3D Shape Reconstruction of Knee Bones from Low Radiation X-ray Images Using Deep Learning

by

Hampali Shamanth

A thesis

presented to the University of Waterloo

in fulfilment of the

thesis requirement for the degree of

Master of Applied Science

in

Mechanical and Mechatronics Engineering

Waterloo, Ontario, Canada, 2021

Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Understanding the bone kinematics of the human knee during dynamic motions is necessary to evaluate the pathological conditions, design knee prosthesis, orthosis and surgical treatments such as knee arthroplasty. Also, knee bone kinematics is essential to assess the biofidelity of the computational models. Kinematics of the human knee has been reported in the literature either using *in vitro* or *in vivo* methodologies. *In vivo* methodology is widely preferred due to biomechanical accuracies. However, it is challenging to obtain the kinematic data *in vivo* due to limitations in existing methods. One of the several existing methods used in such application is using X-ray fluoroscopy imaging, which allows for the non-invasive quantification of bone kinematics.

In the fluoroscopy imaging method, due to procedural simplicity and low radiation exposure, single-plane fluoroscopy (SF) is the preferred tool to study the *in vivo* kinematics of the knee joint. Evaluation of the three-dimensional (3D) kinematics from the SF imagery is possible only if prior knowledge of the shape of the knee bones is available. The standard technique for acquiring the knee shape is to either segment Magnetic Resonance (MR) images, which is expensive to procure, or Computed Tomography (CT) images, which exposes the subjects to a heavy dose of ionizing radiation. Additionally, both the segmentation procedures are time-consuming and labour-intensive.

An alternative technique that is rarely used is to reconstruct the knee shape from the SF images. It is less expensive than MR imaging, exposes the subjects to relatively lower radiation than CT imaging, and since the kinematic study and the shape reconstruction could be carried out using the same device, it could save a considerable amount of time for the researchers and the subjects.

However, due to low exposure levels, SF images are often characterized by a low signal-to-noise ratio, making it difficult to extract the required information to reconstruct the shape accurately. In comparison to conventional X-ray images, SF images are of lower quality and have less detail. Additionally, existing methods for reconstructing the shape of the knee remain generally inconvenient

since they need a highly controlled system: images must be captured from a calibrated device, care must be taken while positioning the subject's knee in the X-ray field to ensure image consistency, and user intervention and expert knowledge is required for 3D reconstruction.

In an attempt to simplify the existing process, this thesis proposes a new methodology to reconstruct the 3D shape of the knee bones from multiple uncalibrated SF images using deep learning. During the image acquisition using the SF, the subjects in this approach can freely rotate their leg (in a fully extended, knee-locked position), resulting in several images captured in arbitrary poses. Relevant features are extracted from these images using a novel feature extraction technique before feeding it to a custom-built Convolutional Neural Network (CNN). The network, without further optimization, directly outputs a meshed 3D surface model of the subject's knee joint. The whole procedure could be completed in a few minutes. The robust feature extraction technique can effectively extract relevant information from a range of image qualities.

When tested on eight unseen sets of SF images with known true geometry, the network reconstructed knee shape models with a shape error (RMSE) of 1.91± 0.30 mm for the femur, 2.3± 0.36 mm for the tibia and 3.3± 0.53 mm for the patella. The error was calculated after rigidly aligning (scale, rotation, and translation) each of the reconstructed shape models with the corresponding known true geometry (obtained through MRI segmentation). Based on a previous study that examined the influence of reconstructed shape accuracy on the precision of the evaluation of tibiofemoral kinematics, the shape accuracy of the proposed methodology might be adequate to precisely track the bone kinematics, although further investigation is required.

Acknowledgements

I thank my supervisor, Dr. Naveen Chandrashekar, for his consistent guidance and support. I am grateful for his optimism and belief in me. Many thanks to Jin Zhu for her constant motivation. If not for the support and care provided by my family, *great* friend Harish Rao, and wonderful roommates Prasannaah Hadagali and Priyanka Hadagali, my output would have been significantly reduced. I am fortunate to have them all!

I would also like to express my gratitude to the online community for deep learning; thank you for compiling many excellent works and sharing the knowledge.

Table of Contents

List of Figures

List of Tables

Chapter 1 Introduction

Understanding the natural and pathological function of the knee joint requires a thorough understanding of the movement of the knee bones during dynamic activities [1]–[3]. Researchers often recreate the movement of the bones using the observed trajectories of markers placed on the skin, which are not rigidly linked to the underlying bones due to the interposed soft tissues. The local mobility of these markers (commonly referred to as soft tissue artifacts) results in errors that may be of the same order of magnitude as the movements of the bones under the investigation [4]. Due to this issue, assessing the biofidelity of several knee computational models could become difficult. Some reliable methods that can accurately track knee bones are photogrammetric analysis using intracortical pins [5] and stereo radiography with bone-implanted radio markers [6]. These provide accurate information, but they are invasive and risky [4]. Bone tracking using X-ray fluoroscopy is a reliable and non-invasive method that quantifies bone kinematics through X-ray image analysis.

Among fluoroscopy methods, single-plane fluoroscopy (SF) is an advantageous technique for studying joint kinematics because it enables subjects to perform activities freely [3], [7], [8], [8]–[15], [15]–[19]. Although bi-planar fluoroscopy (BF) with two orthogonal units could have more precise outcomes, it would place more constraints on the subject's motion making it difficult to perform activities [13]. Moreover, the radiation dose is doubled in BF compared to SF. Since SF devices are more commonly used for intra-interventional imaging during surgeries, they are more easily accessible to researchers who want to study subject-specific knee kinematics.

The lack of depth detail in SF images makes 3D kinematics analysis challenging. So, the critical stage in measuring the knee kinematics is two–dimensional–to–three–dimensional (2D-3D) registration [20], which establishes the poses of the knee bones in the camera's coordinate system. Initially, researchers used a library-based registration technique to track knee implants on fluoroscopic sequences

[9], [10]. These methods used the implant's known shape for tracking. Later, researchers extended the concepts used for tracking implants to track bones through a system for 2D-3D rigid registration[2], [21]–[23]. The 3D model used for the registration is generally generated by segmenting 3D magnetic resonance (MR) or computed tomography (CT) images [9]–[11]. Both the techniques can achieve adequate precision; however, bone segmentation can be labour intensive. Additionally, CT scans expose the subjects to high radiation and MR scans are expensive.

A solution that could greatly simplify the study of knee kinematics is to generate 3D representations of models of the knee joint using the same SF device used for kinematics analysis. The technique of 3D model generation from 2D images is generally referred to as 2D-3D shape reconstruction or simply 3D shape reconstruction. 3D shape reconstruction using SF images is significantly less expensive than MR scans and exposes the subject to substantially less radiation than CT scans [1], [24]. Since the two procedures (3D shape reconstruction and 2D-3D registration) could be performed under the same roof at the same time using the same device, it could save a considerable amount of time for the subjects and that of the researchers.

Currently, there are a few techniques [20], [25]–[27] available to reconstruct the knee shape from SF images. They all involve performing complicated procedures like calibration, image distortion correction before acquiring the images. Also, several constraints might be placed on the subject's movement during the image acquisition in these techniques. Additionally, the current techniques require manual intervention and expert knowledge during 3D reconstruction. Most importantly, due to low radiation, the images generated from SF are of poor quality (less detail), making them challenging to interpret.

This research aims to develop an end-to-end framework to reconstruct the 3D shape of a knee joint from SF images. The framework is intended to assist the researcher in generating 3D models of the

knee joint quickly and accurately in a hassle-free manner. The work outlined in this thesis lays the groundwork for the second stage of 3D kinematic analysis - 2D-3D registration.

In this approach, for reconstructing the 3D shape of a knee, the subject can be asked to freely rotate the leg (in a fully extended, knee-locked position) in an SF device, resulting in several images captured in arbitrary poses. It is believed that free leg rotation is the simplest activity a subject can perform to generate the maximum information required for the complete 3D shape reconstruction. The information from the images is then extracted using a novel feature extraction technique and analyzed using a custom-built Convolutional Neural Network (CNN). Without further optimization, the network produces a meshed 3D shape model of the subject's knee joint. The whole procedure could be completed in a few minutes.

1.1 Hypothesis

The research objective was to reconstruct the 3D shape of the knee joint from multiple uncalibrated single-plane fluoroscopy (SF) images in a quick, accurate and hassle-free manner.

It was hypothesized that using multiple uncalibrated SF images featuring random poses of the knee taken during a subject's free leg rotation, it is possible to reconstruct the 3D shape of the femur, tibia, and patella.

The required accuracy of the reconstructed model for precise evaluation of kinematics (during 2D-3D registration) was estimated from the literature and is discussed in section 2.5.3

Chapter 2 Background and Related Work

This chapter provides the necessary background, including knee morphology, statistical shape model, knee imaging techniques, and existing methods for reconstructing 3D knee shape from 2D medical images. Additionally, it presents the deep learning methods and architectures used in this thesis. This chapter concludes with a discussion outlining the challenges in the existing methods in achieving the research objectives and how these challenges can be addressed.

2.1 Knee Anatomical Background

The application of biomechanical engineering concepts requires prior knowledge of human anatomy. This necessitates specifying the standard terms used to define the relative locations of various parts of the body. This section provides a summary of anatomical nomenclature, knee anatomy and morphology.

2.1.1 Nomenclature

Three anatomical planes are specified to denote cross-sections and motions:

- *coronal/frontal*: division of the body into front and back,
- *sagittal:* division of the body into left and right, and
- *axial*: division of the body into cranial and caudal.

Additionally, the directional pairs are used to characterize the relationship between the position of any two structures:

- *anterior-posterior*: towards the front or back of the body,
- *superior-inferior:* towards the head or feet,

- *medial-lateral:* towards or away from the body's midline,
- *proximal-distal:* closer to or away from the torso, and
- *superficial-deep:* towards or away from the skin.

The anatomical planes and directions are shown in Figure 2.1

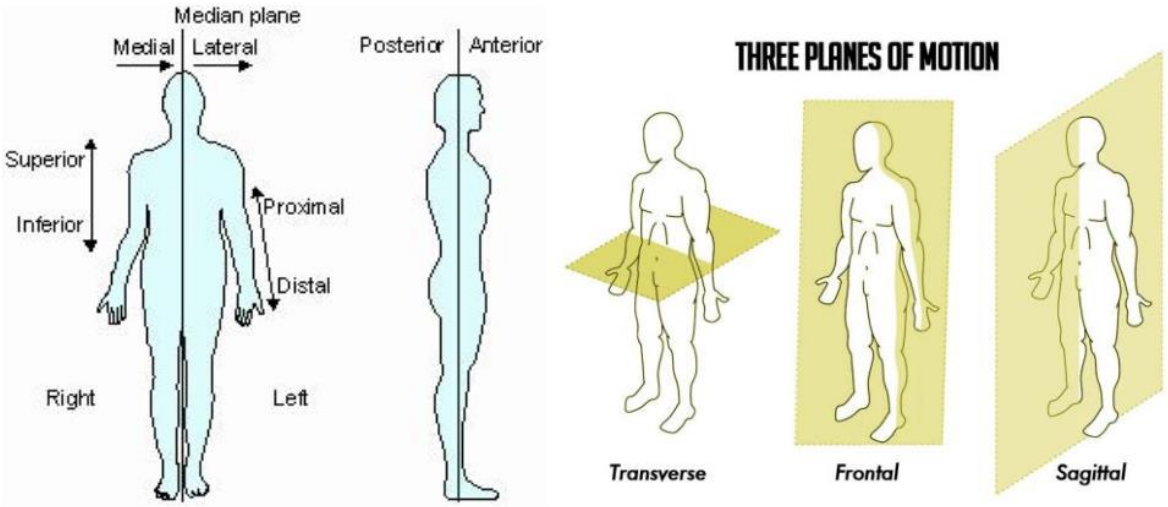


Figure 2.1 Anatomical planes and directions. Adapted from [28]

2.1.2 Knee Anatomy

The knee joint is the largest joint in the human body. It mainly consists of bones and soft tissues like cartilages, menisci, ligaments, tendons, and muscles. The three bones that constitute the knee joint are the femur (thighbone), the tibia (shinbone), and the patella (kneecap). Bones are hard tissues that maintain the structures and keep the shapes of the body. Compared to the soft tissues, bones are high-density structures that are easily visible through X-ray imaging. Figure 2.2 shows the anatomical structure of a human knee joint.

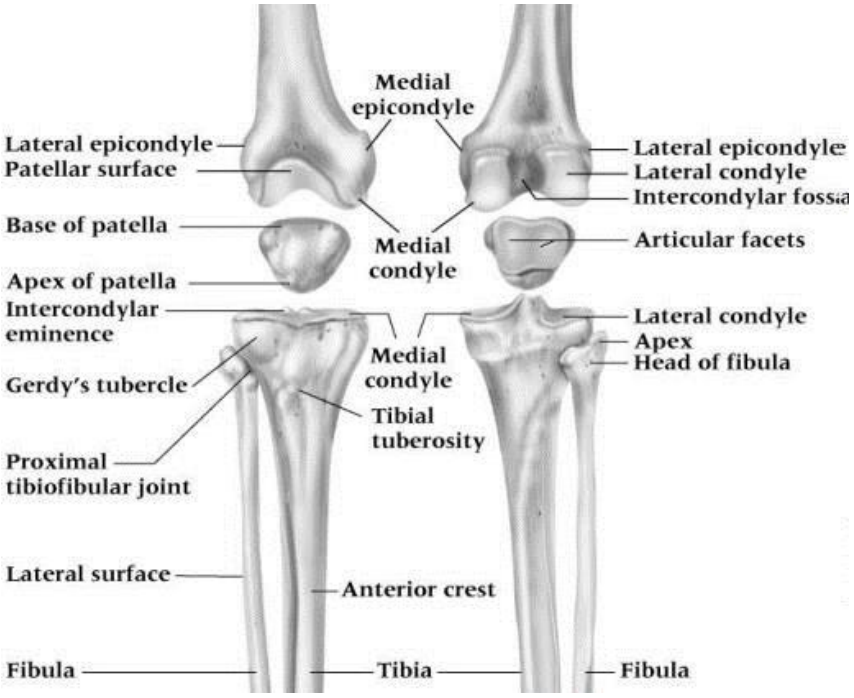


Figure 2.2: Anatomical structure of the human knee joint. Adapted from [29]

2.1.3 Knee Morphology

The knee is considered the most complex joint in the human body. This section gives a brief overview of the knee bony landmarks that help to characterize the complicated shape of the knee.

The distal end of the knee consists of many significant bony landmarks, including the lateral and medial epicondyles [30]. The bony protrusions' more proximal features are seen here. The lateral and medial condyles of the femur are located distal to the epicondyles. These condyles attach to the tibia at the bottom. The patellar surface is located between the condyles and articulates with the patella or knee cap. These landmarks serve as critical attachment points for ligaments and muscles. The intercondylar fossa is located on the posterior aspect of the distal femur. A wide posterior opening is found between the two condyles.

The tibia is the second largest bone in the knee, and it articulates with the femur to form the knee joint's primary component. The tibia is larger at the proximal and distal ends to improve knee and ankle stability. The lateral and medial condyles on the proximal tibia match the corresponding structures on the distal femur. The lateral condyle is smaller than the medial condyle [31]. The intercondylar eminence is a bony ridge that connects the two condyles and blends into the femur's intercondylar fossa.

The patella is a sesamoid bone that forms within the quadriceps muscle tendon [32]. Sesamoid bones are nodules that form in some tendons and are circular or oval. Because of their similarity to sesame seeds, they were initially called sesamoid seeds [33]. They are most often seen where tendons touch the ends of long bones (as in the knee) and help to shield the tendon from wear and tear.

The shape of the knee bones is known to affect the knee's mechanics [34], [35], and the variations in knee shapes have been observed across the population [31], [36]. During fluoroscopic experiments, knee kinematics are measured by registering the representation of the knee shapes to the fluoroscopic image sequences. The availability of quantitative information on the complex shapes of the knee bones is essential to build a shape reconstruction framework. A mathematical model to quantify various shapes of knee bones is discussed in the following Section.

2.2 Statistical Shape Model

Statistical shape models (SSM) [37] are commonly used to describe an object's shape variability. Using a point distribution model, an object in a collection can be represented as a set of labelled points around their mean positions with multiple modes of variations. Principal component analysis is used to describe the variability in shapes. The mathematical model has several advantages, including quantifying data variability, allowing for the development of new 'virtual' geometries that are representative of the population (also known as the training set), and effectively modelling a subject-

specific representation from imperfect or fragmented data sets using less intrusive approaches (e.g. X-ray imaging) [38].

Numerous works based on SSM [37] can be found in biomechanics, including those related to bones [17–19]. SSMs allow population-based assessments of knee joints by characterizing anatomical variations in size and shape between the population [35]. SSMs have traditionally concentrated on individual bones with applications such as implant design [42] and bone joint disorders [43], but they can also be built to study the shape of knee bones and their relative alignments [38].

2.2.1 The University of Denver's Statistical Shape Model

The only publicly available statistical shape model (SSM) for the knee is established by the University of Denver in 2014. It is freely downloadable from the SimTK website [44]. Their study aimed to create a statistical shape and alignment modelling (SSAM), an extension of SSM, that can characterize intrasubject variability in knee bone shapes along with their relative alignments. The SSAM [38] was used in the proposed research to virtual geometries during the deep learning network's training. For more details, the readers may refer to Section 3.3. The remainder of this Section is focused on the SSAM's specifics.

The SSAM [38] was created by segmenting Magnetic Resonance (MR) images of the distal femur, the proximal tibia, the patella bone, and the associated articular surface cartilages of 50 healthy knees (25 males, 25 females). Select soft tissue structures' attachment sites and line-of-actions were also included. Each bone's local coordinate system was described as triads with 1D beam components. The joint was aligned in the MRI as-scanned spot. R3D3 tri elements were used to model bones as 2D surface meshes (Figure 2.3). There were 2384, 1101, and 472 surface nodes (vertices) for the femur, the tibia, and the patella in the training set.

To register the training set geometries to the template mesh, an iterative closest point (ICP) algorithm [45] was used. The subject meshes were reported to be finer than the reference mesh; average edge lengths for the femur and tibia were 0.4 mm and 0.15 mm, respectively, compared to 3.0 mm for the template.

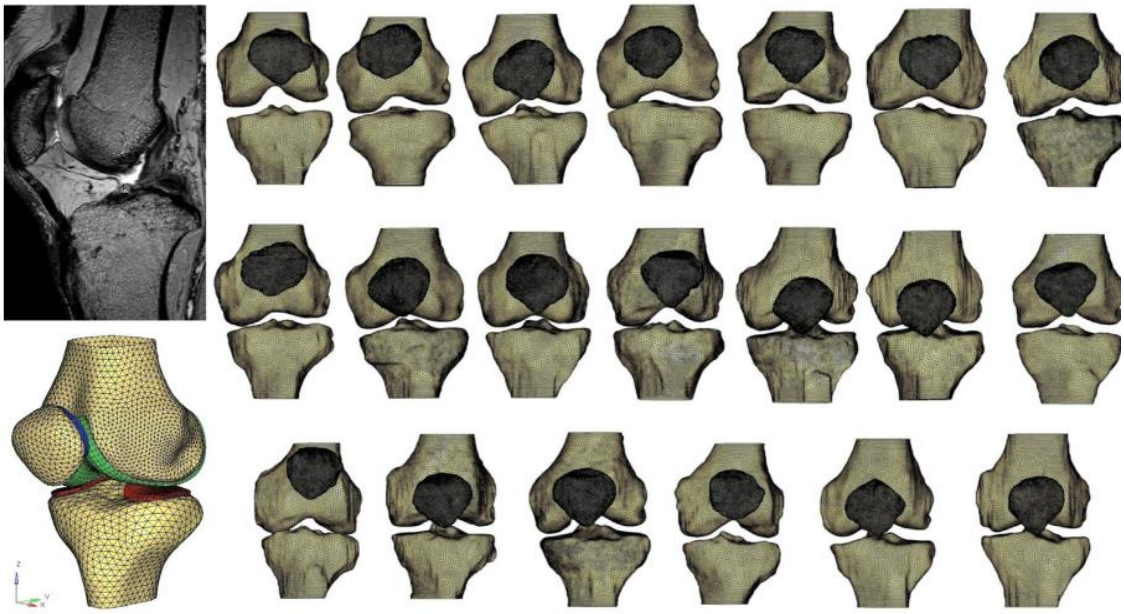


Figure 2.3 The University of Denver's SSM representation. Adapted from [46]

2.3 Knee Imaging Techniques

Medical imaging is often used to provide geometrical details on human knee joints. Computed tomography (CT), magnetic resonance imaging (MRI), positron emission tomography (PET), and X-ray imaging are the most widely used diagnostic imaging techniques [47]. CT and MRI are the most widely used imaging techniques for knee joints. During CT imaging, X-rays are emitted by an X-ray tube, attenuated by the patient, and measured by an X-ray detector. Air appears black due to its low x-ray linear attenuation coefficient; bone appears white due to its much higher x-ray linear attenuation coefficient. Different shades of grey are used to represent different kinds of soft tissues [48]. The MRI

scanner detects radio signals produced by hydrogen atoms in body tissues using a magnetic field and radio waves. Blood, fat, protein, cartilages, bones, and other body parts are expressed in varying degrees of illumination [49]. CT imaging, in general, shows very simple bone structures and is better for regions sensitive to motion, such as the lungs and intestine, but it exposes the subject to heavy ionizing radiation. MRI imaging, however, is safer and more appropriate for soft tissues like ligament and cartilage[50]–[52], but they are more expensive compared to CT imaging.

MRI and CT imaging are pre-interventional 3D imaging techniques. On the other hand, intra-interventional imaging techniques like single-plane, and bi-planar fluoroscopy, provide real-time anatomical information during the surgical procedure. Intra-interventional images are quick to produce and can show the position of a tool or a bone in real-time. In addition, compared to pre-interventional images, these images are exposed to significantly lower radiation. However, such images provide much less detail than pre-interventional 3D modalities [12].

2.3.1 Single-plane Fluoroscopy

This sub-section provides a brief overview of the single-plane Fluoroscopy (SF) in its uses, radiation exposure levels and construction. SF is a low radiation X-ray videography used to study moving body structures. The body part being studied is exposed to a continuous X-ray beam. The beam is captured and transmitted to a digital monitor, displaying the body part and its motion in detail. SF allows doctors to examine a variety of body structures, including the skeletal, digestive, urinary, respiratory, and reproductive systems. SF can be used to examine individual body parts like bones, muscles, and joints, as well as healthy organs like the heart, lungs, and kidneys.

