neural networks and active contour models. *Medical Engineering & Physics, 26*(1), 71-86. doi:10.1016/s1350-4533(03)00137-1

Mitulescu, A., Semaan, I., De Guise, J. A., Leborgne, P., Adamsbaum, C., & Skalli, W. (2001). Validation of the non-stereo corresponding points stereoradiographic 3D reconstruction technique. *Medical and Biological Engineering and Computing, 39*(2), 152-158. doi:10.1007/BF02344797

Mu, Z. (2016). A Fast DRR Generation Scheme for 3D-2D Image Registration Based on the Block Projection Method. Paper presented at the *2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Las Vegas, NV, USA, 609-617.

Muhogora, W., & Pitcher, R. (2016). Defining the diagnostic divide: an analysis of registered radiological equipment resources in a low-income African country. *The Pan African Medical Journal, 25*(99). doi:10.11604/pamj.2016.25.99.9736

Mutsvangwa, T., Wasswa, W., Burdin, V., Borotikar, B., & Douglas, T. S. (2017). Interactive patient-specific 3D approximation of scapula bone shape from 2D X-ray images using landmark-constrained statistical shape model fitting. Paper presented at the *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Seogwipo, South Korea, Vol. 2017, 1816-1819.

Mutsvangwa, T. E. M., Veeraragoo, M., & Douglas, T. S. (2011). Precision assessment of stereo-photogrammetrically derived facial landmarks in infants. *Annals of Anatomy - Anatomischer Anzeiger, 193*(2), 100-105. doi:10.1016/j.aanat.2010.10.008

Newman, M. E. (2005). Power laws, Pareto distributions and Zipf's law. *Contemporary Physics, 46*(5), 323-351. doi:10.1080/00107510500052444

Ng, K.-H., & Yeong, C.-H. (2014). Imaging phantoms: conventional X-ray imaging applications *The Phantoms of Medical and Health Physics* (pp. 91-122). Springer, New York, NY: Springer.

Ohl, X., Stanchina, C., Billuart, F., Skalli, W., , & (2010). Shoulder bony landmarks location using the EOS low-dose stereoradiography system: a reproducibility study. *Surgical and Radiological Anatomy, 32*(2), 153-158. doi:10.1007/s00276-009-0566-z

Pereira, S., Pinto, A., Alves, V., & Silva, C. A. (2016). Brain Tumor Segmentation Using Convolutional Neural Networks in MRI Images. *IEEE Trans Med Imaging, 35*(5), 1240-1251. doi:10.1109/TMI.2016.2538465

Pérez-Pérez, A., Alesan, A., & Roca, L. (1990). Measurement error: Inter-and Intraobserver Variability. An Empiric Study. *International Journal of Anthropology, 5*(2), 129-135. doi:10.1007/bf02442082

Prasoon, A., Petersen, K., Igel, C., Lauze, F., Dam, E., & Nielsen, M. (2013). Deep Feature Learning for Knee Cartilage segmentation- Using a Triplanar Convolutional Neural Network. Paper presented at the *Medical Image Computing and Computer-Assisted Intervention — MICCAI 2013*, Berlin, Heidelberg, Vol. 8150, 246-253.

Reyneke, C. (2019). Voxel registration. Bitbucket. Retrieved from https://bitbucket.org/suiroc/voxelregistration/src/master/

Reyneke, C. J. F., Lüthi, M., Burdin, V., Douglas, T. S., Vetter, T., & Mutsvangwa, T. E. M. (2019). Review of 2-D/3-D Reconstruction Using Statistical Shape and Intensity Models and X-Ray Image Synthesis: Toward a Unified Framework. *IEEE Reviews in Biomedical Engineering, 12*, 269-286. doi: 10.1109/RBME.2018.2876450

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. *International Conference on Medical Image Computing and Computer-Assisted Intervention, 9351*(2015 MICCAI), 234-241. doi:10.1007/978-3-319-24574-4_28

Sarkalkan, N., Weinans, H., & Zadpoor, A. A. (2014). Statistical shape and appearance models of bones. *Bone, 60*(2014), 129-140. doi:10.1016/j.bone.2013.12.006

Shah, N., Bansal, N., & Logani, A. (2014). Recent advances in imaging technologies in dentistry. *World Journal of Radiology, 6*(10), 794-807. doi:10.4329/wjr.v6.i10.794

Shvets, A., Iglovikov, V., Rakhlin, A., & Kalinin, A. (2018). Angiodysplasia Detection and Localization Using Deep Convolutional Neural Networks. Paper presented at the *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*, Orlando, FL, USA, 612--617.

Simonyan, K., & Zisserman, A. (2014). Very Deep Convolutional Networks for Large-Scale Image Recognition. *CoRR, abs/1409.1556*.

Staub, D., & Murphy, M. J. (2013). A digitally reconstructed radiograph algorithm calculated from first principles. *Medical Physics 40*(1), 011902. doi:10.1118/1.4769413.

Stegmann, M. B., & Gomez, D. D. (2002). A Brief Introduction to Statistical Shape Analysis. *Informatics and Mathematical Modelling, Technical University of Denmark (DTU), 15*, 11.