SF images are a series of low-radiation X-ray images in which the bone appears much darker than the soft tissues around it, enabling direct observation and study of its movement under dynamic and weight-bearing conditions. A standard one-minute protocol exposes the subject between 0.6 to 1.8 rem of

radiation [1]. The International Commission on Radiological Protection (ICRP) recommends that the practical dose limit for the public and radiation users be expressed as one mSv per year [53]. This means a subject can be exposed to up to 20,000 knee SF images [54] a year and can remain within safe limits.

SF is often performed using a C-arm fluoroscopy device (Figure 2.4). Due to its simplicity of construction, it can be found in hospital operating rooms across the world. The C-shaped arm that connects the X-ray source and the X-ray detector is the source of the name. C-arms can perform radiography, but they are often used for fluoroscopic intraoperative imaging during surgical, orthopedic, and emergency procedures. The devices can produce high-resolution X-ray images in real-time, enabling the physician to track progress and make any necessary adjustments right away.

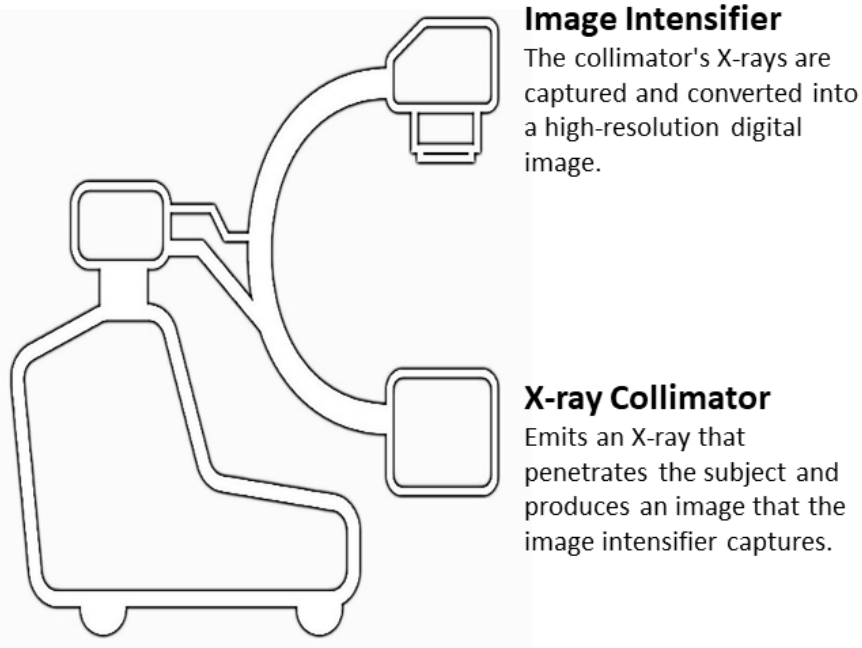
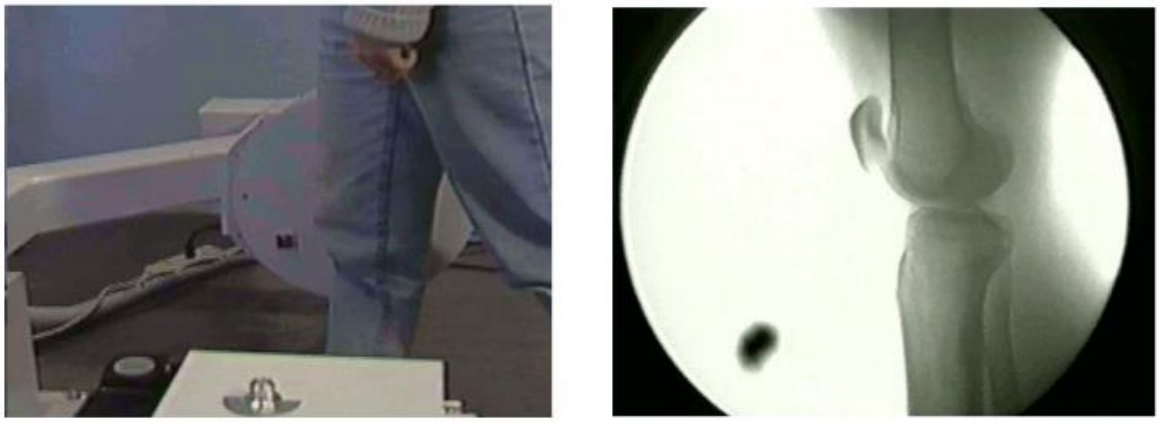


Figure 2.4: Construction of a typical C-arm fluoroscopy machine

The disadvantage of a low-radiation X-ray scan is a low-quality image (

Figure 2.5) with significant information loss. In comparison to conventional X-ray radiography images, SF images exhibit increased quantum (mottle) and structural noise [55] due to the lower number of X-ray photons emitted.



(a) (b)

Figure 2.5: Fluoroscopic imaging (a) fluoroscopic machine (b) an example of the fluoroscopic image of a knee. Adapted from [12]

2.4 Deep Learning

Numerous machine learning and signal processing approaches employ shallow architectures of one or two levels of nonlinear feature transformations [56]. Support vector machines (SVM), logistic and kernel regressions, Gaussian mixture models (GMM), and maximum entropy models are examples of shallow architectures. These architectures have proven extremely successful and have shown promise in various straightforward or well-constrained problems [56].

Their basic modelling and representational capabilities, on the other hand, could be insufficient to cope with more complex real-world applications where a high degree of flexibility is required. For such applications, deep learning algorithms and architectures can be powerful and efficient.

Deep learning comprises deep learning training algorithms that use multi-layer networks of nonlinear processing units at each layer to learn features [56]. In deep learning, several learned representations refer to different levels of abstraction; the levels correspond to hierarchical latent features, and higher-level features are derived from lower-level features [56].

Generally, machine learning models have been trained to perform useful tasks using features derived manually from raw data or features learned from other basic machine learning models [57], [58]. Computers automatically learn useful interpretations and features from raw data in deep learning, obviating this time-consuming and manual phase. The primary feature that all deep learning approaches have in common is their focus on feature learning: automatically learning representations of data [56]. This is the primary distinction between approaches based on deep learning and those based on more "traditional" machine learning. The discovery of features and the performance of a task are combined into a single problem, and thus both are enhanced during the same training period.

Multilayer deep neural networks, recurrent neural networks, convolutional neural networks (CNN), deep belief networks, auto encoders, and deep stacking networks are examples of state-of-the-art deep learning architectures [56]. The interest in deep learning in medical imaging is primarily driven by convolutional neural networks (CNNs) [59], a powerful technique for learning useful interpretations of images and other structured data. Prior to the advent of efficient CNNs, these features had to be manually engineered or generated using less effective machine learning models.

2.4.1 Convolutional Neural Networks

In principle, when applying neural networks to pictures, one may use the basic feedforward neural networks. However, connecting every node in one layer to every node in the next layer is highly inefficient [60]. Pruning connections based on domain information, i.e. the structure of images, results in significantly improved efficiency. A CNN is a form of artificial neural network specifically designed

to preserve spatial relationships in data using very few connections between the layers [61]. The input to a CNN is structured in a grid and then fed through layers that maintain these relationships, with each layer operates on a small region of the previous layer [62]. CNNs are capable of forming highly efficient representations of input data, making them well-suited for image-based tasks. A CNN is composed of several layers of convolutions and activations, often interspersed with pooling layers, and is trained using backpropagation and gradient descent in the same way as regular artificial neural networks [61]. Additionally, CNNs usually have completely linked layers at the conclusion that compute the output results.

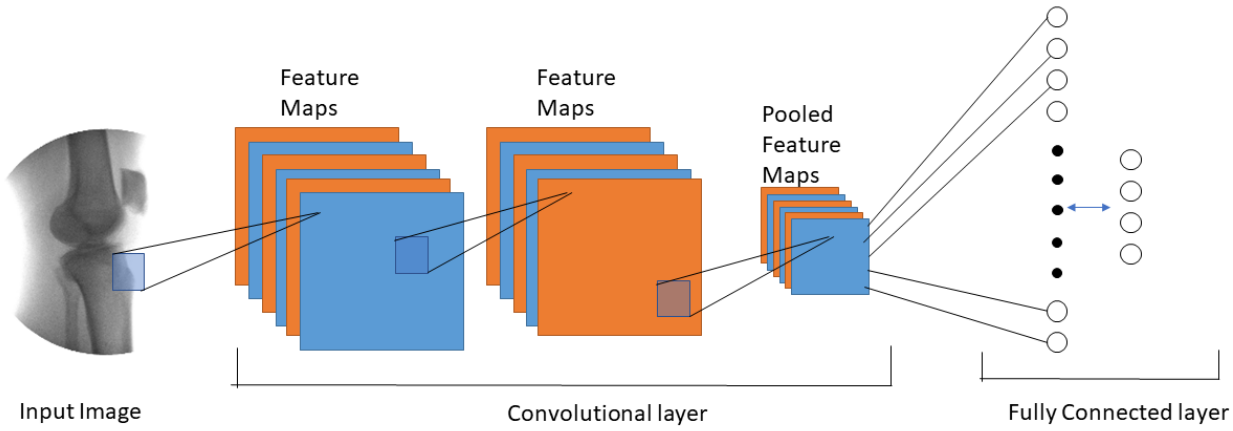


Figure 2.6: Building blocks of a typical CNN

Convolutional layers: Convolutional layers combine the behaviour of previous layers with a collection of small, parameterized filters, usually of size 3X3, and store them in a tensor $T^{(i,j)}$Where *j* is the filter number, and *i* is the layer number. By sharing the identical weights across the entire input domain, i.e. translational equivariance at each layer, one may significantly reduce the number of weights that must be learned [61]. The reason for this weight-sharing is that features that appear in one area of the picture are almost certain to appear in other areas. If one has a filter that can detect horizontal lines, then that

could be used to detect lines anywhere they appear in an image—applying all convolutional filters to a convolutional layer at all positions in the input results in a tensor of feature maps [63].

Pooling: After feeding the data through one or more convolutional layers, each function map is usually pooled in a pooling layer. Pooling operations take input from small grid regions and generate unique numbers for each region. Typically, the maximum function (max-pooling) or the average function (average pooling) is used to calculate the number. Due to the fact that small changes in the input image result in small changes in the activation maps, the pooling layers provide translational invariance to the CNN [64]. Using convolutions with increased stride lengths is another way to achieve the downsampling effect of pooling [64]. By removing the pooling layers, the network architecture is simplified without sacrificing efficiency.

Dropout: A fundamental concept that significantly improved the efficiency of CNNs. When multiple models are averaged together in an ensemble, one usually obtains better output than individual models[65]. By deleting neurons randomly during training, one ends up with slightly different networks for each batch of training data, and the trained network's weights are tuned based on optimization of multiple network variations.

Batch Normalization: These layers are usually positioned after activation layers, resulting in normalized activation maps for each training batch by subtracting the mean of the data and then dividing by the standard deviation[66]. As the training batch passes through these layers, the network is forced to adjust its activations to zero mean and unit standard deviation, which regularizes the network, speeds up training, and reduces the reliance on careful parameter initialization.

Activation Layer: Nonlinear activation functions are used to feed feature maps from a convolutional layer. As a result, the entire neural network is capable of approximating almost every nonlinear function[67]. The activation functions are typically rectified linear units, or ReLUs defined as

$$ReLU(z) = \max(0, z) \quad (1)$$

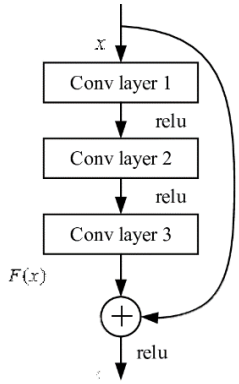
or variants such as leaky ReLUs or parametric ReLUs. For more detail on these and other activation mechanisms, see [68], [69].

Architectures of Convolutional Neural Network

The elements discussed in the preceding Section are combined in increasingly complex and interconnected ways or are even replaced by more convenient operations in designing new and improved CNN architectures. When building a CNN for a specific task, many considerations must be taken into account, including understanding the task at hand and the criteria to be met, determining the optimum way to feed data to the network, and maximizing computing and memory usage within a given budget. In the early stages of modern deep learning, as in Lenet[59] and AlexNet[62], one appeared to use fundamental variations of the building blocks. Later network architectures became significantly more complex, with each generation building on previous architectures' ideas and experiences, resulting in state-of-the-art changes. Table 1 summarises many well-known CNN architectures, demonstrating how the building blocks can be integrated and the field progresses. Usually, these neural networks are implemented in one or more of a limited number of frameworks that dominate machine learning research, all of which are based on the cuDNN library and NVIDIA's CUDA platform [70]. Today's deep learning approaches are almost exclusively implemented using Pytorch [71], a framework developed by Facebook Research, Tensor-Flow [72], a framework developed by Google Research, Keras [73], a deep learning library that was recently integrated into TensorFlow.

Table 1: Some of the popular CNN architectures

Architecture	Description
AlexNet [62]	AlexNet was born out of the desire to boost the ImageNet challenge performance. This was one of the first Deep convolutional networks to attain a high level of accuracy on the 2012 ImageNet LSVRC-2012 challenge, achieving an accuracy of 84.7 percent, compared to the second-best accuracy of 73.8 percent. Convolutional layers and receptive fields were used to investigate the concept of spatial similarity within an image frame. Five Convolutional (CONV) layers and three Fully Connected (FC) layers comprise the network. The ReLU is used as the activation. The figure below contains the structural information of each layer in the network.
VGGNet [74]	VGGNet was developed in response to the need to decrease the number of parameters in the Convolutional layers and speed up training. VGG16 contains 138 million parameters in total. The critical point to remember here is that all conv kernels are 3x3 in size, while maxpool kernels are 2x2 in size with a stride of two. The rationale for using fixed-size kernels is that all Alexnet's variable-size convolutional kernels (11x11, 5x5, 3x3) can be repeated using multiple 3x3 kernels as building blocks. The replication is done in terms of the kernels' receptive region.
ResNet [75]	ResNet introduced skip connections, which enables the training of far deeper networks. A 152-layer deep ResNet won the 2015 ILSVRC competition, and the authors also trained a version with 1001 layers successfully. Having skip connections in addition to the typical path enables the network to simply copy

	activations between layers (or, more specifically, between ResNet blocks), preserving information as data passes through the layers. Certain features function best in shallow networks, while others need greater depth. Both are facilitated by the skip connections, which increase the network's versatility when fed input data. ResNets perform a type of boosting because the skip connections cause the network to learn residuals. 
U-Net [76]	U-Net is a common and efficient network for 2D image segmentation. When an input image is fed into the system, it is first downsampled using a "standard" CNN and then upsampled using transpose convolutions to its original scale. Additionally, based on ResNet concepts, skip connections link features from the downsampling to the upsampling routes.

Domain Shift

Domain shift [77]–[79] is the difference in data distribution between the testing dataset (source domain) and the test dataset (target domain) for a machine learning model. The domain shift problem often occurs in real-world implementations of different machine learning algorithms and can result in substantial performance degradation. The domain shift problem often plagues deep learning methods

in medical image analysis. In the context of this thesis, domain shift is common in single-plane fluoroscopic images due to variations in the fluoroscopic devices, imaging protocols, and radiation doses.

Unlike regular image processing, which uses large-scale labelled databases such as ImageNet [62], the absence of labelled data poses a significant obstacle in developing accurate and stable machine learning models in medical image analysis. Medical image labelling is particularly expensive, time-consuming, and repetitive, and it necessarily requires the involvement of surgeons, radiologists, and other specialists. An intuitive solution is to generate synthetic data for training the deep learning model; however, if the training dataset's domain changes from that of the real-world dataset, the accuracy of the deep learning model's performance can suffer significantly [80].

Domain adaptation (DA) has received much coverage in recent years as a promising approach to the domain shift problem. A DA technique aims to minimize the distribution differences in the images among different but related domains. Many researchers have used domain adaptation approaches to address various medical image processing challenges [80]. Generally, the DA is constructed using a machine learning algorithm that needs extensive training. [78]. In rare cases, researchers have shown that feeding handcrafted features to the CNN can enhance the performance [81]–[89]. A 2018 study [90] uses Gabor filter [91], a linear filter used for texture analysis, to extract facial features for a face-related task. In this thesis, a multi-scale feature extraction technique based on the Canny edge detection algorithm [92] was used to transform images from different sources (virtual or real) into a common representative image. It acts as a preprocessing step for the images before being fed into the deep learning network.

2.5 3D Shape Reconstruction

Numerous 3D model reconstruction techniques exist for converting 2D medical images to 3D models. This section introduces the segmentation procedure - a conventional technique for the 3D model reconstruction of the knee joint from MR and CT images. Later, the state-of-the-art techniques for reconstruction of the knee joint from X-ray radiographs are discussed.

2.5.1 Reconstruction from MR/CT Images

A process known as segmentation is used to create three-dimensional (3D) models from MR or CT image volumes. To separate an image volume into several corresponding regions of interest, a label field is applied to image voxels (3D pixels) with similar characteristics. A 3D model is generated when the segmentation process is repeated on all the images (or slices) in the scan. Segmentation is typically performed using medical image processing tools such as Scan, Mimics, Amira, and 3D Slicer (www.slicer.org). 3D Slicer was chosen for image segmentation in the current study because it freely offers all of the required manual and automated image segmentation resources.

Automatic segmentation methods can be divided into four types: thresholding, area expanding, clustering, and registration-based techniques[93]. Thresholding is determined by a greyscale intensity value in an image used to divide segments into two classes, one above and one below the threshold value [94]. In the area expanding technique, a seed is put in an image's region of interest, and the increasing region algorithm expands the seed by finding neighbouring areas of the same strength. Clustering, like area expanding, employs a distance feature to similar group pixels or image patches into a single subset [94]. Registration-based techniques use the principle of image registration, where a transformation is applied to all points in an image to coordinate it with another image of the same anatomical component. Registration-based techniques have the potential to be fully automated [95].

The 3D model generated from MR or CT image segmentation is considered a gold standard due to its accurate representation of the anatomical geometry [96]. A 2012 study [97] reports an average error of 0.23 mm for 3D models of lower limbs generated from MR image segmentation. To evaluate the performance of the proposed 3D reconstruction framework, the reconstructed model was compared with the model generated through the segmentation of corresponding MR images.

2.5.2 Reconstruction from X-ray Images

The two most widely used medical imaging methods for 3D model reconstruction are computed tomography (CT) scanning and magnetic resonance imaging (MRI). Although CT scanning is an effective and widely used form of 3D imaging, it exposes the patient to a high radiation dose [50]. This procedure could result in unnecessary risk of radiation exposure in many disorders where repeated scans are needed for clinical follow-up exams. Additionally, CT scans are expensive and are done with the patient lying down, altering the global outline of bone structures [50], [98]. This becomes much more important as it comes to illnesses affecting the spine and the knees. Methods based on MRI, on the other hand, may not have the issue of high irradiating dose; however, they are more expensive than CT imaging, and they are not appropriate for patients with ferromagnetic materials due to the potential risks associated with movement or dislodgment of these items, as well as artifacts caused by metallic objects [95].

As a result, numerous studies have been conducted to introduce alternative approaches that remove these limitations while still providing realistic 3D knowledge. Because of its precision, low irradiating dosage, and potential to be applied in a standing or weight-bearing position, photogrammetry for making 3D models from radiographic images is currently considered a viable alternative solution. A 2015 study [99] provides a detailed overview of the various state-of-the-art photogrammetry-based techniques for 3D reconstruction from radiographic images. The study's work

is briefly summarised in this Section after introducing the general methodology followed in the 3D reconstruction from radiographic images.

General methodology

The important aspects involved in the 3D reconstruction from radiographic images are as follows:

1. *Calibration of the Imaging Device*: Single-plane fluoroscopy (SF) or X-ray imaging may be represented as a perspective projection model, with the parameters calculated using a calibration method. It is widely acknowledged that accurate intra-procedural calibration is needed for tracking and reconstruction purposes [100]. The calibration is performed for two main purposes:

    i. Determination of intrinsic parameters: During reconstruction, these parameters are required to build a perspective projection model. The intrinsic parameters are classified into two categories: those related to the pixel scale and those related to the focal length (distance between the X-ray source and the image intensifier). Depending on the electronic circuitry, the pixel size of the radiographic images is normally constant during their lifetime. The positioning of the X-ray source, on the other hand, varies from pose to pose due to the mechanical design's stability and the weight of the X-ray source [100]. The intrinsic calibration of the C-arm is position-dependent. As a result, traditional fluoroscopic guidance approaches need calibration of the C-arm at each pose, a major disadvantage for intraoperative navigation. Traditionally, an image intensifier is connected to a calibration phantom, and an X-ray image is taken and sent to a calibration evaluation [100].

ii. Image distortion correction: Distortion of the image may lead to reconstruction error and must be corrected. Radial distortion of the detector, the earth's magnetic field, and perspective distortion contribute to image distortion of radiographic images [99], [101].

2. *Image Acquisition***:** Depending on the application, images may or may not have markers as benchmarks during acquisition. Many kinds of markers, including those surgically inserted into the bone and those directly stuck to the tissue [99]. Although using markers improves accuracy, it also has some major drawbacks, such as inconvenience for patients [99]. Also, there might be severe constraints on placing the knee in a particular position and orientation during the image acquisition.

3. *Feature Extraction/ Image Enhancement***:** X-ray imaging could be extremely noisy, especially when captured from low radiation devices like SF. Due to the small number of X-ray photons present in SF images, there is a significantly higher quantum noise [100]. It is also worth noting that the superimposition of bony structures on X-ray images degrades image clarity. A robust feature extraction/image enhancement technique is essential to extract relevant features while still suppressing the noise efficiently.

4. *3D Reconstruction*: Several methods exist to reconstruct a 3D model from calibrated multiple X-ray images followed by feature extraction or image enhancement. The following section summarises the existing methods for the 3D reconstruction from X-ray images.

Summary of Previous Research

The existing reconstruction methods from X-ray images can be classified into six types: (1) point-based methods, (2) contour-based methods, (3) statistical shape modelling-based methods, (4) parametric

methods, (5) hybrid methods and (6) deep learning-based methods. This classification and the information presented below complements the literature analysis performed by [99].

1. Point-based methods: The methods of this class are based on the identification of low-level primitives, such as points, and their matching on multi-view radiographs. The two key types of 3D reconstruction point-based techniques are Stereo Corresponding Point (SCP) based Techniques[102], [103] and Non-Stereo corresponding Point (NSCP) based Techniques [104]. The SCP method requires correspondence in identified points between images, while the NSCP method does not. These approaches rely heavily on the ability of the specialist operator to identify precise points. As a result, their reproducibility is low. Because of the manual recognition, these techniques take between 2- 4 hours to reconstruct [99].

2. Contour-based Methods: The principle is to link recognizable 2D contours from radiographs to 3D lines identified on the surface of a reference object. Then an elastic 3D model deformation in relation to 2D contours is performed on multiple images [99]. In contrast to point-based approaches, contour-based methods use a higher level of geometric primitives, eliminating the issue of identifying matching points and decreasing the need for user interaction [105], [106]. Despite the appropriate precision of contour-based techniques, 3D reconstruction takes a long time for these approaches [100]. The time taken for these 3D reconstruction methods was estimated to be about 15-35 minutes for lower limbs [100].

3. Statistical Shape Modelling-based Methods: Another approach is to use statistical shape analysis and modelling [20], [25], [37], [42], [105], [107]–[115] for the 3D shape reconstruction. Only a limited number of sparse landmark vertices on the bone's surface is required in this method. Since the knowledge in those sparse landmark vertices is insufficient for a complete 3D surface reconstruction, a priori information is essential. From a collection of training surfaces that reflect

reasonable variations of the surfaces the bone of interest, a statistical model can be built (Section 2.2). The statistical model is then used as prior information to match the subject's anatomy using landmark vertices acquired from radiographs. To summarize, the statistical shape model-based method aims to extrapolate a complete and relatively accurate 3D anatomical surface from a sparse or incomplete set of 3D vertices. Due to their ability to accurately represent structures, model-based approaches are relatively reliable, with recorded reconstruction errors of 1-3 mm [116].

4. Parametric Methods: Rather than using the whole collection of points suggested by SSM modelling, statistics are conducted on anatomical descriptive parameters (DP) derived from the surface of interest[114]. These methods are suitable for generating an initial 3D reconstruction that is both fast and stable. These approaches define a generalized parametric model for the representation of the object of interest. According to the considered bony structure, this generalized parametric model comprises various geometric primitives such as points, lines, circles, and spheres [100].

5. Hybrid Methods: This class includes various types of reconstruction models, hence known as hybrid methods[100]. This class's methods will have a variety of properties depending on how they are combined and used. An example[109] would be the combination of the statistical and image-based biplanar reconstruction process.

6. Deep Learning Methods: Recently, deep learning techniques are being employed to reconstruct the shape of a predefined object from multiple images [117], [118]. Geometrically, it is possible to reconstruct a three-dimensional surface by triangulating two or more images, assuming the cameras' relative positions are known[119]. The following recent studies show how to reconstruct shapes from two or more images using deep learning techniques: Multi-view

Stereopsis[120] and 3D-R2N2 [30] both employ RNNs to combine features from multiple images in order to reconstruct three-dimensional models. In comparison, Pix2Vox[122] reconstructs a single volume from each image and combines them in a context-aware layer. Pixel2mesh++ [123] extends the work [117] by refining it using enhanced 2D features sampled from the projections of the 3D points on the multiple images. In medical imaging, 3D models were reconstructed from simulated bi-planar X-ray images of a single spine vertebra using a deep network using an autoencoder scheme [124]. A recent approach to reconstruct 3D knee geometry uses end-to-end 3D convolutional layers with skip connections, allowing for quicker and more precise reconstruction of multi-class bones for the knee joint [125].

2.5.3 Estimation of Target Accuracy for 3D Reconstruction

3D models generated from the segmentation of MR/CT images are widely regarded as the gold standard to which other methods of 3D reconstruction are evaluated. The root mean square (RMS) distance between MRI/CT segmented bone surface points (commonly referred to as ground truth) and the reconstructed model's surface points is used to measure the shape reconstruction error. Prior to the error measurement, the reconstructed surface is matched with the MRI/CT segmented surface using a 3D/3D rigid registration.

In the context of this thesis, the reconstructed shape of the knee bones should be accurate enough to precisely register the 3D shape to the 2D images during the kinematic analysis. Since the work on the second stage (2D-3D registration) is in progress at the time of writing this thesis, it was decided to set an estimated accuracy from the literature.

A 2014 study [28] investigated the influence of the accuracy of 3D shape reconstruction on the precision of the kinematic tracking of the knee joint. From the study, the author would like to point out two findings that helped set a target error for the reconstructed model:

1. The ability to precisely match the edges in fluoroscopic sequences plays a greater role in assessing kinematic precision than the overall 3D shape accuracy of the model.
2. In one of their experiments, they discovered that the 3D model with the worst RMS error of 1.37 mm for the femur and 1.64 mm for the tibia had comparable kinematic accuracy to the ground truth model.

2.6 Summary

In this chapter, the main concepts in the literature, the state-of-the-art and previous work related to assessing 3D reconstruction from radiographic images were reviewed. Also, the technical background: knee morphology, statistical shape model, knee imaging techniques, deep learning concepts and architectures related to this thesis were introduced.To recap, the objective of the research was to develop a methodology to reconstruct the shape of a knee in a quick, accurate and hassle-free manner from multiple single-plane fluoroscopy (SF) images captured during the free leg rotation. This section addresses the technical challenges involved in using the existing methods to accomplish the research objectives. The section concludes by outlining the approach used in the proposed study to overcome these challenges.

2.6.1 Challenges

Three major challenges need to be addressed to achieve the research objectives using the existing methods. (C1) SF images are generally of poor quality and hence contain less detail. So, it is challenging to extract relevant features for accurate reconstruction; (C2) It is difficult to use existing techniques for reconstructing the shape from images taken during the arbitrary motion since they were not designed

for this purpose.; (C3) the existing methods demand images captured from the calibrated device. Calibration is a tedious procedure.

Each of the challenges is discussed in detail below:

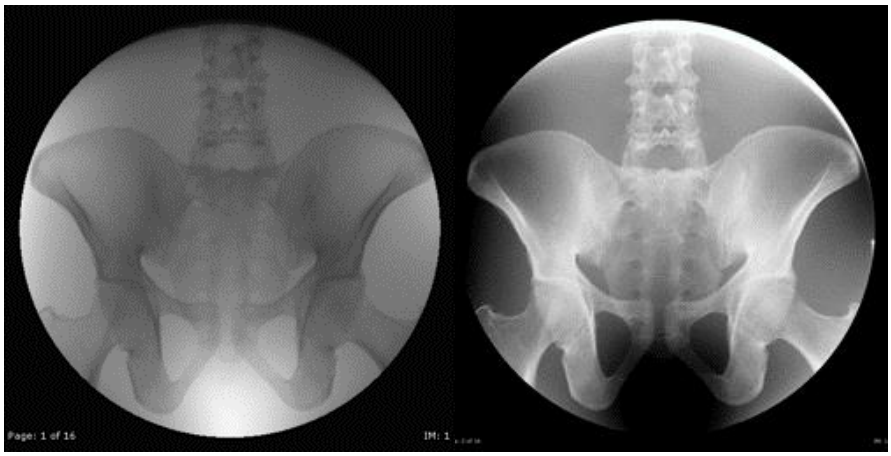
C1. Poor Image Quality

Unlike X-ray radiography (XR) images, SF images are not generated for diagnostic interpretation but to roughly identify the location of the surgical tools and the bones during surgeries. Since an information-rich image is not a critical requirement, the SF devices generate radiation on the order of 50-100 times lesser than XR devices[126]. Lower radiation results in higher mottle(quantum noise) [55], leading to information-poor or low-quality images[127]. This makes it challenging to extract features from the images to perform any further interpretation[128]. Figure 2.7 shows the differences in the image qualities of a pelvis phantom when captured using an SF device and an XR device. Moreover, along with the mottle, one has to deal with low contrast resolution and signal non-homogeneity when analyzing SF image sequences [11].

The information loss at the pixel level can be seen in Figure 2.8. It is hard to extract accurate enough features to reconstruct a precise 3D model at this noise level. Most of the methods discussed in the previous section require clearly defined features from radiographic images for accurate reconstruction. Since SF images are characterized by a low signal-to-noise ratio[129], extracting relevant information for reconstruction is challenging.

Although many techniques exist for the reconstruction of 3D knee models from information-rich XR images[105], [110], [112], [115], [124], [125], [130]–[144], only a few exist for information-poor SF images[20], [25]–[27]. All the existing techniques for SF images are based on statistical shaped modelling methods. In these methods, usually contours of the bones are extracted either using active shape modelling[37] or active contour modelling [145] before fitting a shape model. The two

techniques extract only the bones' external edges, which are imaged in a particular pose on which it is trained. Employing these techniques would be challenging to extract edges when the knee bones are imaged in arbitrary poses.



(a) (b)

Figure 2.7: SF image (a) and XR image (b) of a pelvis phantom captured in the same position. Adapted from [126]

C2. Shape Reconstruction from Free Motion

It is important to make the *in vivo* experiment as constraint-free as possible to keep it subject-friendly. It is believed that allowing the subjects to conduct tasks at their discretion may simplify the experiment process. However, since the images generated this way include a vast range of arbitrary poses, interpretation becomes more challenging.

The structure from motion (SfM) technique seems like a natural choice to reconstruct the 3D shape of an object from a set of images captured in different poses[146]. In medical imaging, SfM has been used for close-range photogrammetry applications; an example would be the 3D reconstruction of the whole stomach from endoscope video [147]. However, in SF images, poor features and the translucent nature of the bones can cause complications as the incorrect features can be linked during the automated

feature matching operation[148]. Also, difficulty in point feature tracking between adjacent frames of SF imagery has been reported in several studies [11], [149]–[151].

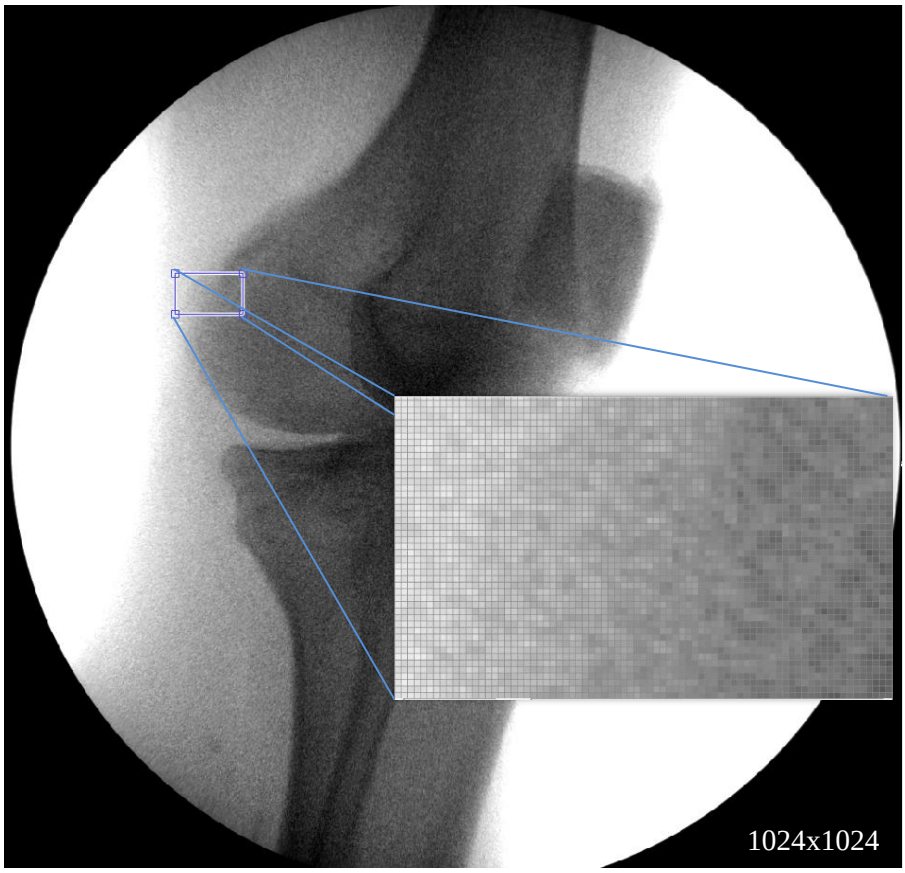


Figure 2.8: A pixel-level view of a fluoroscopic image. The edges at the pixel level are hard to detect.

A 2003 study proposes a methodology for the 3D reconstruction of the computed tomography(CT) from images generated by fluoroscopy along an arbitrary path [152]. However, it requires metallic beads to be attached to the skin as markers to estimate the poses. The pose information is then used in the 3D reconstruction. Although using markers could improve precision, it would cause inconvenience for the subjects [99].

C3. Tedious Calibration Procedure

SF or XR imaging is modelled as a perspective projection with the parameters estimated using a calibration method during reconstruction. It is widely agreed that precise calibration is needed for reconstruction to be effective [27], [110], [113], [139], [142], [144]. However, the calibration procedure is tedious and adds to system complexity[100]. Hence, it is a major obstacle in making the *in vivo* experiments hassle-free.

2.6.2 Approach

The following strategies were followed to address the challenges addressed previously:

A1. Development of A Novel Feature Extraction Technique

A novel feature extraction technique called 'Multi-layer Canny' (Section 3.4) was developed to effectively extract the relevant features required for the 3D shape reconstruction. By combining the concept of the image pyramid [153] and Canny's edge operator algorithm [92], the technique blends global and local (pixel-level) knowledge to detect features while still suppressing the noise**.**

A2. Training the CNN to be Pose-invariant

To accurately reconstruct the shape of a knee from a set of images taken in a wide range of arbitrary poses, a CNN was trained on a massive collection (30,000) of synthetic images generated using huge randomization of pose parameters (Section 3.3). The trained pose-invariant CNN can be used to reconstruct the shape from a wide variety of knee poses.

A3. Use of Multiple Images to Reduce the Effect of Calibration Error

The calibration is performed for two purposes: to determine the intrinsic parameters and to correct the image distortion. Using a theoretical framework, authors in the 2007 paper [100] titled "C-arm

Calibration - is it Really Necessary?" argue the C-arm (single-plane fluoroscopy) calibration could be avoided in certain situations. They studied the influence of the calibration error on the accuracy of 3D shape reconstruction. The authors argue that increasing the number of images will improve the accuracy of the 3D reconstruction despite calibration errors. The 3D shape, in particular, was found to be more stable even for large calibration errors (100 mm error in focal length). Since the requirement of this proposed study is only *shape reconstruction* and not the full-scale model reconstruction of the knee, it was decided to use a large number of images with a generic set of calibration parameters.

In one of the experiments described in the study [100], an object's reconstruction accuracy improved by 50% when six images were used instead of one despite an error of 100 mm in the focal length parameter. To nullify the effect of calibration error on the performance of the proposed framework, it was decided to use 16 images for the reconstruction with a generic set of calibration parameters.

Chapter 3 Methodology

At the end of the previous chapter, three challenges to achieving the research objectives were discussed, and the appropriate solutions were proposed. This chapter discusses the development of a framework for the implementation of these solutions. The goal is to reconstruct the 3D shape of the femur, tibia and patella from a set of single-plane fluoroscopic images featuring arbitrary poses of the knee captured during the free leg rotation. The input and the output of the framework are shown in Figure 3.1.

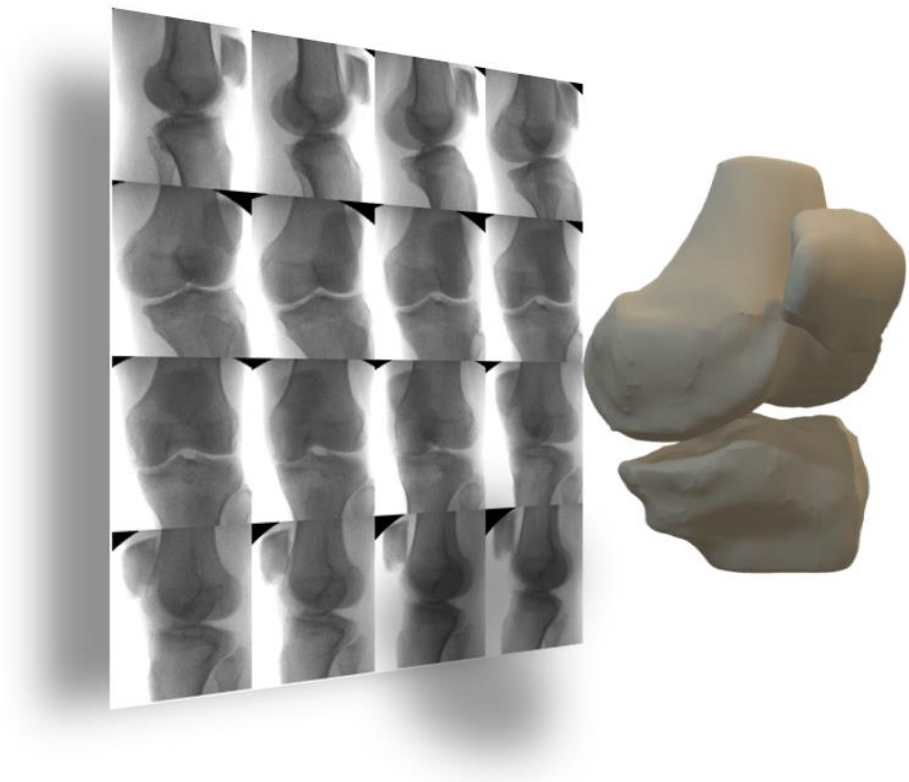


Figure 3.1: The input and the output of the proposed reconstruction framework. The input is a set of single-plane fluoroscopic images featuring arbitrary poses of the knee captured during the free leg rotation, and the output is the 3D shape model of the femur, tibia and patella.

3.1 Method Overview

The problem of 3D shape reconstruction from a sequence of fluoroscopy images was viewed as a mapping *M* from a sequence of fluoroscopy images $\{I_i\}$to a set of spatial coordinates V, corresponding to the 3D structure of the knee bone $(M : \{I_i\} \rightarrow V)$. The mapping function is modelled using a custom-built convolutional neural network (CNN). A supervised learning approach using Procrustes shape distance [154] is followed to *train* the neural network's weights. Once trained, the neural network is provided with a new fluoroscopy image sequence, and the corresponding 3D shape of the bone is predicted. Supervised learning approaches rely on large amounts of annotated data, i.e., network input and output pairs$(\{I_i\}, V)$ with different knee shapes, poses, capture conditions etc. to learn the weights of the network.

Annotating fluoroscopy images with their corresponding 3D shapes is difficult and often involves tedious manual work. Due to the non-availability of such annotated datasets, the CNN is trained using synthetic (computer-generated) fluoroscopic images (Figure 3.2 (c)), whose 3D shapes are generated from a statistical shape model. However, synthetic fluoroscopic images are visually and statistically different from real fluoroscopic images (Figure 3.2 (a)), and hence a network directly trained on synthetic images does not generalize well to real fluoroscopic images. In order to overcome this challenge, a feature extraction technique was proposed, which, when applied on synthetic and real images, results in images with similar visual and statistical characteristics as shown in (Figure 3.2 (b) and (d)). The feature extracted synthetic images and their corresponding 3D bone shapes are used to train the CNN. During the evaluation of real fluoroscopic images, the features of the real images are first extracted and are provided as input to the trained network.

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

Figure 3.2: Comparison of real and synthetic image. (a) Real X-ray image. (b) Feature-extracted real image. (c) Synthetic image (Randomly generated) (d) Feature-extracted synthetic image

The rest of the Section is organized as follows:

- The real X-ray image acquisition process is described in Sec 3.2. The real X-ray images and corresponding MRI images are used to evaluate the accuracy of the proposed 3D shape reconstruction method.
- The synthetic image generation process that mimics the real X-ray image acquisition setup in a virtual environment is discussed in Sec 3.3
- A new feature extraction technique used to obtain visually and statistically similar real and synthetic images is introduced in Sec 3.4
- The convolutional neural network architecture for reconstructing the 3D shape of knee bones from a sequence of feature-extracted fluoroscopic images is discussed in Sec 3.5. The trained network is used to evaluate real X-ray images.