TakÁCs, B. (1998). Comparing face images using the modified Hausdorff distance. *Pattern Recognition, 31*(12), 1873-1881. doi:10.1016/S0031-3203(98)00076-4

Victor, J., Van Doninck, D., Labey, L., Innocenti, B., Parizel, P., & Bellemans, J. (2009). How precise can bony landmarks be determined on a CT scan of the knee? *The knee, 16*(5), 358-365. doi:10.1016/j.knee.2009.01.001

Wasswa, W. (2016). *3D approximation of scapula bone shape from 2D X-ray images using landmark-constrained statistical shape model fitting.* (MSC (Med)), University of Cape Town. Retrieved from http://hdl.handle.net/11427/23777

Wernick, M. N., Yang, Y., Brankov, J. G., Yourganov, G., & Strother, S. C. (2010). Machine Learning in Medical Imaging. *IEEE Signal Processing Magazine, 27*(4), 25-38. doi:10.1109/MSP.2010.936730

Xiao, R., Yang, J., Fan, J., Ai, D., Wang, G., & Wang, Y. (2016). Shape context and projection geometry constrained vasculature matching for 3D reconstruction of coronary artery. *Neurocomputing, 195*(2016), 65-73. doi:10.1016/j.neucom.2015.08.110

Yang, L., Ye, L.-G., Ding, J.-B., Zheng, Z.-J., & Zhang, M. (2016). Use of a full-body digital X-ray imaging system in acute medical emergencies: a systematic review. *Emerg Med J, 33*(2), 144-151. doi:10.1136/emermed-2014-204270

Yu, W., Chu, C., Tannast, M., & Zheng, G. (2016). Fully automatic reconstruction of personalized 3D volumes of the proximal femur from 2D X-ray images. *International Journal of Computer Assisted Radiology and Surgery, 11*(9), 1673-1685. doi:10.1007/s11548-016-1400-9

Zhang, B., Sun, S., Sun, J., Chi, Z., & Xi, C. (2010). 3D Reconstruction Method from Biplanar Radiography Using DLT Algorithm: Application to the Femur. Paper presented at the *2010 First International Conference on Pervasive Computing, Signal Porcessing and Applications*, Harbin, China, 251-254.

Zhang, J., Lv, L., Shi, X., Wang, Y., Guo, F., Zhang, Y., & Li, H. (2013). 3-D reconstruction of the spine from biplanar radiographs based on contour matching using the hough transform. *IEEE Transactions on Biomedical Engineering, 60*(7), 1954-1964. doi:10.1109/TBME.2013.2246788

Zhang, W., Itoh, K., Tanida, J., & Ichioka, Y. (1990). Parallel distributed processing model with local space-invariant interconnections and its optical architecture. *Applied Optics, 29*(32), 4790-4797. doi:10.1364/AO.29.004790

Zhang, X., Guo, Y., & Du, P. (2011). The Contour Detection and Extraction for Medical image of the Knee joint. Paper presented at the *2011 5th International Conference on Bioinformatics and Biomedical Engineering*, Wuhan, China, 1-4.

Zhang, Z. (2000). A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 22*(11), 1330-1334. doi:10.1109/34.888718

Zhang, Z., Liu, Q., & Wang, Y. (2018). Road Extraction by Deep Residual U-Net. *IEEE Geoscience and Remote Sensing Letters, 15*(5), 749-753. doi:10.1109/LGRS.2018.2802944

Zheng, G., Dong, X., Zhang, X., & Nolte, L. P. (2007). Automated detection and segmentation of diaphyseal bone fragments from registered C-arm images for long bone fracture reduction. *Computer Methods and Programs in Biomedicine, 87*(1), 1-11. doi:10.1016/j.cmpb.2007.03.002

Zou, K. H., Warfield, S. K., Bharatha, A., Tempany, C. M. C., Kaus, M. R., Haker, S. J., . . . Kikinis, R. (2004). Statistical validation of image segmentation quality based on a spatial overlap index. *Academic radiology, 11*(2), 178-189. doi:10.1016/s1076-6332(03)00671-8

Appendix A: Reconstructed mesh evaluation results

Results shown in the Table A.1 are average results of 10 reconstructed meshes for LAT and AP predicted images with 8 points from the contour of the scapula.

Table A.1: Surface-to-surface distance errors (mm)

Reconstructed mesh instance	Predicted Mesh	Average Hausdorff distance	Hausdorff distance	Average distance
Predicted AP	**0**	1.28	4.94	1.27
	1	1.72	6.92	1.72
	2	1.28	5.89	1.26
	3	1.05	4.46	1.03
	4	1.19	4.14	1.19
Predicted LAT	**0**	1.58	9.78	1.58
	1	1.76	8.47	1.75
	2	1.47	7.25	1.43
	3	1.21	4.92	1.19
	4	1.63	5.49	1.55
Mean		**1.42**	**6.23**	**1.40**

Results shown in the Table A.2 are average results of 10 reconstructed meshes for LAT and AP predicted images with 6 corresponding reproducible scapula landmarks.

Table A.2: Surface-to-surface distance errors (mm)

Reconstructed mesh instance	Predicted Mesh	Average Hausdorff distance	Hausdorff distance	Average distance
Predicted AP	**0**	1.64	8.19	1.58
	1	1.89	8.06	1.78
	2	1.55	6.77	1.47
	3	4.86	21.51	4.86
	4	1.50	6.06	1.50
Predicted LAT	**0**	1.84	9.08	1.75
	1	1.57	7.50	1.57
	2	1.67	6.71	1.57
	3	1.21	4.99	1.19
	4	1.92	6.50	1.89
Mean		**1.96**	**8.54**	**1.91**