The overview of the approach followed to train and evaluate the deep neural network is shown in Figure 3.3

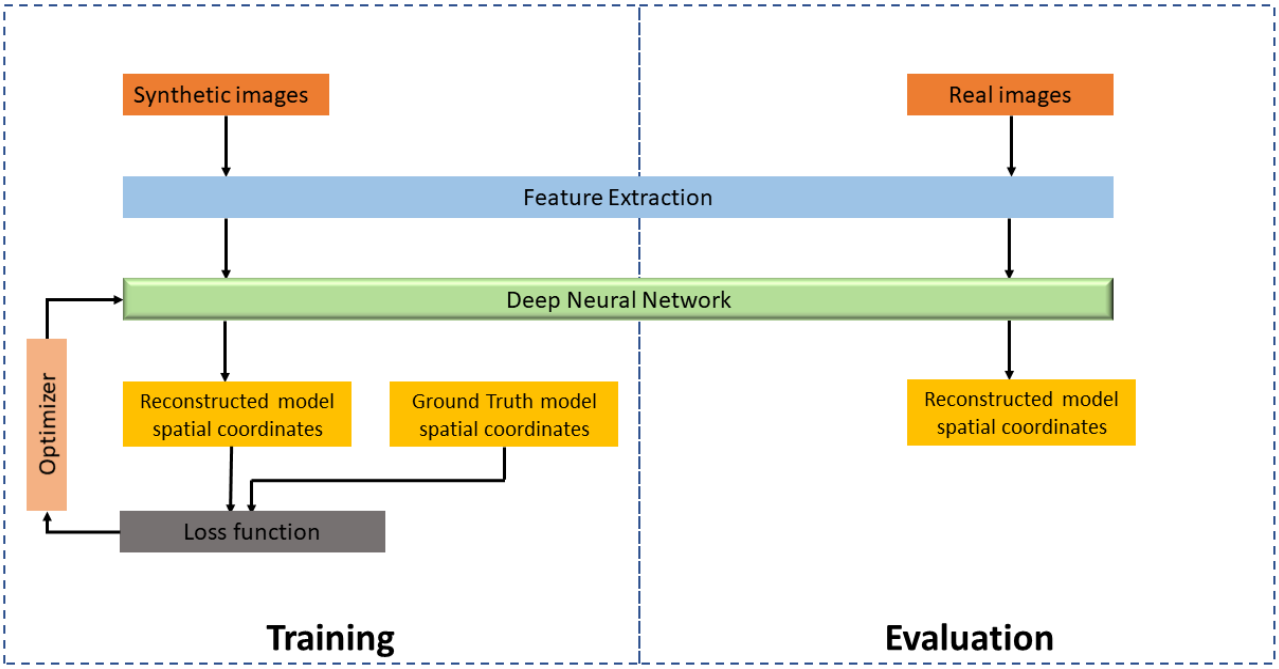


Figure 3.3: Overview of the approach

3.2 About the Acquisition of Real Images

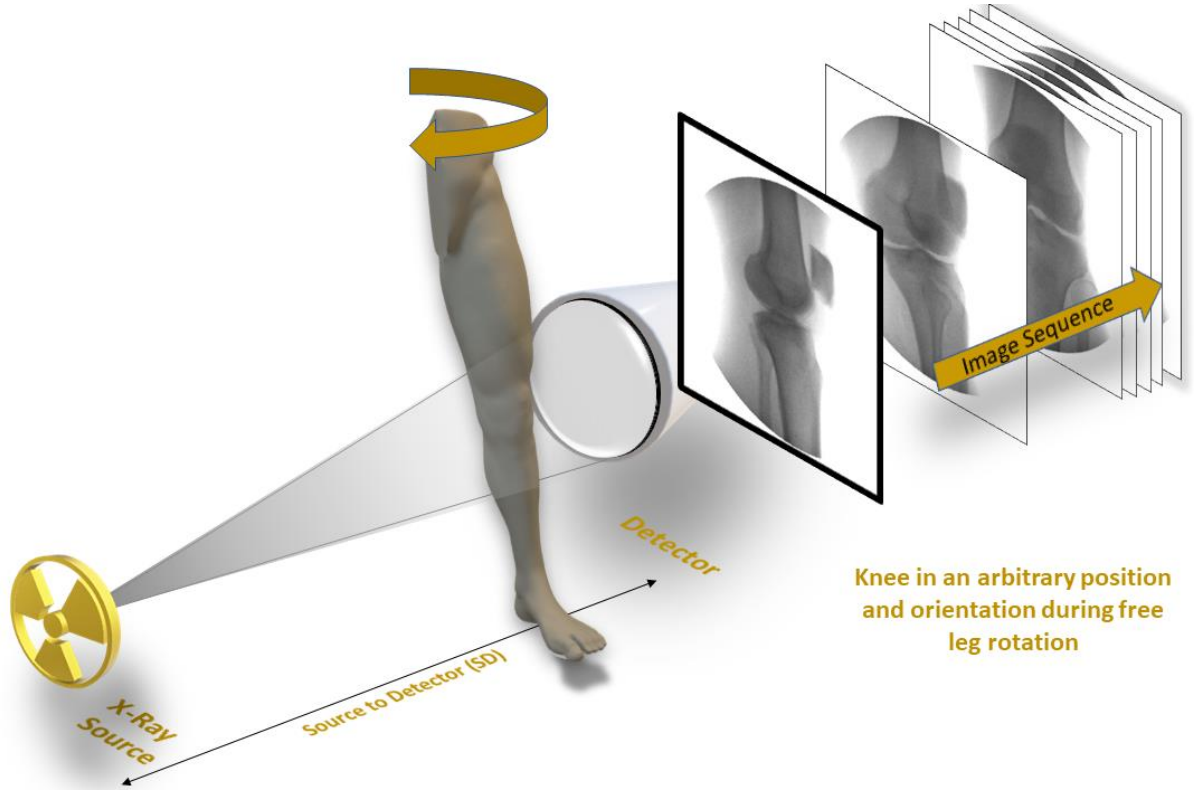


Figure 3.4: A general case of a real-time image acquisition setup using an X-ray fluoroscopy device. As the leg rotates freely around the tibial (mechanical) axis, an image sequence is formed.

This section describes the procedures involved in the real X-ray image acquisition system such that it can be replicated in the virtual environment to generate synthetic images.

A diverse collection of data is required to train a versatile network capable of reconstructing knee geometry under a range of circumstances. However, it is challenging to acquire a large training dataset of annotated X-ray images with variations indicative of those found in real-world scenarios. It is time-consuming and costly, especially given the variability of uncontrolled environments. An example would be the arbitrary variations in the orientation and location of a subject's knee in relation to the X-ray source's reference frame. Instead of acquiring large sets of X-ray images, the convolutional neural network (CNN) training and test sets were created using synthetic images

combined with ground-truth shape details. The technique of massive image creation (hundreds of thousands of synthetic images) with huge randomization of input parameters was used rather than waste a lot of money and time acquiring real X-ray images.

However, to determine the input parameters of synthetic image generation, it is necessary to understand the processes involved in the real system of X-ray image acquisition. The process would begin with a subject with an unknown knee geometry placing his/her leg at an arbitrary distance from the X-ray source. The orientation and position of the knee in the X-ray beam field would also be arbitrary and could keep varying during the activity. The subject would then freely rotate the leg for about 180º around the tibial (mechanical) axis by pivoting his/her heels on the ground. The fluoroscopy device would capture this activity, which would generate an image sequence of the knee in different poses. The process is tried to capture in Figure 3.4.

The parameters that influence the content of the X-ray image generation can be broadly categorized into three types:

1. **Shape Parameters**: Parameters that quantitively reflect the shape of the femur, tibia, patella, and the relative arrangement.
2. **Pose parameters**: Extrinsic rotations (Euler angles) of the knee measured with respect to a general fixed coordinate system. The Euler angles are typically denoted as α, β and γ.
3. **Scene parameters**: 3D distances of the knee's local coordinate system measured from the X-ray source's reference frame. SD, the distance between the source and the detector plane, is also considered as a parameter (see Figure 3.4), although it remains fixed for a given device.

As discussed earlier, the deep neural network was fully trained on synthetic data. To determine the network's performance against real X-ray images, it was tested against a publicly available dataset [155]

on Zenodo.org. The datasets include the fluoroscopic image sequences of ten subjects captured during the activity of the free leg rotation. Selected poses of subject 1 are shown in Figure 3.5. In addition, corresponding high-resolution MRI images of the knees are also included. The MRI images were downloaded and segmented to construct three-dimensional models that serve as a ground truth. More details on MRI segmentation are discussed in Section 4.1

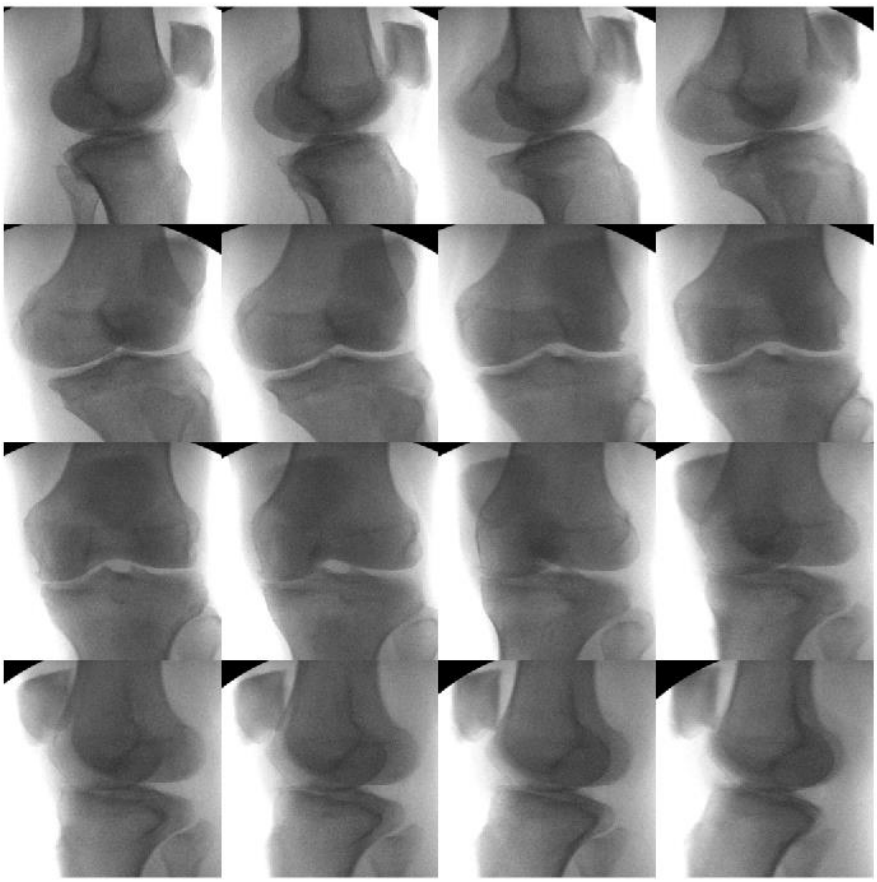


Figure 3.5: Some of the poses captured during the free leg rotation of subject 1.

3.3 Generation of Synthetic Images

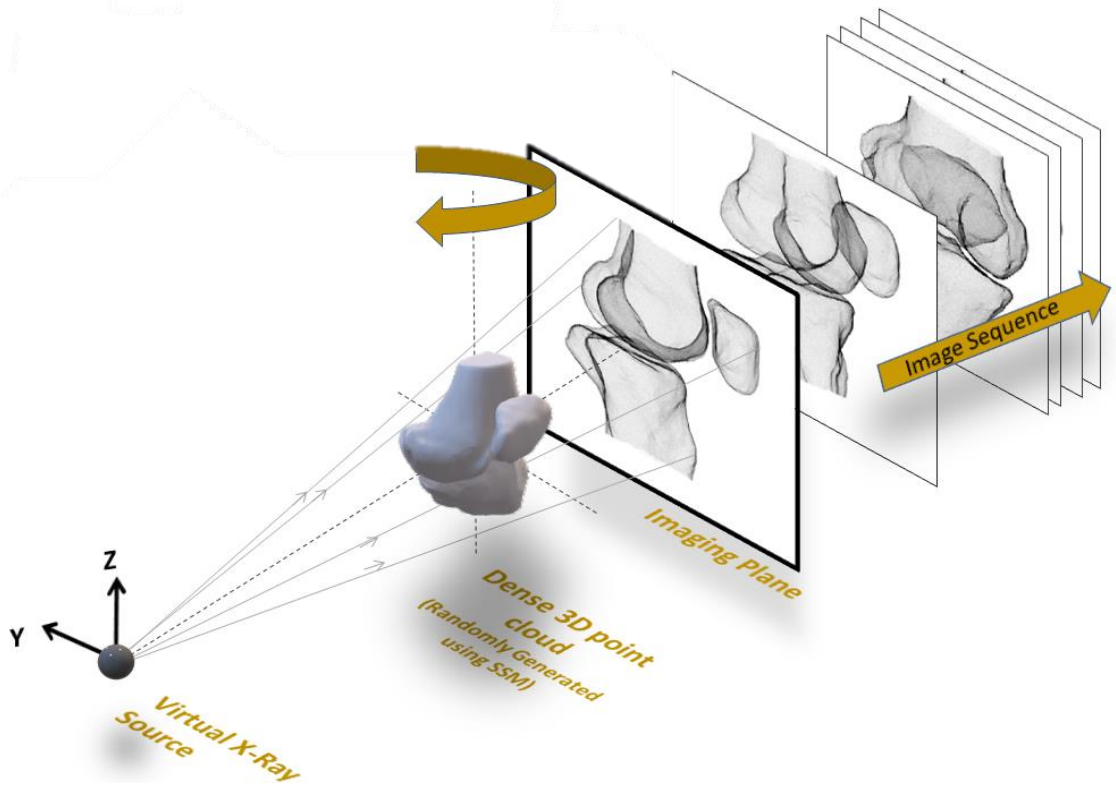


Figure 3.6: Setup of a virtual environment. The dense 3D point cloud is projected onto the plane using a perspective projection scheme. The randomly created model can be translated and rotated to any pose in the virtual environment to generate a corresponding two-dimensional image.

The simplest way to avoid problems with content inconsistencies between real and synthetic images is to replicate real processes as closely as possible. This Section will address the steps taken during the research to ensure that the content in the simulated image used for training is as accurate as possible.

One method for learning with only synthetic data is to produce training examples with a lot of variation to enable the network to learn invariances. FlowNet[156], which uses highly modified and unrealistic composite images between real images and rendered objects, is likely the first application of such an approach with CNNs. Despite the fact that our proposed research is in a completely different

area, a similar strategy to generate high variety by massively randomizing key input parameters (pose and scene parameters) is used, as discussed in the previous Section.

3D Model Creation and Placement

A new model can be developed by randomly varying the shape parameters in a statistical shape model (SSM). This study uses the SSM created by the University of Denver (Sec **Error! Reference source not found.**) to create all the synthetic models required for the training. The SSM was created using the distal femur, proximal tibia, patella bone, and associated articular surface cartilages from 50 healthy knees (25 males, 25 females). Select soft tissue structures' attachment sites and line-of-actions were also included. Each bone's local coordinate system is described as triads with 1D beam components. The joint is aligned in the MRI as-scanned spot. R3D3 tri elements are used to model bones as 2D surface meshes (Figure 2.3). There were 2384, 1101, and 472 nodes for femur, tibia and patella in the training set, respectively.

Every new model would be some linear deformation of the template mesh used to create the SSM. The SSM was created using Principal Component Analysis (PCA), so theoretically, an infinite number of unique models can be generated in the PCA space. Since the SSM was created using 50 knee training models, any new model can be represented by

$$S_0 = \bar{S} + P\sigma \tag{2}$$

where S_0 is a specific 3D model, $\bar{S}$ is the mean model, P is the principal orthogonal components, and σ is a set of 50 randomly generated shape parameters. To place the newly generated model in a particular position and orientation, transformation T is used such that:

$$S = T(S_0) = RS_0 + t \tag{3}$$

where R is a rotation matrix $R = R_\gamma * R_\beta * R_\alpha$, $R_\alpha = \begin{bmatrix} 1 & 0 & 1 \\ 0 & \cos\alpha & -\sin\alpha \\ 1 & \sin\alpha & \cos\alpha \end{bmatrix}$ $R_\beta = \begin{bmatrix} \cos\beta & 0 & \sin\beta \\ 0 & 1 & 0 \\ -\sin\beta & 0 & \cos\beta \end{bmatrix}, R_\gamma = \begin{bmatrix} \cos\gamma & -\sin\gamma & 0 \\ \sin\gamma & \cos\gamma & 0 \\ 0 & 0 & 1 \end{bmatrix}$ and t is a translation vector, $t = \begin{bmatrix} x \\ y \\ z \end{bmatrix}$

Along with the 50 shape parameters, the three pose parameters (α, β and γ) and the three scene parameters (x,y,z) are randomly varied to generate each synthetic image according to the details shown in Figure 3.8. The range's maximum and minimum values reflect highly improbable conditions in the real-world scenario. For example, a shape parameter of the value of 3.0 would likely result in an unrealistic knee model.

Setup of the Virtual Environment

A virtual environment was needed to generate large datasets of synthetic images belonging to diverse models in various poses. The fluoroscopic system was modelled after a pinhole camera in the virtual environment, with the imaging plane at the image intensifier and the focal point (origin) at the X-ray source (Figure 3.7). Using the techniques discussed in the previous Section, every randomly generated model was randomly placed in the virtual environment to get projected onto the imaging plane. Let $S \in \Re^3$ be the point cloud of the 3D model and let $X = [\acute{x}, \acute{y}, \acute{z}], X \in S$ be the spatial coordinates measured from the origin. The projection onto the 2D plane can be expressed as

$$M_{2d} = P * X \tag{4}$$

where M_{2d} is the 2D coordinate array of the 3D model after perspective projection. P is a perspective projection matrix of the shape (2x3) and a function of the focal length (SD in Figure 3.7). The technique to rendering the project points M_{2d} as a digital image is discussed in the next subsection.

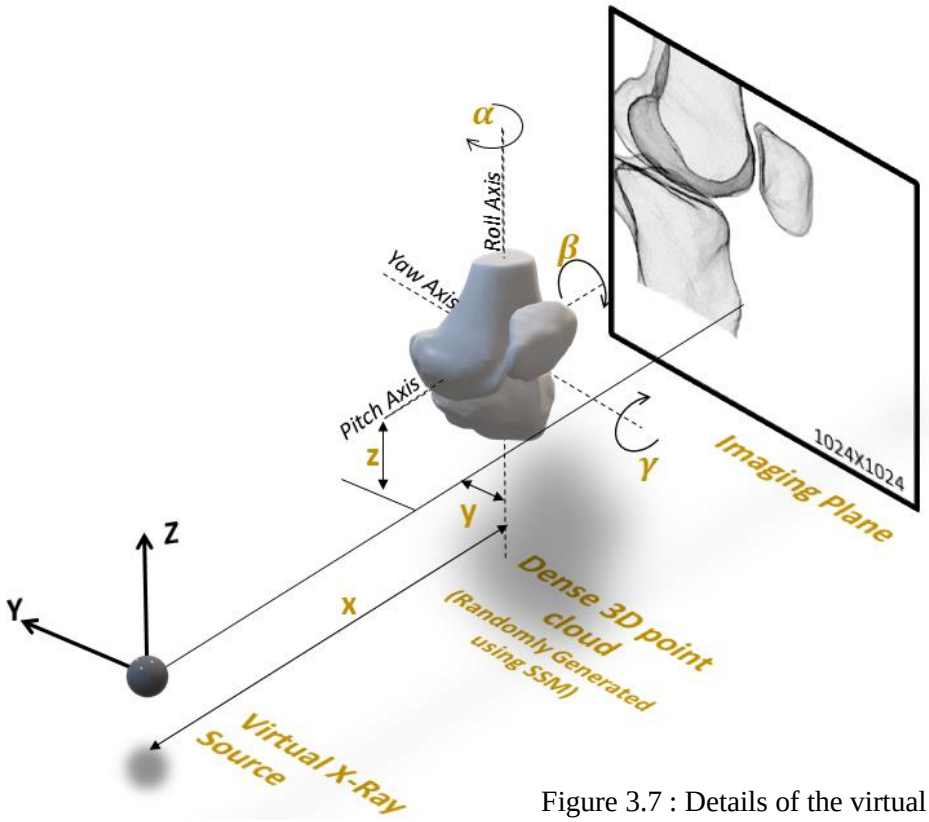
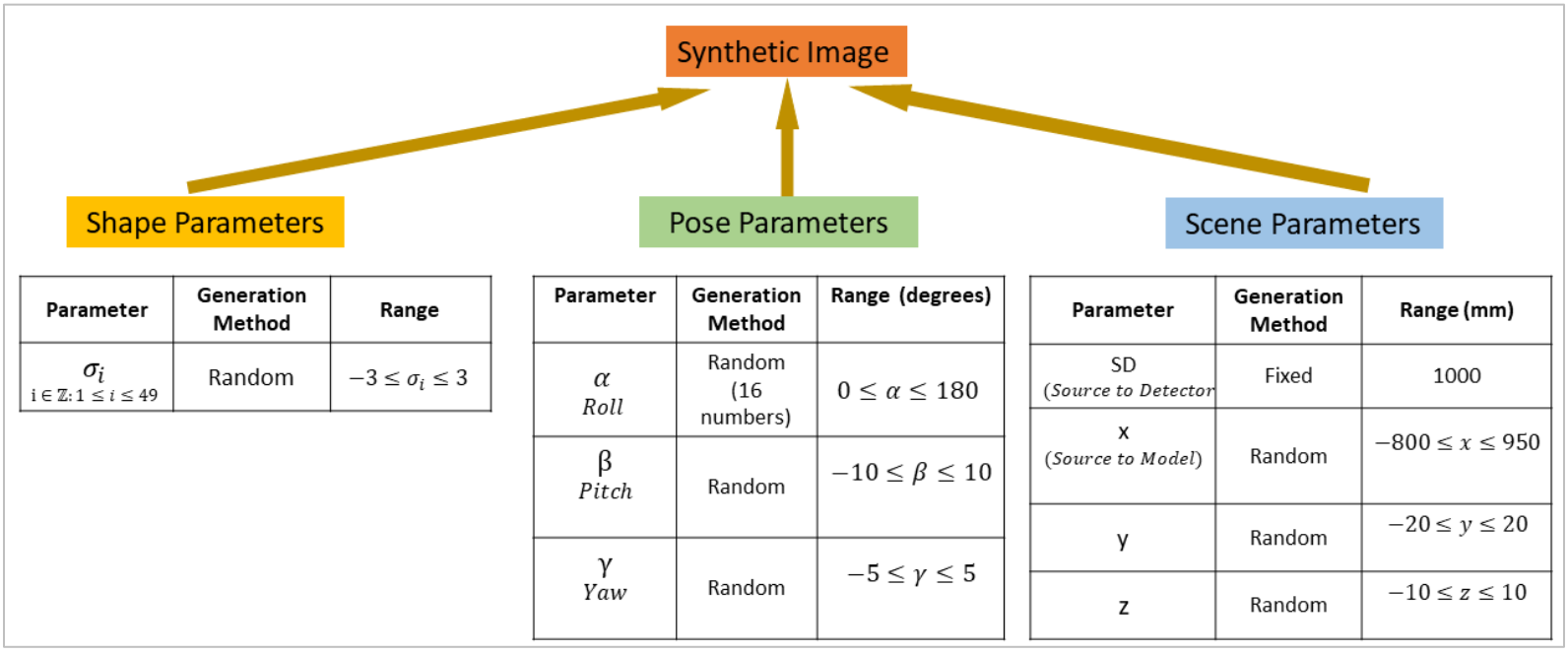


Figure 3.7 : Details of the virtual environment



Parameter	Generation Method	Range
σ_i $i \in \mathbb{Z}: 1 \leq i \leq 49$	Random	$-3 \leq \sigma_i \leq 3$

Parameter	Generation Method	Range (degrees)
α *Roll*	Random (16 numbers)	$0 \leq \alpha \leq 180$
β *Pitch*	Random	$-10 \leq \beta \leq 10$
γ *Yaw*	Random	$-5 \leq \gamma \leq 5$

Parameter	Generation Method	Range (mm)
SD (*Source to Detector*)	Fixed	1000
x (*Source to Model*)	Random	$-800 \leq x \leq 950$
y	Random	$-20 \leq y \leq 20$
z	Random	$-10 \leq z \leq 10$

Figure 3.8: Recipe to generate a synthetic image.

Image Rendering

A smoother and well-detailed image can be generated by projecting a denser 3D point cloud onto the imaging plane. The denser point cloud can be created through random sampling of the mesh points. In this study, the knee bone model was sampled to 3 million points. To capture the projected points, S_{2d} (Eq. 5) into an image of size 1024x1024, a 2D histogram with a million (1024^2) bins was employed. A 2D histogram, also known as a density heatmap, is a 2-dimensional generalization of a histogram that is computed by grouping a collection of points defined by their x and y coordinates into bins and applying a sum or count function to compute the colour of the tile representing the bin and thus creating an image.

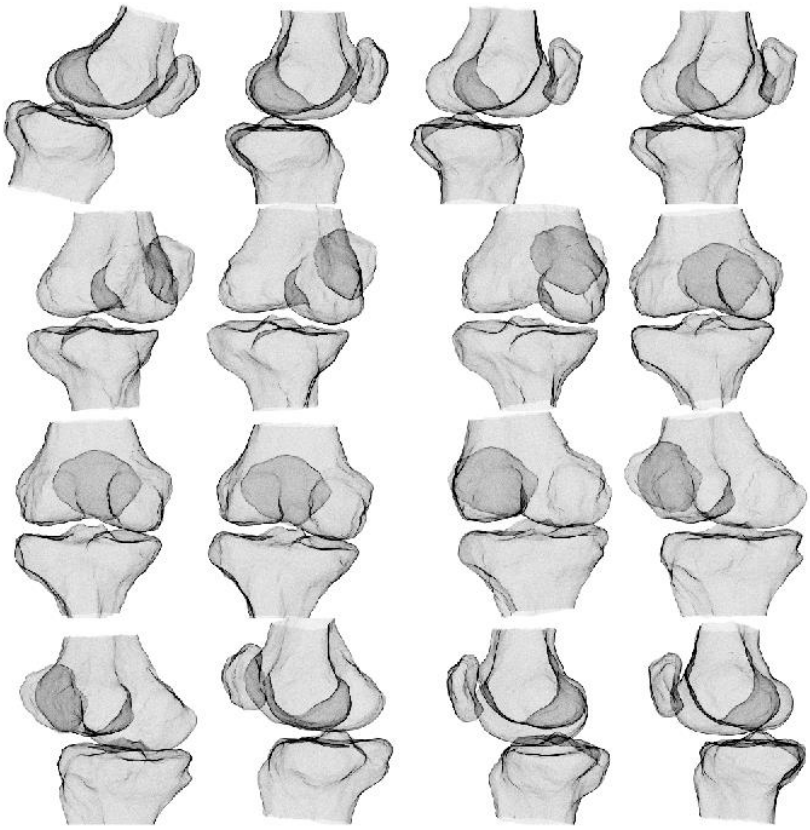


Figure 3.9: Random poses of a randomly generated model using the SSM. The poses are arranged in the increasing order of roll angle α

3.4 Feature Extraction

A feature in computer vision is a quantifiable piece of data in an image that is unique to a particular entity. It may be a specific colour in a picture or a particular form in the image, such as a line, an edge, or an image fragment. A useful feature is one that allows artifacts to be distinguished from one another.

Usually, input images contain an excessive amount of detail that is not needed for learning. As a result, the first move after preprocessing the image is to simplify it by extracting critical details and discarding irrelevant ones. The essential features of an X-ray image required in this research are the ones that depict the shape of the femur, tibia and patella and their relative arrangement, while the undesirable features are the soft-tissue artifacts and the quantum noise [55].

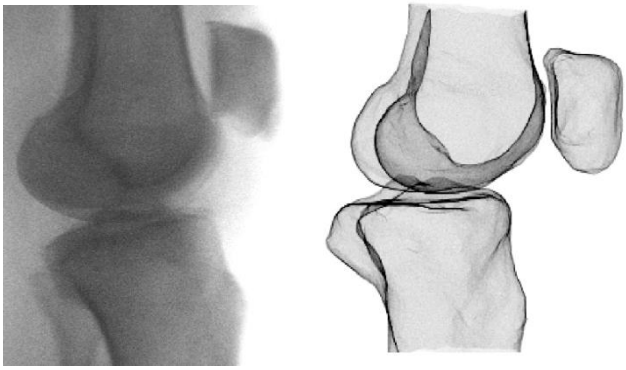


Figure 3.10: Different styles of real and (randomly generated) synthetic image

As seen in Figure 3.7, the synthetic image and real X-ray image have distinct styles. It is troublesome because, in order for a deep neural network to perform efficiently, the training and validation data sets must share a common style domain. To resolve this problem, a novel technique for extracting desirable features from images independent of the style domain they were produced. The domain gap between the synthetic and actual X-ray images is tried to eliminate by feeding the network only the feature extracted data. Though Canny's edge operator [92] is a natural choice for feature

extraction and generally accurate, its efficiency suffers due to the high degree of noise present in X-ray fluoroscopy images. As shown in Figure 3.11, the edges are not continuous, and a considerable noise is captured as false edges. Varying the threshold and fine-tuning parameters makes little difference since the Signal to Noise ratio is inherently higher [55].

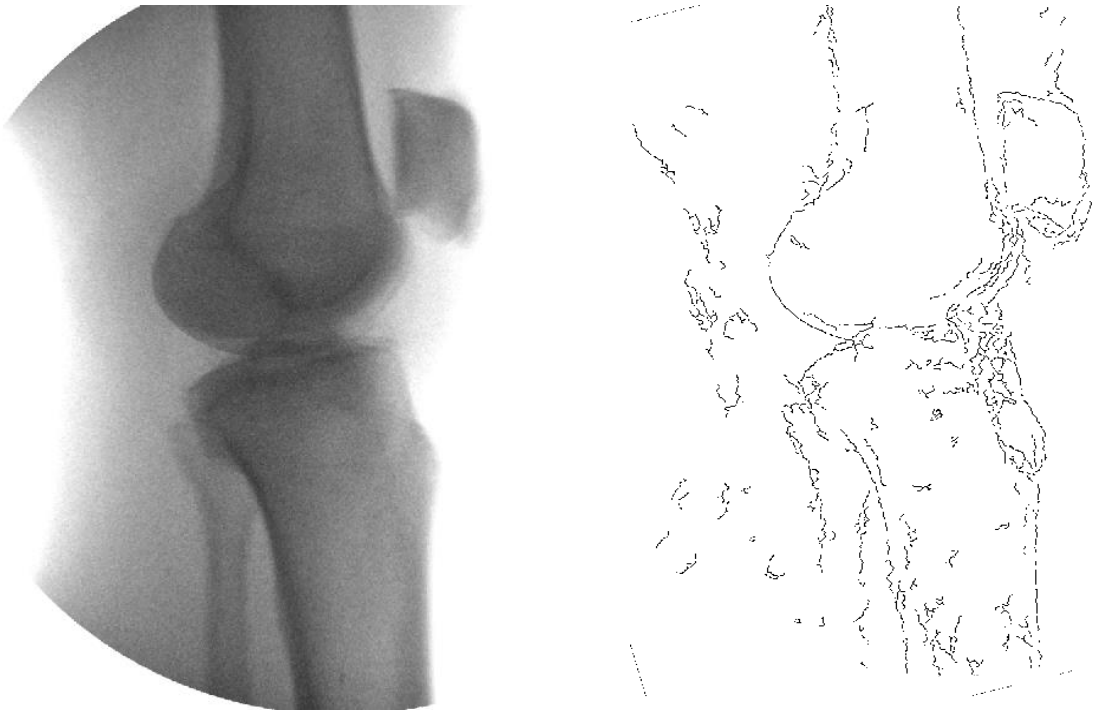


Figure 3.11: Poor performance of Canny' edge detection on a fluoroscopic X-ray image

Multi-layer Canny Operator

Humans seem to interpret edges easily, although it is usually very challenging for an edge detection algorithm. One explanation is that human perception has global knowledge, while the edge operator only has information found within a 3x3 or 5x5 neighbourhood. The distinction is

perhaps better understood by imagining a mouse trapped in a labyrinth and a bird gazing down from the clouds. A new robust technique is presented to accurately extract edges by combining the concept of the image pyramid [153] and Canny's edge operator algorithm [92] to blend global and local knowledge to detect features while still suppressing the noise. Figure 3.12 illustrates the multi-layer canny algorithm's workflow, whereas Figure 3.13 demonstrates how, despite a large domain gap at the input, the algorithm produces identical styled images.

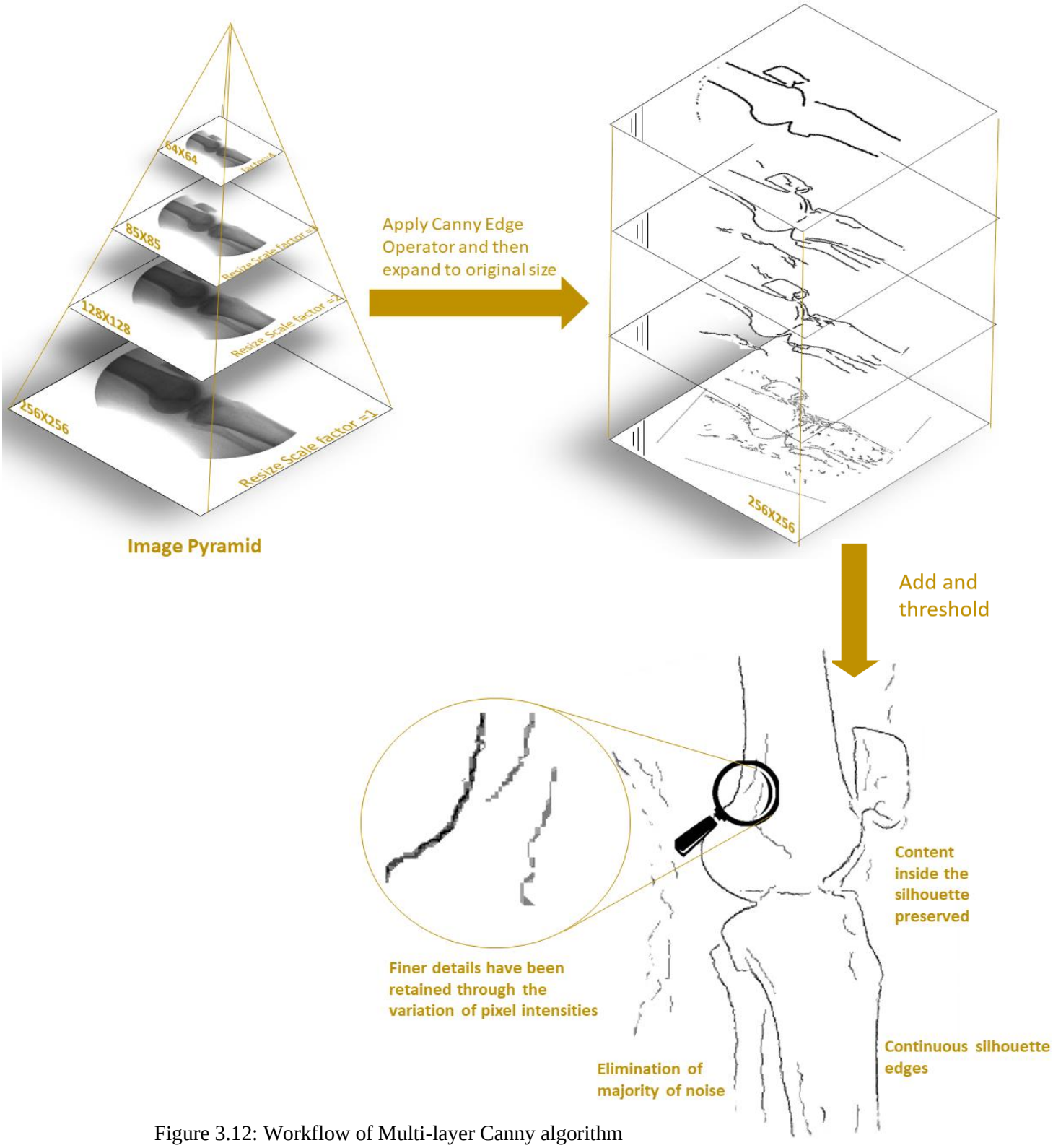


Figure 3.12: Workflow of Multi-layer Canny algorithm

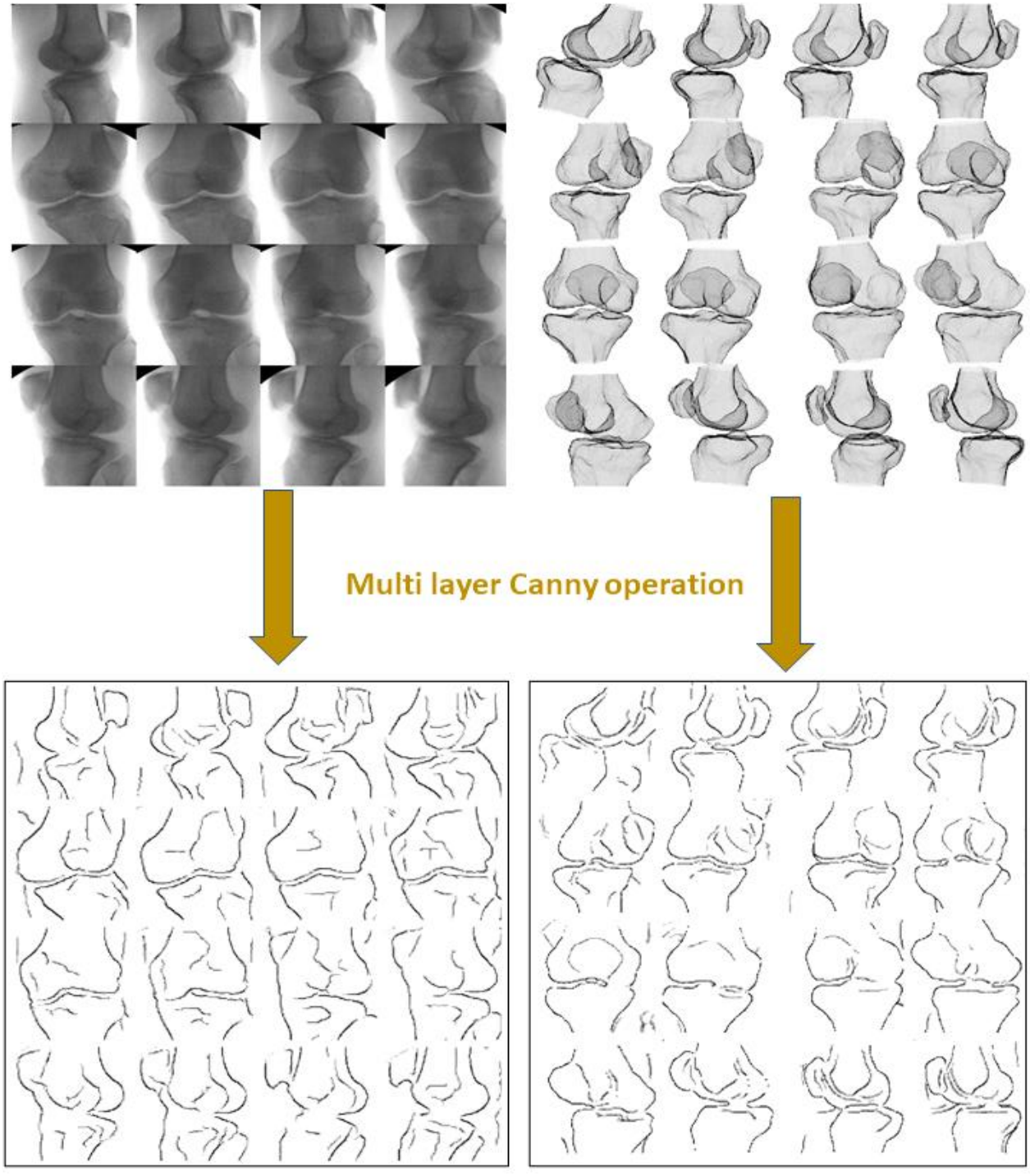


Figure 3.13: The feature-extracted real and (randomly generated) synthetic images are stylistically similar

3.5 Training of Custom-Built CNN

Since a CNN needs a vast volume of annotated training data and careful hyperparameter tuning, training from scratch is usually difficult and time-consuming. To generalize well and prevent overfitting, the learning rates used during training should ensure sufficient convergence. Transfer learning is an alternative to learning from scratch that involves fine-tuning CNNs that have already been trained in another domain to a target domain, such as the medical one. However, significant variations between the source and target domains can restrict information transfer, potentially reducing the efficiency of fine-tuned CNNs. Since our featured-extracted synthetic images are uniquely styled and structured, any pre-trained CNN existing in the literature is unlikely to be suitable for training. However, with enough labelled training data and carefully chosen hyper-parameters, custom-built CNNs can outperform fine-tuned CNNs and hand-crafted alternatives [57], [63]. Custom-built CNNs are extremely effective in a variety of medical applications. Some of the recent studies that use custom-built CNNs for knee image analysis are multi-stream CNNs [157] for cartilage segmentation, real-time 2D/3D registration using CNN regressors [158] for volar plate detection and femur bone age assessment in X-ray images using deep learning [159]. Motivated by these methods, our CNN is trained from the ground up to reconstruct the 3D knee bone geometry from multiple X-ray image poses.

A feature-extracted synthetic image containing 16 poses of a knee geometry organized in ascending order of roll angle was fed into the network as the input. Each pose has a resolution of 256x256 pixels, resulting in a 1024x1024 total image size (Figure 3.14). The network produces a 12576-long vector as its output. In the following step, the long vector can be reshaped into an array of spatial coordinates for the vertices of the knee joint model. This architecture is constructed so that the vertices are implicitly constrained to conform to the vertices of the reference mesh used to create the

statistical form model (SSM). As a result, the output of the network is simply the vertices of a predetermined mesh.

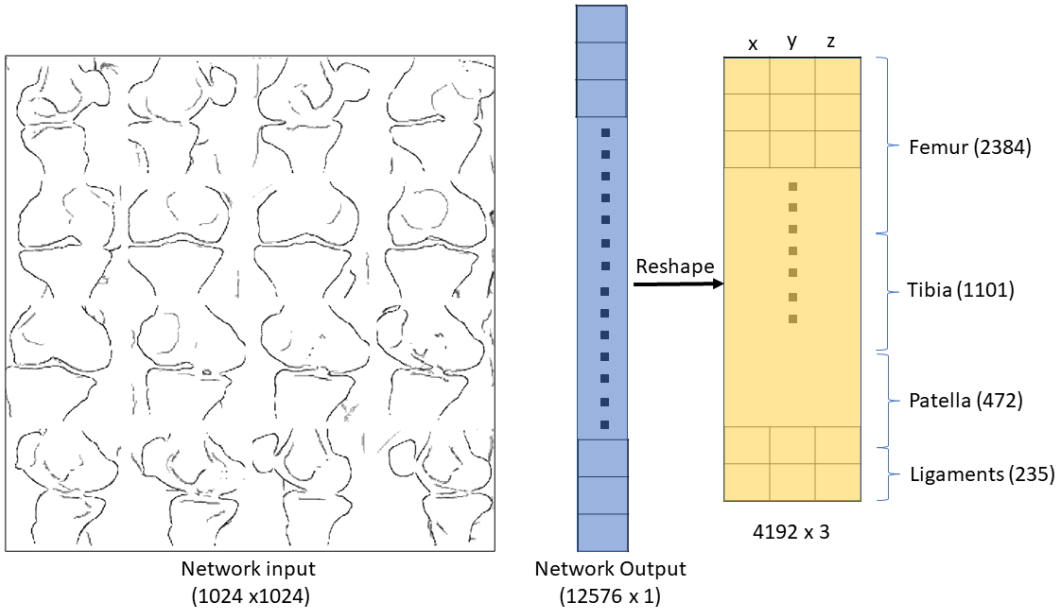


Figure 3.14: The network's input and output

3.5.1 Architecture

A 23 layer deep ResNet [75] inspired CNN architecture was custom built for the task of reconstruction. The details of the architecture are shown in Figure 3.15. Batch normalization and a ReLU activation layer accompanied each convolutional layer. Every third skip connection used a max-pooling operation. Two fully connected layers with a 12576-long vector output followed the final pooling layer. A drop-out layer with a drop-out ratio of 0.1 was added for the fully connected layer to avoid overfitting.

Due to the GPU's limitations, a batch size of 16 was used during training. The network was trained using Google Colab resources (GPU: NvidiaK80/ T4 with 12 GB of GPU memory). The massive datasets were uploaded to the online servers through Google Drive. The programs were written in python, and pytorch[71] was used as the machine learning framework. Wandb [160]was one of many APIs used during the coding that had a major impact on the results Wandb is a machine learning tracking tool. Integrating the Wandb library into every level of the code greatly aided in consolidating and tracking results.

Using the approach described in section 3.3, a set of 30,000 images were produced from 10,000 uniquely generated models from the SSM for training the network. Another 600 images were created to serve as a test set. The network was trained for 50 epochs for 24 hours using the Adam optimizer with a learning rate of 0.04.

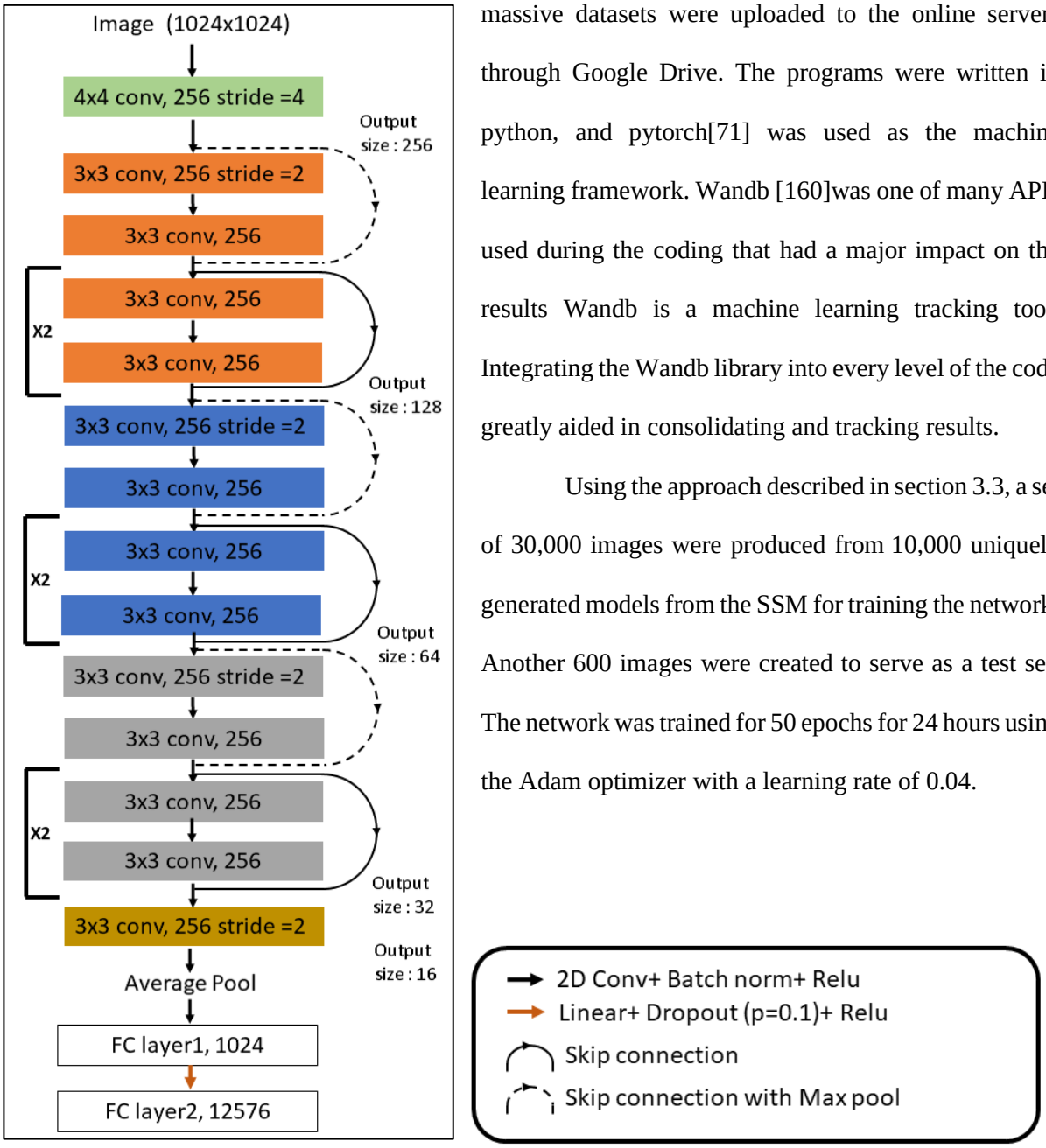


Figure 3.15: Network architecture

3.5.2 Loss Function

At first glance, it might look appropriate to use a regression metric such as Mean Squared Error (MSE) (Figure 3.16 (a)) as a loss function since the objective of the neural network is to minimize the difference between the 12576-long vector output and the ground truth. However, it should be noted that the images (real or synthetic) are generated by perspective projections of 3D geometry and therefore lack scale detail. Additionally, the orientation and location of the reconstructed model and the ground model do not have to be similar. Taking these points into account, it was determined that the Procrustes shape distance (Figure 3.16 (b)) is more appropriate for the task. Given that the output and ground truth models' vertices are implicitly constrained to be in exact correspondence, the Procrustes shape distance seems to be a reasonable preference.

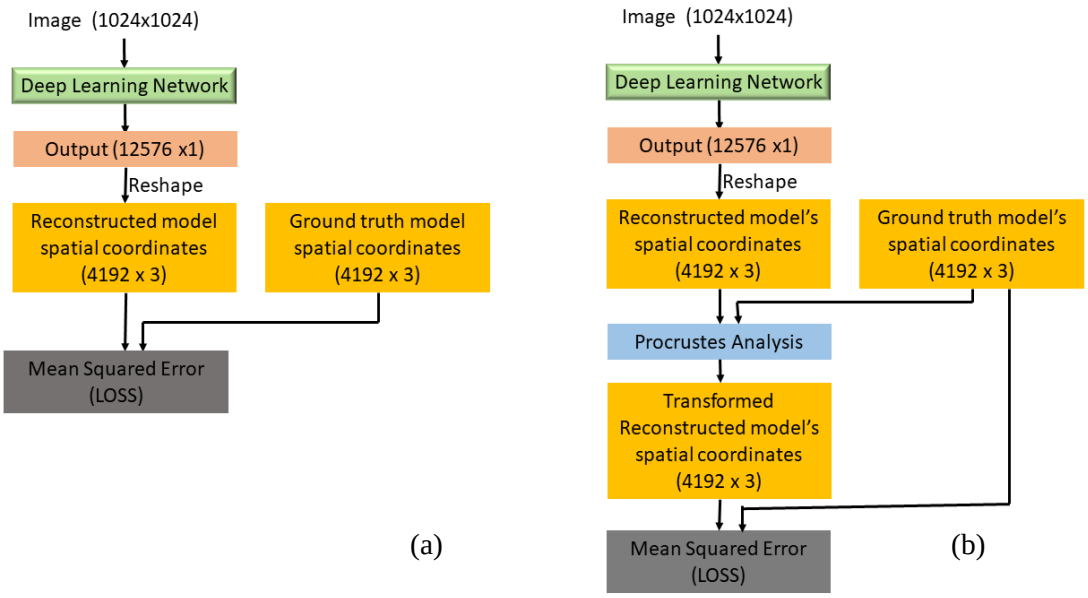


Figure 3.16: Loss Functions: (a) Mean Squared Error Loss. (b) Procrustes shape distance Loss

To compare MSE loss to the Procrustes shape distance metric, the network was trained for 20 epochs in each case. Figure 3.17 compares the two loss functions by plotting the loss values for the test cases against the epochs. The Procrustes shape distance metric outperforms MSE by a large margin.

Figure 3.17: Comparison of two loss functions

Chapter 4 Evaluation

The preceding chapter covered the process of training the custom-built CNN for 3D reconstruction. This chapter addresses the method used for evaluating the accuracy of a CNN's reconstruction. The accuracy of the current method is compared and debated with the existing methods.

4.1 Dataset

The neural network, which was only trained on synthetic images, was validated against a publicly accessible dataset [155] on Zenodo.org to see how well it performed against actual X-ray images. The databases include fluoroscopic image sequences of ten subjects taken during the free leg rotation activity. Additionally, the study included accompanying research-grade Magnetic Resonance (MR) images of the knees. The MR images were downloaded and segmented to build three-dimensional ground truth models. Only eight datasets were considered for the model's validation since two (Subject 8 and 9) of the ten subjects' MRI scans were unsuitable for segmentation due to consistency issues. The appendix comprises a selection of X-ray images from each of the eight subjects.

4.1.1 MRI Segmentation: Ground Truth Creation

As discussed earlier, the ground truth for each dataset was created by segmenting the MR images using 3D slicer software (version 4.11). The method of segmenting MR images in 3D Slicer is depicted in Figure 4.1.

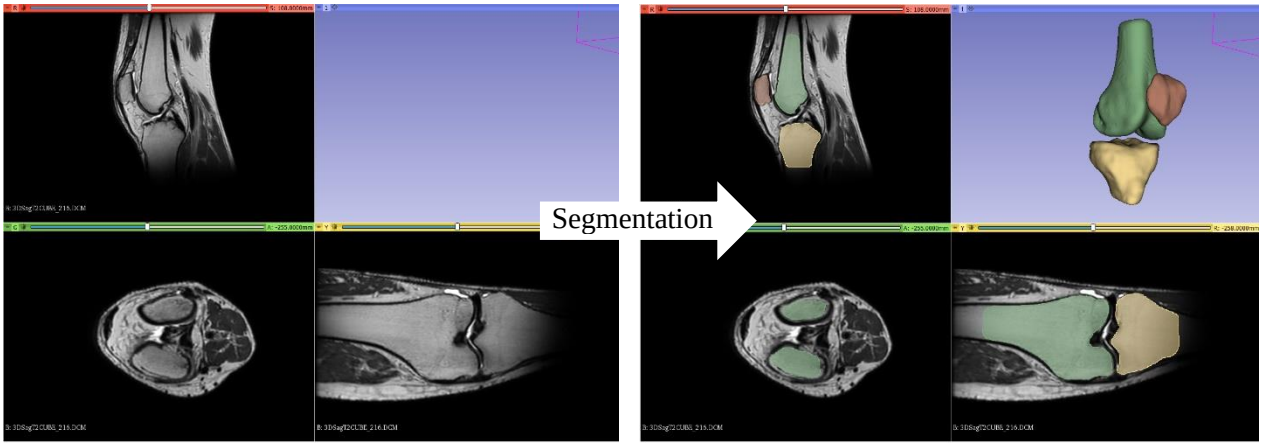


Figure 4.1: Segmentation process of bones in 3D Slicer

The geometry created by 3D Slicer was coarse and contained jagged edges referred to as staircase artifacts, a common characteristic of segmentation. Thus, prior to meshing, the surfaces were smoothed, and the geometries' borders were trimmed using the Surface Wizard and Curve Wizard functions in SolidWorks 2018 (Dassault Systèmes, Waltham, MA, USA). Smoothing was done with caution to avoid distorting the bone's initial geometrical structure or contours, and only the surface defects were removed (Figure 4.2).

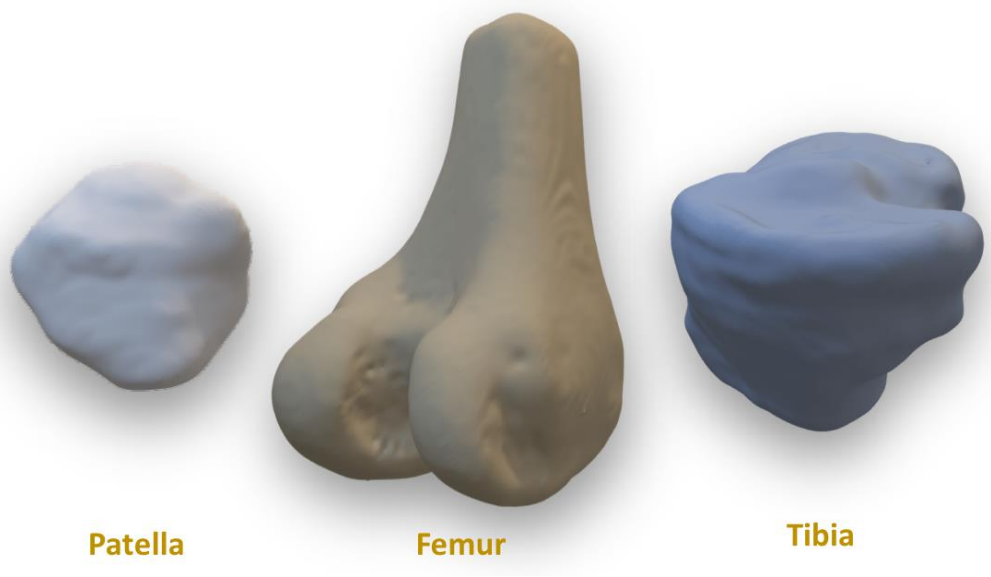


Figure 4.2: Models used as ground truth. These models were created by segmenting MR images

4.2 Validation

The 3D shape reconstructed from real X-ray images was compared with the 3D knee model obtained by MR image segmentation using the Iterative closest point (ICP) algorithm [161]. The two models were transformed to point clouds before being fed into the ICP algorithm. The ICP aligns the reconstructed model to the ground truth by adjusting the rotation and translation of the reconstructed model. After aligning the two models, the Root Mean Square Error (RMSE) for the complete knee and individual bones was determined. The validation protocol is depicted in Figure 4.3.

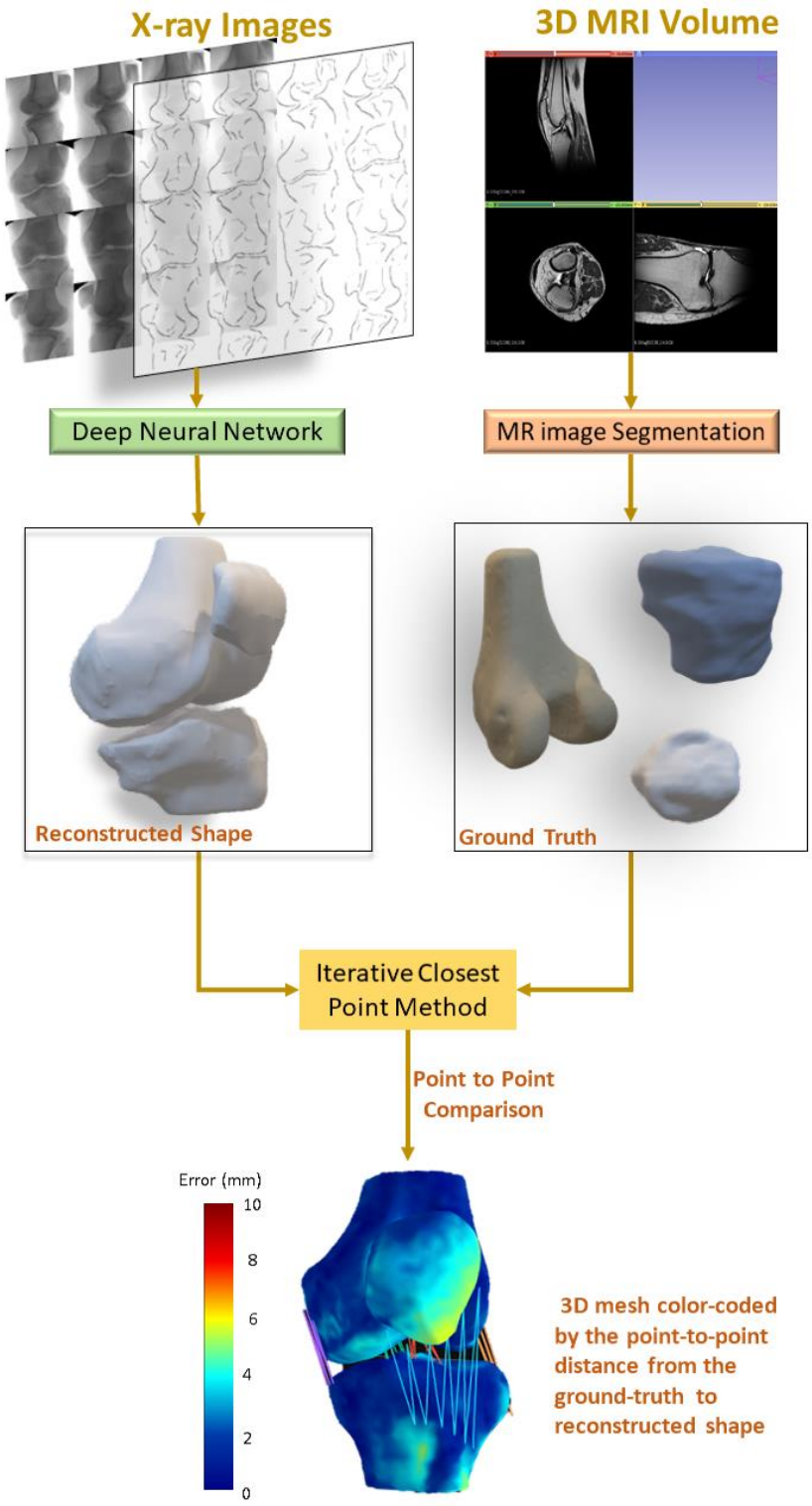


Figure 4.3: Validation protocol for real test cases

The validation protocol was followed to validate the neural network against eight datasets. The results are shown in Table 2: Root Mean Squared Error (mm) of eight subjects' bone The reconstructed model for subject 1 is shown in Figure 4.4. Similar results of the rest of the subjects are shown in the Appendix.

Table 2: Root Mean Squared Error (mm) of eight subjects' bone shapes.

SUBJECTS	**FEMUR**	**TIBIA**	**PATELLA**	**ALL**
SUBJECT 1	1.9	2.9	3.1	2.3
SUBJECT 2	1.7	2.7	4.2	2.5
SUBJECT 3	1.5	2.1	2.5	1.8
SUBJECT 4	1.7	1.9	3.3	1.9
SUBJECT 5	2.4	2.2	3.1	2.5
SUBJECT 6	2.0	2.5	3.6	2.3
SUBJECT 7	2.3	2.7	3.0	2.5
SUBJECT 10	1.8	1.9	4.1	2.1
MEAN	**1.9**	**2.3**	**3.3**	**2.2**

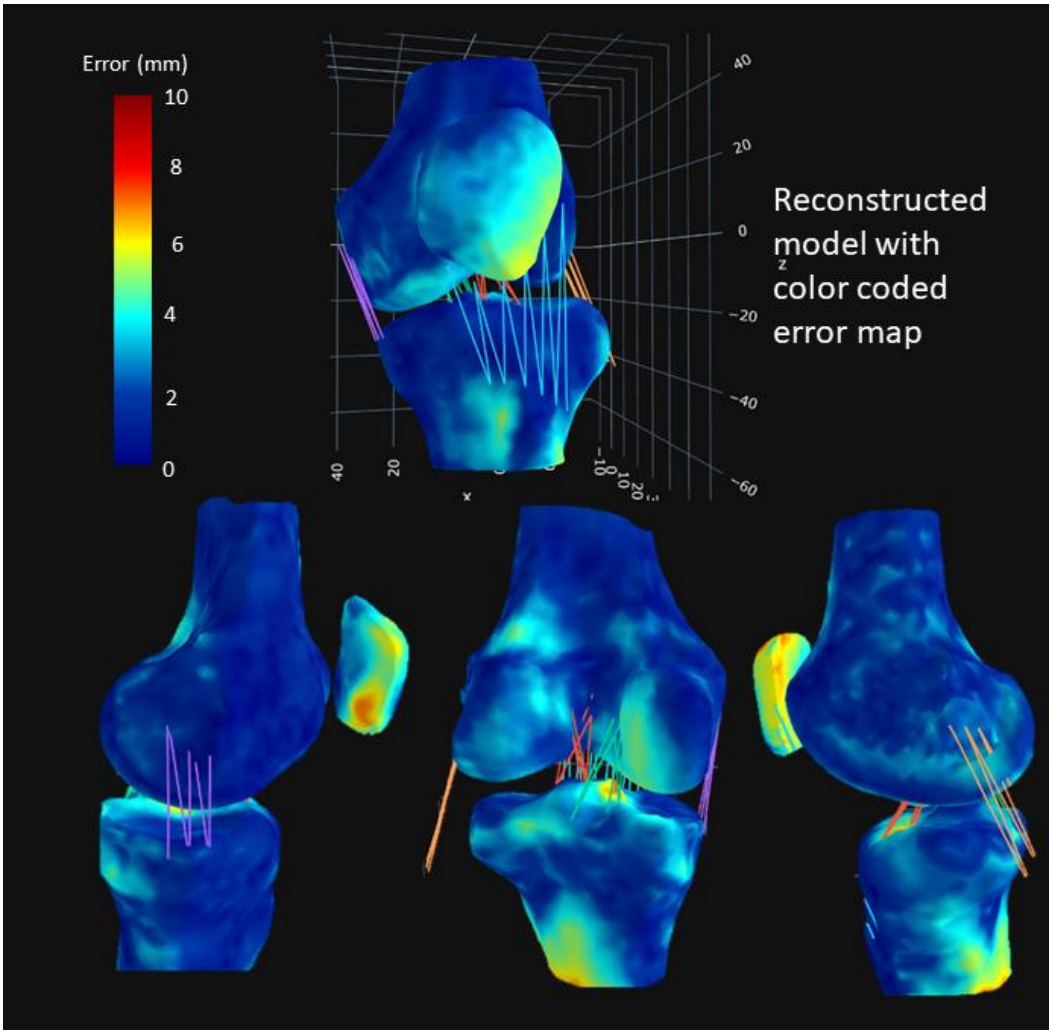


Figure 4.4: Reconstructed knee model of subject 1

4.3 Discussion

An automated procedure for reconstructing the femur, tibia, and patella from single-plane fluoroscopic X-ray sequences without the need for additional 3-D imaging acquisition is demonstrated. The three-dimensional reconstruction was accomplished using a deep neural network. Reconstruction findings for eight subjects performing free leg rotation sequences are presented to evaluate the accuracy

of the proposed method. The results were validated by comparing the reconstructed model with the gold standard MRI segmented model.

The hypothesis was proven correct- it is possible to reconstruct the 3D shape of the femur, tibia, and patella from multiple uncalibrated SF images featuring random poses of the knee taken during a subject's free leg rotation. According to Table 2, the proposed deep neural network has an overall accuracy of 2.2 mm with 1.9 mm, 2.3 mm, and 3.3 mm for the femur, tibia, and patella bones, respectively. Based on the discussion in section 2.5.3, the reported accuracy might be sufficient for a precise evaluation of knee kinematics, although a detailed investigation is required.

The errors in knee bone 3-D reconstruction can be seen in detail through the colour-coded error maps presented in Appendix. The surface errors are evenly distributed in the geometry and are not location-specific. On the other hand, the patella has the largest error since its appearance rate in the X-ray image sequence is significantly smaller than the other two bones.

4.3.1 Comparison to Related Works

Table 3 compares the research performance of different existing methods to the 3D bone reconstruction from single-plane fluoroscopy (SF) images. While the accuracy of the proposed method is marginally lower than that of the current literature, the method is more adaptable than others. It appears that the proposed method is the first method to reconstruct the 3D shape of the knee using deep learning, while the rest of the methods mostly focus on some deformation of the statistical shape model (SSM). As a result, the computation time and the user intervention for the 3D reconstruction of the proposed method are very small. The proposed CNN was trained and tested on Google Colab Also, the proposed method is the only method that does not require calibrated images for the 3D reconstruction.

Table 3: Comparison to related works

Study	**Reconstruction error (mm)**	**Test cases**	**Is calibration required?**	**Is reconstruction automated?**	**Reconstruction time/test case**
Wu **[12]**	Femur: 1.15 Tibia: 1.19	Five knee joints	Yes	No	20 min
Kurazume et al. **[26]**	~1	Two proximal femurs	Yes	No	1 min
Zheng et al.**[27]**	1.44	One distal femur (cadaveric)	Yes	No	Not reported
Tang et al.**[162]**	1.8	Two distal femurs	Yes	No	Not reported
Hurvitz et al. **[17]**	1.15	One distal femur (cadaveric)	Yes	No	Not reported
Current work	Femur:1.9 Tibia:2.3 Patella: 3.3	Eight knee joints	No	Yes	~1 sec

Generally, the existing methods aim to extrapolate a complete 3D anatomical surface from sparse or incomplete 3D information obtained from 2D radiographic images. This is done by deforming the SSM by optimizing the values of a handful of shape parameters. The deep learning model is trained to optimize every vertex of a 3D model (4192 vertices) in the proposed method to match the ground truth (known shape model). The model builds a mapping between each pixel of the input image (containing 16 poses) to the vertices of the reconstructed 3D model. Since it could learn the complex relationship between each vertex and the input image, it can be trained for more accurate reconstruction.

The main purpose of the proposed methodology is to quickly and accurately build the shape of the knee so that it could be used during the pose registration of the individual bones during the kinematic study. Nevertheless, the full-scale model could be generated by updating a reference dimension in the shape model. A reference dimension could be the femoral bicondylar width which can be measured from a pre-operative AP radiograph [163]. It should be noted that the accuracy of measurement of the reference dimension could affect the accuracy of the full-scale model.

4.3.2 Feeding hand-crafted features for the convolutional neural network

Feeding feature extracted data to a CNN is not uncommon. It is a common practice when the images are sourced from different domains. To reduce the domain shift (Section 2.4.1), researchers have either used domain adaptation technique [78], [80], [164]–[166] or image transformation technique [81]–[89]. In the proposed framework to unify the real and synthetic image into a common representative image, an image transformation technique was employed. The transformed image is still an implicit representation of the ground truth. Moreover, feature

extraction is necessary to deal with single-plane fluoroscopy (SF) images that suffer from low signal-to-noise ratio, low contrast resolution and signal non-homogeneity [11].

4.3.3 Systematic Errors

In the context of this thesis, systemic errors refer to inaccuracies that occur as a result of the virtual environment's inability to capture a particular phenomenon. The author could think of two systematic errors:

i. Relative movement between the knee bones during the free leg rotation: In the virtual environment, different poses are generated by rigidly rotating and translating the knee bones. As a result, the reconstruction accuracy could degrade if the subjects in the real world do not maintain rigidity of the knee during the free leg rotation. The subjects may be instructed to rotate their leg to address this error while keeping their knee in the locked position [167]. In the locked position, relative motion between the knee bones is restricted.

ii. Calibration errors: The calibration is performed for two purposes: to determine the intrinsic parameters and correct the image distortion. Using a theoretical framework, authors in the 2007 study [100] argue that the C-arm (single-plane fluoroscopy) calibration could be avoided in certain situations. The authors argue that the reconstructed shape, in particular, was found to be more stable even for large calibration errors (100 mm error in focal length) when a large number of images were considered. So, to nullify the effect of calibration error on the performance of the proposed framework, it was decided to use 16 images for the reconstruction with a generic set of calibration parameters.

4.4 Limitations

The present thesis has some acknowledged limitations, including the following:

1. Although the deep neural network might generate a highly nonlinear mapping between the input images and the output model, the network is trained on models produced from a single Statistical Shape Model (SSM), an intrinsically linear model [168]. SSMs are essentially low pass filters, and it is highly improbable that they can capture high-frequency details in X-ray images [169]. Since the network is trained exclusively on SSM models, every reconstructed model is mostly limited to some linear combination of the SSMs' training examples.

2. Since the SSM used to train the convolutional neural network lacks information about the fibula bone, the network is completely ignorant of the fibula. As a result, the fibula silhouettes had to be manually erased from the feature-extracted images during the test cases.

3. In addition to the spatial coordinates of the knee bones, the deep neural network also outputs predictions for ligament attachment sites. However, due to the low visibility of soft tissues in MRI scans, the ligament attachment point could not be segmented, and hence it could not be validated. Nonetheless, the predicted ligament attachment sites for all eight subjects could be seen in the images of the Appendix.

Chapter 5 Conclusions

5.1 Summary

This study aimed to accurately reconstruct the three-dimensional shape of knee bones by analyzing a small number of knee poses captured during a subject's free leg rotation using an uncalibrated fluoroscopy device. Since annotated X-ray images are scarce, the neural network had to be trained only on synthetic images. Since radiation dose settings differ between fluoroscopy devices and system operators, creating quality variability, the reconstruction output should be reliable across a wide range of image quality.

An end-to-end framework is presented to reconstruct the 3D shape of the knee geometry from low radiation X-ray images. First, the mechanisms involved in acquiring a real X-ray image are investigated. It is then repeated in a virtual environment to generate hundreds of thousands of randomized images. The present work develops a novel domain-invariant feature extraction strategy for extracting desirable features from simulated and real-world images of varying quality. Finally, a 23-layer deep custom-built convolutional neural network(CNN) is trained using the massive feature-extracted synthetic dataset. Real-world X-ray images are evaluated using the trained network. The method is adaptable enough to deal with a wide spectrum of poses and low-quality images, rendering it suitable for a broad range of applications. The findings show that the deep learning model is highly effective and produces good-quality reconstructions with a high degree of generalization.

5.2 Future Directions

Several steps could be taken to overcome the limitations listed in the previous section. Possible improvements to improve the overall performance of the framework is discussed below:

1. Since the Statistical shape model (SSM) is at the core of the presented 3D shape reconstruction framework, any improvements in the SSM could directly boost the performance of the CNN. The SSM could be improved in a variety of ways:
    - Inclusion of more and diverse training examples: The SSM used in this study is built using 50 training examples [44]. More training examples could result in capturing more shape variations.
    - Resolution enhancement: Currently, the SSM mesh is represented by 3957 nodes (2384 for the femur, 1101 for the tibia and 472 for the patella) [38]. Higher resolutions could result in capturing high-frequency details.
    - The PCA-based SSM could be reformulated as Gaussian process morphable models (GPMMs) [170] or Kernel-PCA (K-PCA)[171]. These nonlinear models are more effective at capturing shape variations in the training examples.
    - Inclusion of Fibula
2. The MRI images used to create the ground truth seem to be acquired in T2 sequences with low visibility of soft tissues. In the future, acquiring MRI scans in the T1 sequence can increase soft tissue visibility, allowing for identifying ligament attachment sites [172].

References

[1] A. Sharma, R. Komistek, and M. Mahfouz, “In Vivo Kinematics Evaluation in Flexion of Patients Implanted With Primary TKA,” *Techniques in Knee Surgery*, vol. 10, no. 2, pp. 66–72, Jun. 2011, doi: 10.1097/BTK.0b013e31821cabb2.

[2] D. A. Dennis, M. R. Mahfouz, R. D. Komistek, and W. Hoff, “In vivo determination of normal and anterior cruciate ligament-deficient knee kinematics,” *J Biomech*, vol. 38, no. 2, pp. 241–253, Feb. 2005, doi: 10.1016/j.jbiomech.2004.02.042.

[3] C. Li, A. Hosseini, T.-Y. Tsai, Y.-M. Kwon, and G. Li, “Articular contact kinematics of the knee before and after a cruciate retaining total knee arthroplasty,” *J Orthop Res*, vol. 33, no. 3, pp. 349–358, Mar. 2015, doi: 10.1002/jor.22764.

[4] V. Camomilla, R. Dumas, and A. Cappozzo, “Human movement analysis: The soft tissue artefact issue,” *J Biomech*, vol. 62, pp. 1–4, Sep. 2017, doi: 10.1016/j.jbiomech.2017.09.001.

[5] M. A. Lafortune, P. R. Cavanagh, H. J. Sommer, and A. Kalenak, “Three-dimensional kinematics of the human knee during walking,” *Journal of Biomechanics*, vol. 25, no. 4, pp. 347–357, Apr. 1992, doi: 10.1016/0021-9290(92)90254-X.

[6] S. Tashman, D. Collon, K. Anderson, P. Kolowich, and W. Anderst, “Abnormal rotational knee motion during running after anterior cruciate ligament reconstruction,” *Am J Sports Med*, vol. 32, no. 4, pp. 975–983, Jun. 2004, doi: 10.1177/0363546503261709.

[7] R. D. Komistek, D. A. Dennis, and M. Mahfouz, “In vivo fluoroscopic analysis of the normal human knee,” *Clin Orthop Relat Res*, no. 410, pp. 69–81, May 2003, doi: 10.1097/01.blo.0000062384.79828.3b.

[8] J. B. Stiehl, R. D. Komistek, D. A. Dennis, R. D. Paxson, and W. A. Hoff, “Fluoroscopic analysis of kinematics after posterior-cruciate-retaining knee arthroplasty,” *J Bone Joint Surg Br*, vol. 77, no. 6, pp. 884–889, Nov. 1995.

[9] S. A. Banks and W. A. Hodge, “Accurate measurement of three-dimensional knee replacement kinematics using single-plane fluoroscopy,” *IEEE Trans Biomed Eng*, vol. 43, no. 6, pp. 638–649, Jun. 1996, doi: 10.1109/10.495283.

[10] W. A. Hoff, R. D. Komistek, D. A. Dennis, S. M. Gabriel, and S. A. Walker, “Three-dimensional determination of femoral-tibial contact positions under in vivo conditions using fluoroscopy,” *Clin Biomech (Bristol, Avon)*, vol. 13, no. 7, pp. 455–472, Oct. 1998, doi: 10.1016/s0268-0033(98)00009-6.

[11] A. Wong, N. M. Dunk, and J. P. Callaghan, “A systematic approach to feature tracking of lumbar spine vertebrae from fluoroscopic images using complex-valued wavelets,” *Comput Methods Biomech Biomed Engin*, vol. 12, no. 5, pp. 607–616, Oct. 2009, doi: 10.1080/10255840902802891.

[12] J. Wu, “2D-3D Registration of Knee Joint from Single Plane X-ray Fluoroscopy Using Nonlinear Shape Priors,” p. 143.

[13] S. Acker *et al.*, “Accuracy of single-plane fluoroscopy in determining relative position and orientation of total knee replacement components,” *J Biomech*, vol. 44, no. 4, pp. 784–787, Feb. 2011, doi: 10.1016/j.jbiomech.2010.10.033.

[14] J. Hermans, P. Claes, J. Bellemans, D. Vandermeulen, and P. Suetens, “Robust initialization for 2D/3D registration of knee implant models to single-plane fluoroscopy,” in *Medical Imaging 2007: Image Processing*, Mar. 2007, vol. 6512, p. 651208. doi: 10.1117/12.708163.

[15] A. Saveh *et al.*, “Knee Dynamic Analysis Based on 2D-to-3D Registration of Fluoroscopic and Angiographic Images,” *undefined*, 2012. /paper/Knee-Dynamic-Analysis-Based-on-2D-to-3D-of-and-Saveh-Zali/b29116b6ef04a3be0f9bf14c6243488d92dea67d (accessed Sep. 09, 2020).

[16] T.-W. Lu, T.-Y. Tsai, M.-Y. Kuo, H.-C. Hsu, and H.-L. Chen, “In vivo three-dimensional kinematics of the normal knee during active extension under unloaded and loaded conditions using single-plane fluoroscopy,” *Medical Engineering & Physics*, vol. 30, no. 8, pp. 1004–1012, Oct. 2008, doi: 10.1016/j.medengphy.2008.03.001.

[17] A. Hurvitz and L. Joskowicz, “Registration of a CT-like atlas to fluoroscopic X-ray images using intensity correspondences,” *Int J CARS*, vol. 3, no. 6, pp. 493–504, Dec. 2008, doi: 10.1007/s11548-008-0264-z.

[18] A. H. Saveh, H. R. Katouzian, and M. Chizari, “Measurement of an intact knee kinematics using gait and fluoroscopic analysis,” *Knee Surg Sports Traumatol Arthrosc*, vol. 19, no. 2, pp. 267–272, Feb. 2011, doi: 10.1007/s00167-010-1190-5.

[19] L. Zollei, E. Grimson, A. Norbash, and W. Wells, “2D-3D rigid registration of X-ray fluoroscopy and CT images using mutual information and sparsely sampled histogram estimators,” in *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, Dec. 2001, vol. 2, p. II–II. doi: 10.1109/CVPR.2001.991032.

[20] J. Wu and M. R. Mahfouz, “Reconstruction of knee anatomy from single-plane fluoroscopic x-ray based on a nonlinear statistical shape model,” *J Med Imaging (Bellingham)*, vol. 8, no. 1, p. 016001, Jan. 2021, doi: 10.1117/1.JMI.8.1.016001.

[21] W. Anderst, R. Zauel, J. Bishop, E. Demps, and S. Tashman, “Validation of three-dimensional model-based tibio-femoral tracking during running,” *Med Eng Phys*, vol. 31, no. 1, pp. 10–16, Jan. 2009, doi: 10.1016/j.medengphy.2008.03.003.

[22] M. R. Torry *et al.*, “Knee kinematic profiles during drop landings: a biplane fluoroscopy study,” *Med Sci Sports Exerc*, vol. 43, no. 3, pp. 533–541, Mar. 2011, doi: 10.1249/MSS.0b013e3181f1e491.

[23] E. S. Abebe *et al.*, “The effects of femoral graft placement on in vivo knee kinematics after anterior cruciate ligament reconstruction,” *J Biomech*, vol. 44, no. 5, pp. 924–929, Mar. 2011, doi: 10.1016/j.jbiomech.2010.11.028.

[24] “Strategies for Reducing Radiation Dose in CT.” https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2743386/ (accessed Apr. 30, 2021).

[25] Y.-R. Chen, H.-Y. Lin, and M.-L. Wang, “3-D model reconstruction from C-arm images,” in *2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, Oct. 2016, pp. 002454–002459. doi: 10.1109/SMC.2016.7844607.

[26] R. Kurazume *et al.*, “3D reconstruction of a femoral shape using a parametric model and two 2D fluoroscopic images,” *Computer Vision and Image Understanding*, vol. 113, no. 2, pp. 202–211, Feb. 2009, doi: 10.1016/j.cviu.2008.08.012.

[27] G. Zheng, M. A. G. Ballester, M. Styner, and L.-P. Nolte, “Reconstruction of patient-specific 3D bone surface from 2D calibrated fluoroscopic images and point distribution model,” *Med Image Comput Comput Assist Interv*, vol. 9, no. Pt 1, pp. 25–32, 2006, doi: 10.1007/11866565_4.

[28] R. Bakker, “The Effect of Sagittal Plane Mechanics on Anterior Cruciate Ligament Strain During Jump Landing,” p. 104.

[29] adolph, “Knee Anatomy,” *SlideServe*, Oct. 09, 2014. https://www.slideserve.com/adolph/knee-anatomy (accessed May 15, 2021).

[30] “Bone - Bone morphology,” *Encyclopedia Britannica*. https://www.britannica.com/science/bone-anatomy (accessed May 15, 2021).

[31] “Three-dimensional Morphology of the Knee Reveals Ethnic Differences.” https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3237843/ (accessed May 14, 2021).

[32] S. F. Fucentese, A. von Roll, P. P. Koch, D. R. Epari, B. Fuchs, and P. B. Schottle, “The patella morphology in trochlear dysplasia--a comparative MRI study,” *Knee*, vol. 13, no. 2, pp. 145–150, Mar. 2006, doi: 10.1016/j.knee.2005.12.005.

[33] “Unit VII.” https://courses.vcu.edu/DANC291-003/unit_7.htm (accessed May 15, 2021).

[34] H. Iwaki, V. Pinskerova, and M. A. Freeman, “Tibiofemoral movement 1: the shapes and relative movements of the femur and tibia in the unloaded cadaver knee,” *J Bone Joint Surg Br*, vol. 82, no. 8, pp. 1189–1195, Nov. 2000, doi: 10.1302/0301-620x.82b8.10717.

[35] L. M. Smoger *et al.*, “Statistical modeling to characterize relationships between knee anatomy and kinematics,” *Journal of Orthopaedic Research*, vol. 33, no. 11, pp. 1620–1630, 2015, doi: 10.1002/jor.22948.

[36] B. Yue, K. M. Varadarajan, S. Ai, T. Tang, H. E. Rubash, and G. Li, “Differences of Knee Anthropometry Between Chinese and White Men and Women,” *The Journal of Arthroplasty*, vol. 26, no. 1, pp. 124–130, Jan. 2011, doi: 10.1016/j.arth.2009.11.020.

[37] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, “Active Shape Models-Their Training and Application,” *Computer Vision and Image Understanding*, vol. 61, no. 1, pp. 38–59, Jan. 1995, doi: 10.1006/cviu.1995.1004.

[38] C. Rao, C. K. Fitzpatrick, P. J. Rullkoetter, L. P. Maletsky, R. H. Kim, and P. J. Laz, “A statistical finite element model of the knee accounting for shape and alignment variability,” *Medical Engineering & Physics*, vol. 35, no. 10, pp. 1450–1456, Oct. 2013, doi: 10.1016/j.medengphy.2013.03.021.

[39] M. Roberts, T. F. Cootes, and J. E. Adams, “Vertebral morphometry: semiautomatic determination of detailed shape from dual-energy X-ray absorptiometry images using active appearance models,” *Invest Radiol*, vol. 41, no. 12, pp. 849–859, Dec. 2006, doi: 10.1097/01.rli.0000244343.27431.26.

[40] C. Lindner, S. Thiagarajah, J. M. Wilkinson, G. A. Wallis, T. F. Cootes, and arcOGEN Consortium, “Accurate bone segmentation in 2D radiographs using fully automatic shape model matching based on regression-voting,” *Med Image Comput Comput Assist Interv*, vol. 16, no. Pt 2, pp. 181–189, 2013, doi: 10.1007/978-3-642-40763-5_23.

[41] M. G. Roberts, E. M. B. Pacheco, R. Mohankumar, T. F. Cootes, and J. E. Adams, “Detection of vertebral fractures in DXA VFA images using statistical models of appearance and a semi-automatic segmentation,” *Osteoporos Int*, vol. 21, no. 12, pp. 2037–2046, Dec. 2010, doi: 10.1007/s00198-009-1169-6.

[42] Z. Zhu and G. Li, “Construction of 3D human distal femoral surface models using a 3D statistical deformable model,” *Journal of Biomechanics*, vol. 44, no. 13, pp. 2362–2368, Sep. 2011, doi: 10.1016/j.jbiomech.2011.07.006.

[43] T. L. Bredbenner, T. D. Eliason, R. S. Potter, R. L. Mason, L. M. Havill, and D. P. Nicolella, “Statistical shape modeling describes variation in tibia and femur surface geometry between Control and Incidence groups from the osteoarthritis initiative database,” *J Biomech*, vol. 43, no. 9, pp. 1780–1786, Jun. 2010, doi: 10.1016/j.jbiomech.2010.02.015.

[44] “SimTK: Statistical Shape Model of the Knee: Downloads.” https://simtk.org/frs/?group_id=960 (accessed Apr. 14, 2021).

[45] K. S. Arun, T. S. Huang, and S. D. Blostein, “Least-Squares Fitting of Two 3-D Point Sets,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-9, no. 5, pp. 698–700, Sep. 1987, doi: 10.1109/TPAMI.1987.4767965.

[46] C. Rao, “Representing Intersubject Variability with a Statistical Shape and Alignment Model of the Knee,” p. 94.

[47] “Statistics of Medical Imaging - 1st Edition - Tianhu Lei - Routledge.” https://www.routledge.com/Statistics-of-Medical-Imaging/Lei/p/book/9781138114746 (accessed May 15, 2021).

[48] “Image Processing with MATLAB: Applications in Medicine and Biology - 1.” https://www.routledge.com/Image-Processing-with-MATLAB-Applications-in-Medicine-and-Biology/Demirkaya-Asyali-Sahoo/p/book/9780849392467 (accessed May 15, 2021).

[49] “Practical Body MRI.” https://www.cambridge.org/core/books/practical-body-mri/A10DC95DBA9958AECFE05C5FB8C5FEDC (accessed May 15, 2021).

[50] “A new method to compare image quality in CT and MRI images | IEEE Conference Publication | IEEE Xplore.” https://ieeexplore.ieee.org/document/5966736 (accessed May 15, 2021).

[51] B. M. Frye, A. A. Najim, J. B. Adams, K. R. Berend, and A. V. Lombardi, “MRI is more accurate than CT for patient-specific total knee arthroplasty,” *Knee*, vol. 22, no. 6, pp. 609–612, Dec. 2015, doi: 10.1016/j.knee.2015.02.014.

[52] Y. Uozumi, K. Nagamune, and K. Mizuno, “Computer-Aided Segmentation System of Posterior Cruciate Ligament in Knee Joint from CT and MRI Using Anatomical Information: A Pilot Study of System Configuration,” in *2015 IEEE International Conference on Systems, Man, and Cybernetics*, Oct. 2015, pp. 2295–2298. doi: 10.1109/SMC.2015.401.

[53] J. Vennart, “The 1990 recommendations of the International Commission on Radiological Protection,” *J. Radiol. Prot.*, vol. 11, no. 3, pp. 199–203, Sep. 1991, doi: 10.1088/0952-4746/11/3/006.

[54] T. Tharmviboonsri, K. Riansuwan, A. Nitising, and B. Mahaisavariya, “Radiation exposure during 3D fluoroscopy of the knee: an experimental study,” *Eur J Trauma Emerg Surg*, vol. 38, no. 3, pp. 307–311, Jun. 2012, doi: 10.1007/s00068-011-0165-7.

[55] “Image Noise.” http://www.sprawls.org/ppmi2/NOISE/#Fluoroscopy (accessed Apr. 14, 2021).

[56] L. Deng and D. Yu, “Deep Learning: Methods and Applications,” *SIG*, vol. 7, no. 3–4, pp. 197–387, Jun. 2014, doi: 10.1561/2000000039.

[57] G. Litjens *et al.*, “A Survey on Deep Learning in Medical Image Analysis,” *Medical Image Analysis*, vol. 42, pp. 60–88, Dec. 2017, doi: 10.1016/j.media.2017.07.005.

[58] H. Zhang and B. Dong, “A Review on Deep Learning in Medical Image Reconstruction,” *arXiv:1906.10643 [physics]*, Jun. 2019, Accessed: Apr. 04, 2021. [Online]. Available: http://arxiv.org/abs/1906.10643

[59] “Gradient-based learning applied to document recognition | IEEE Journals & Magazine | IEEE Xplore.” https://ieeexplore.ieee.org/document/726791 (accessed Apr. 13, 2021).

[60] S. Albawi, T. A. Mohammed, and S. Al-Zawi, “Understanding of a convolutional neural network,” in *2017 International Conference on Engineering and Technology (ICET)*, Aug. 2017, pp. 1–6. doi: 10.1109/ICEngTechnol.2017.8308186.

[61] K. O’Shea and R. Nash, “An Introduction to Convolutional Neural Networks,” *arXiv:1511.08458 [cs]*, Dec. 2015, Accessed: May 17, 2021. [Online]. Available: http://arxiv.org/abs/1511.08458

[62] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet classification with deep convolutional neural networks,” *Commun. ACM*, vol. 60, no. 6, pp. 84–90, May 2017, doi: 10.1145/3065386.

[63] “Convolutional Neural Networks for Medical Image Analysis: Full Training or Fine Tuning? | IEEE Journals & Magazine | IEEE Xplore.” https://ieeexplore.ieee.org/document/7426826 (accessed Apr. 22, 2021).

[64] J. T. Springenberg, A. Dosovitskiy, T. Brox, and M. Riedmiller, “Striving for Simplicity: The All Convolutional Net,” *arXiv:1412.6806 [cs]*, Apr. 2015, Accessed: Apr. 13, 2021. [Online]. Available: http://arxiv.org/abs/1412.6806

[65] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, “Dropout: A Simple Way to Prevent Neural Networks from Overfitting,” *Journal of Machine Learning Research*, vol. 15, no. 56, pp. 1929–1958, 2014.

[66] S. Ioffe and C. Szegedy, “Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift,” in *International Conference on Machine Learning*, Jun. 2015, pp. 448–456. Accessed: Apr. 13, 2021. [Online]. Available: http://proceedings.mlr.press/v37/ioffe15.html

[67] M. Leshno, V. Y. Lin, A. Pinkus, and S. Schocken, *Multilayer feedforward networks with a non-polynomial activation function can approximate any function*. 1992.

[68] K. He, X. Zhang, S. Ren, and J. Sun, “Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification,” *arXiv:1502.01852 [cs]*, Feb. 2015, Accessed: Apr. 13, 2021. [Online]. Available: http://arxiv.org/abs/1502.01852

[69] D.-A. Clevert, T. Unterthiner, and S. Hochreiter, “Fast and Accurate Deep Network Learning by Exponential Linear Units (ELUs),” *arXiv:1511.07289 [cs]*, Feb. 2016, Accessed: Apr. 13, 2021. [Online]. Available: http://arxiv.org/abs/1511.07289

[70] “NVIDIA cuDNN | NVIDIA Developer.” https://developer.nvidia.com/cudnn (accessed May 18, 2021).

[71] “PyTorch.” https://www.pytorch.org (accessed Apr. 20, 2021).

[72] “TensorFlow.” https://www.tensorflow.org/ (accessed May 17, 2021).

[73] “Keras: the Python deep learning API.” https://keras.io/ (accessed May 17, 2021).

[74] K. Simonyan and A. Zisserman, “Very Deep Convolutional Networks for Large-Scale Image Recognition,” *arXiv:1409.1556 [cs]*, Apr. 2015, Accessed: Apr. 13, 2021. [Online]. Available: http://arxiv.org/abs/1409.1556

[75] “Deep Residual Learning for Image Recognition | IEEE Conference Publication | IEEE Xplore.” https://ieeexplore.ieee.org/document/7780459 (accessed Apr. 13, 2021).

[76] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation,” in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, Cham, 2015, pp. 234–241. doi: 10.1007/978-3-319-24574-4_28.

[77] “Dataset Shift in Machine Learning | The MIT Press.” https://mitpress.mit.edu/books/dataset-shift-machine-learning (accessed May 17, 2021).

[78] “Can We Trust Deep Learning Based Diagnosis? The Impact of Domain Shift in Chest Radiograph Classification | SpringerLink.” https://link.springer.com/chapter/10.1007%2F978-3-030-62469-9_7 (accessed May 17, 2021).

[79] “[PDF] Dataset shift in classification: terminology, benchmarks and methods | Semantic Scholar.” https://www.semanticscholar.org/paper/Dataset-shift-in-classification%3A-terminology%2C-and-Moreno-Torres/7f367c8aabfed4460f965cd6e353608d49eb6b8f (accessed May 17, 2021).

[80] H. Guan and M. Liu, “Domain Adaptation for Medical Image Analysis: A Survey,” *arXiv:2102.09508 [cs, eess]*, Feb. 2021, Accessed: May 17, 2021. [Online]. Available: http://arxiv.org/abs/2102.09508

[81] S. Mardanisamani *et al.*, “Crop Lodging Prediction From UAV-Acquired Images of Wheat and Canola Using a DCNN Augmented With Handcrafted Texture Features,” 2019, pp. 0–0. Accessed: May 18, 2021. [Online]. Available: https://openaccess.thecvf.com/content_CVPRW_2019/html/CVPPP/Mardanisamani_Crop_Lodging_Prediction_From_UAV-Acquired_Images_of_Wheat_and_Canola_CVPRW_2019_paper.html

[82] N. Ghatwary, X. Ye, and M. Zolgharni, “Esophageal Abnormality Detection Using DenseNet Based Faster R-CNN With Gabor Features,” *IEEE Access*, vol. 7, pp. 84374–84385, 2019, doi: 10.1109/ACCESS.2019.2925585.

[83] C.-L. Liu, B. Xiao, W.-H. Hsaio, and V. S. Tseng, “Epileptic Seizure Prediction With Multi-View Convolutional Neural Networks,” *IEEE Access*, vol. 7, pp. 170352–170361, 2019, doi: 10.1109/ACCESS.2019.2955285.

[84] M. Choudhary, V. Tiwari, and V. U., “Iris anti-spoofing through score-level fusion of handcrafted and data-driven features,” *Applied Soft Computing*, vol. 91, p. 106206, Jun. 2020, doi: 10.1016/j.asoc.2020.106206.

[85] D. Heo, J. Y. Nam, and B. C. Ko, “Estimation of Pedestrian Pose Orientation Using Soft Target Training Based on Teacher–Student Framework,” *Sensors*, vol. 19, no. 5, Art. no. 5, Jan. 2019, doi: 10.3390/s19051147.

[86] H. J. Kwon, H. Il Koo, J. W. Soh, and N. Ik Cho, “Age Estimation Using Trainable Gabor Wavelet Layers In A Convolutional Neural Network,” in *2019 IEEE International Conference on Image Processing (ICIP)*, Sep. 2019, pp. 3626–3630. doi: 10.1109/ICIP.2019.8803442.

[87] D. P. Van Hoai and V. T. Hoang, “Feeding Convolutional Neural Network by hand-crafted features based on Enhanced Neighbor-Center Different Image for color texture classification,” in *2019 International Conference on Multimedia Analysis and Pattern Recognition (MAPR)*, May 2019, pp. 1–6. doi: 10.1109/MAPR.2019.8743528.

[88] M. Jbene, A. D. el Maliani, and M. El Hassouni, “Fusion of Convolutional Neural Network and Statistical Features for Texture classification,” in *2019 International Conference on Wireless Networks and Mobile Communications (WINCOM)*, Oct. 2019, pp. 1–4. doi: 10.1109/WINCOM47513.2019.8942469.

[89] A. Majid, M. A. Khan, M. Yasmin, A. Rehman, A. Yousafzai, and U. Tariq, “Classification of stomach infections: A paradigm of convolutional neural network along with classical features fusion and selection,” *Microscopy Research and Technique*, vol. 83, no. 5, pp. 562–576, 2020, doi: https://doi.org/10.1002/jemt.23447.

[90] S. Hosseini, S. H. Lee, and N. I. Cho, “Feeding Hand-Crafted Features for Enhancing the Performance of Convolutional Neural Networks,” *arXiv:1801.07848 [cs]*, Jan. 2018, Accessed: May 18, 2021. [Online]. Available: http://arxiv.org/abs/1801.07848

[91] “Gabor filters as texture discriminator | SpringerLink.” https://link.springer.com/article/10.1007/BF00204594 (accessed May 18, 2021).

[92] L. Ding and A. Goshtasby, “On the Canny edge detector,” *Pattern Recognition*, vol. 34, pp. 721–725, 2001.

[93] N. Baka *et al.*, “Evaluation of automated statistical shape model based knee kinematics from biplane fluoroscopy,” *Journal of Biomechanics*, vol. 47, no. 1, pp. 122–129, Jan. 2014, doi: 10.1016/j.jbiomech.2013.09.022.

[94] L. Hao and W. Kheng, “Registration-Based Segmentation of Medical Images,” 2006. /paper/Registration-Based-Segmentation-of-Medical-Images-Hao-Kheng/d0486f8a34776a8a3f6010b14ec550a631317ad9 (accessed Apr. 15, 2021).

[95] A. Aprovitola and L. Gallo, “Knee bone segmentation from MRI: A classification and literature review,” *Biocybernetics and Biomedical Engineering*, vol. 36, no. 2, pp. 437–449, Jan. 2016, doi: 10.1016/j.bbe.2015.12.007.

[96] D. A. Hussien, F. M. Abed, and A. A. Hasan, “Stereo Photogrammetry vs Computed Tomography for 3D Medical Measurements,” *Karbala International Journal of Modern Science*, vol. 5, no. 4, Dec. 2019, doi: 10.33640/2405-609X.1130.

[97] K. Rathnayaka *et al.*, “Quantification of the accuracy of MRI generated 3D models of long bones compared to CT generated 3D models,” *Med Eng Phys*, vol. 34, no. 3, pp. 357–363, Apr. 2012, doi: 10.1016/j.medengphy.2011.07.027.

[98] G. B. Sharma *et al.*, “Radiological method for measuring patellofemoral tracking and tibiofemoral kinematics before and after total knee replacement,” *Bone Joint Res*, vol. 1, no. 10, pp. 263–271, Oct. 2012, doi: 10.1302/2046-3758.110.2000117.

[99] S. Hosseinian and H. Arefi, “3D RECONSTRUCTION FROM MULTI-VIEW MEDICAL X-RAY IMAGES – REVIEW AND EVALUATION OF EXISTING METHODS,” *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, vol. XL-1/W5, pp. 319–326, Dec. 2015, doi: 10.5194/isprsarchives-XL-1-W5-319-2015.

[100] A. K. Jain, M. An, N. Chitphakdithai, G. Chintalapani, and G. Fichtinger, “C-arm calibration: is it really necessary?,” San Diego, CA, Mar. 2007, p. 65092U. doi: 10.1117/12.712301.

[101] W. A. Wallace and F. Johnson, “Detection and correction of geometrical distortion in X-ray fluoroscopic images,” *Journal of Biomechanics*, vol. 14, no. 2, pp. 123–125, Jan. 1981, doi: 10.1016/0021-9290(81)90171-8.

[102] “Morphometric evaluations of personalised 3D reconstructions and geometric models of the human spine | SpringerLink.” https://link.springer.com/article/10.1007/BF02510968 (accessed May 16, 2021).

[103] M. J. Pearcy, “Stereo radiography of lumbar spine motion,” *Acta Orthop Scand Suppl*, vol. 212, pp. 1–45, 1985, doi: 10.3109/17453678509154154.

[104] D. Mitton *et al.*, “3D reconstruction of the pelvis from bi-planar radiography,” *Comput Methods Biomech Biomed Engin*, vol. 9, no. 1, pp. 1–5, Feb. 2006, doi: 10.1080/10255840500521786.

[105] S. Laporte, W. Skalli, J. A. de Guise, F. Lavaste, and D. Mitton, “A biplanar reconstruction method based on 2D and 3D contours: application to the distal femur,” *Comput Methods Biomech Biomed Engin*, vol. 6, no. 1, pp. 1–6, Feb. 2003, doi: 10.1080/1025584031000065956.

[106] “3D reconstruction method from biplanar radiography using non-stereocorresponding points and elastic deformable meshes | SpringerLink.” https://link.springer.com/article/10.1007/BF02344767 (accessed May 16, 2021).

[107] M. Rudemo, “Statistical Shape Analysis. I. L. Dryden and K. V. Mardia, Wiley, Chichester 1998. No. of pages: xvii+347. Price: £60.00.ISBN 0-471-95816-6,” *Statistics in Medicine*, vol. 19, no. 19, pp. 2716–2717, 2000, doi: https://doi.org/10.1002/1097-0258(20001015)19:19<2716::AID-SIM590>3.0.CO;2-O.

[108] K. T. Rajamani, M. A. G. Ballester, L.-P. Nolte, and M. Styner, “A novel and stable approach to anatomical structure morphing for enhanced intraoperative 3D visualization,” in *Medical Imaging 2005: Visualization, Image-Guided Procedures, and Display*, Apr. 2005, vol. 5744, pp. 718–725. doi: 10.1117/12.595175.

[109] S. Kadoury, F. Cheriet, and H. Labelle, “Personalized X-ray 3-D reconstruction of the scoliotic spine from hybrid statistical and image-based models,” *IEEE Trans Med Imaging*, vol. 28, no. 9, pp. 1422–1435, Sep. 2009, doi: 10.1109/TMI.2009.2016756.

[110] N. Baka *et al.*, “2D–3D shape reconstruction of the distal femur from stereo X-ray imaging using statistical shape models,” *Medical Image Analysis*, vol. 15, no. 6, pp. 840–850, Dec. 2011, doi: 10.1016/j.media.2011.04.001.

[111] M. A. Baldwin, J. E. Langenderfer, P. J. Rullkoetter, and P. J. Laz, “Development of subject-specific and statistical shape models of the knee using an efficient segmentation and mesh-morphing approach,” *Comput Methods Programs Biomed*, vol. 97, no. 3, pp. 232–240, Mar. 2010, doi: 10.1016/j.cmpb.2009.07.005.

[112] Y. Chaibi *et al.*, “Fast 3D reconstruction of the lower limb using a parametric model and statistical inferences and clinical measurements calculation from biplanar X-rays,” *Computer Methods*

in Biomechanics and Biomedical Engineering, vol. 15, pp. 457–66, May 2012, doi: 10.1080/10255842.2010.540758.

[113] W. Yu, M. Tannast, and G. Zheng, "Non-rigid free-form 2D–3D registration using a B-spline-based statistical deformation model," *Pattern Recognition*, vol. 63, pp. 689–699, Mar. 2017, doi: 10.1016/j.patcog.2016.09.036.

[114] S. Quijano, A. Serrurier, B. Aubert, S. Laporte, P. Thoreux, and W. Skalli, "Three-dimensional reconstruction of the lower limb from biplanar calibrated radiographs," *Medical Engineering & Physics*, vol. 35, no. 12, pp. 1703–1712, Dec. 2013, doi: 10.1016/j.medengphy.2013.07.002.

[115] H. Lamecker, T. H. Wenckebach, and H.- Hege, "Atlas-based 3D-Shape Reconstruction from X-Ray Images," in *18th International Conference on Pattern Recognition (ICPR'06)*, Aug. 2006, vol. 1, pp. 371–374. doi: 10.1109/ICPR.2006.279.

[116] K. Morooka, M. Nakamoto, and Y. Sato, "A Survey on Statistical Modeling and Machine Learning Approaches to Computer Assisted Medical Intervention: Intraoperative Anatomy Modeling and Optimization of Interventional Procedures," *IEICE Transactions on Information and Systems*, vol. E96.D, pp. 784–797, Apr. 2013, doi: 10.1587/transinf.E96.D.784.

[117] N. Wang, Y. Zhang, Z. Li, Y. Fu, W. Liu, and Y.-G. Jiang, "Pixel2Mesh: Generating 3D Mesh Models from Single RGB Images," *arXiv:1804.01654 [cs]*, Aug. 2018, Accessed: Apr. 05, 2021. [Online]. Available: http://arxiv.org/abs/1804.01654

[118] "3D ShapeNets: A deep representation for volumetric shapes | IEEE Conference Publication | IEEE Xplore." https://ieeexplore.ieee.org/document/7298801 (accessed Apr. 13, 2021).

[119] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed. Cambridge: Cambridge University Press, 2004. doi: 10.1017/CBO9780511811685.

[120] A. Kar, C. Häne, and J. Malik, "Learning a Multi-View Stereo Machine," *arXiv:1708.05375 [cs]*, Aug. 2017, Accessed: Apr. 13, 2021. [Online]. Available: http://arxiv.org/abs/1708.05375

[121] C. B. Choy, D. Xu, J. Gwak, K. Chen, and S. Savarese, “3D-R2N2: A Unified Approach for Single and Multi-view 3D Object Reconstruction,” *arXiv:1604.00449 [cs]*, Apr. 2016, Accessed: Jan. 12, 2021. [Online]. Available: http://arxiv.org/abs/1604.00449

[122] H. Xie, H. Yao, X. Sun, S. Zhou, and S. Zhang, “Pix2Vox: Context-aware 3D Reconstruction from Single and Multi-view Images,” *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 2690–2698, Oct. 2019, doi: 10.1109/ICCV.2019.00278.

[123] C. Wen, Y. Zhang, Z. Li, and Y. Fu, “Pixel2Mesh++: Multi-View 3D Mesh Generation via Deformation,” *arXiv:1908.01491 [cs]*, Aug. 2019, Accessed: Apr. 13, 2021. [Online]. Available: http://arxiv.org/abs/1908.01491

[124] “Using Bi-planar X-Ray Images to Reconstruct the Spine Structure by the Convolution Neural Network,” *springerprofessional.de*. https://www.springerprofessional.de/en/using-bi-planar-x-ray-images-to-reconstruct-the-spine-structure-/17218070 (accessed Apr. 05, 2021).

[125] Y. Kasten, D. Doktofsky, and I. Kovler, “End-To-End Convolutional Neural Network for 3D Reconstruction of Knee Bones From Bi-Planar X-Ray Images,” *arXiv:2004.00871 [cs, eess]*, Aug. 2020, Accessed: Jan. 12, 2021. [Online]. Available: http://arxiv.org/abs/2004.00871

[126] “Fluoroscopy and Digital Photospot | Radiology | SUNY Upstate Medical University.” https://www.upstate.edu/radiology/education/rsna/fluoro/fluoro.php (accessed May 12, 2021).

[127] P. J. Gamage, “Quantum noise | Radiology Reference Article | Radiopaedia.org,” *Radiopaedia*. https://radiopaedia.org/articles/quantum-noise (accessed May 12, 2021).

[128] H. Elgstrom, “Assessment of image quality in x-ray fluoroscopy based on Model observers as an objective measure for quality control and image optimization,” p. 57.

[129] C. Wu, P. Zhang, Y. Xu, and J. Yao, “Combined spatial and temporal deep learning for image noise reduction of fluoroscopic x-ray sequences,” in *Medical Imaging 2020: Physics of Medical Imaging*, Mar. 2020, vol. 11312, p. 1131240. doi: 10.1117/12.2549693.

[130] “Patient-specific reconstruction of volumetric computed tomography images from a single projection view via deep learning | Nature Biomedical Engineering.” https://www.nature.com/articles/s41551-019-0466-4?proof=t (accessed Apr. 14, 2021).

[131] “Atlas-based 3D-Shape Reconstruction from X-Ray Images | IEEE Conference Publication | IEEE Xplore.” https://ieeexplore.ieee.org/document/1698910 (accessed Apr. 13, 2021).

[132] K. Sohan and M. A. Yousuf, “3D Bone Shape Reconstruction from 2D X-ray Images Using MED Generative Adversarial Network,” in *2020 2nd International Conference on Advanced Information and Communication Technology (ICAICT)*, Nov. 2020, pp. 53–58. doi: 10.1109/ICAICT51780.2020.9333477.

[133] J. Čavojská, J. Petrasch, D. Mattern, N. J. Lehmann, A. Voisard, and P. Böttcher, “Estimating and abstracting the 3D structure of feline bones using neural networks on X-ray (2D) images,” *Commun Biol*, vol. 3, no. 1, p. 337, Dec. 2020, doi: 10.1038/s42003-020-1057-3.

[134] H. Kim, K. Lee, D. Lee, and N. Baek, “3D Reconstruction of Leg Bones from X-Ray Images using CNN-based Feature Analysis,” in *2019 International Conference on Information and Communication Technology Convergence (ICTC)*, Oct. 2019, pp. 669–672. doi: 10.1109/ICTC46691.2019.8939984.

[135] X. Ying, H. Guo, K. Ma, J. Wu, Z. Weng, and Y. Zheng, “X2CT-GAN: Reconstructing CT from Biplanar X-Rays with Generative Adversarial Networks,” *arXiv:1905.06902 [cs, eess]*, May 2019, Accessed: Apr. 14, 2021. [Online]. Available: http://arxiv.org/abs/1905.06902

[136] P. Henzler, V. Rasche, T. Ropinski, and T. Ritschel, “Single-image Tomography: 3D Volumes from 2D Cranial X-Rays,” *arXiv:1710.04867 [cs]*, Nov. 2018, Accessed: Apr. 04, 2021. [Online]. Available: http://arxiv.org/abs/1710.04867

[137] F. D’Isidoro, “How does the hip joint move? Techniques and applications,” Doctoral Thesis, ETH Zurich, 2018. doi: 10.3929/ethz-b-000332702.

[138] B. Goswami and S. Kr., “3D Modeling of X-Ray Images: A Review,” *IJCA*, vol. 132, no. 7, pp. 40–46, Dec. 2015, doi: 10.5120/ijca2015907566.

[139] W. Yu, C. Chu, M. Tannast, and G. Zheng, “Fully automatic reconstruction of personalized 3D volumes of the proximal femur from 2D X-ray images,” *Int J Comput Assist Radiol Surg*, vol. 11, no. 9, pp. 1673–1685, Sep. 2016, doi: 10.1007/s11548-016-1400-9.

[140] V. Karade and B. Ravi, “3D femur model reconstruction from biplane X-ray images: a novel method based on Laplacian surface deformation,” *Int J CARS*, vol. 10, no. 4, pp. 473–485, Apr. 2015, doi: 10.1007/s11548-014-1097-6.

[141] R. Gheno *et al.*, “Three-dimensional measurements of the lower extremity in children and adolescents using a low-dose biplanar X-ray device,” *Eur Radiol*, vol. 22, no. 4, pp. 765–771, Apr. 2012, doi: 10.1007/s00330-011-2308-y.

[142] G. Zheng, “Statistical shape model-based reconstruction of a scaled, patient-specific surface model of the pelvis from a single standard AP x-ray radiograph,” *Med Phys*, vol. 37, no. 4, pp. 1424–1439, Apr. 2010, doi: 10.1118/1.3327453.

[143] T. Cresson, D. Branchaud, R. Chav, B. Godbout, and J. A. de Guise, “3D shape reconstruction of bone from two x-ray images using 2D/3D non-rigid registration based on moving least-squares deformation,” San Diego, California, USA, Mar. 2010, p. 76230F. doi: 10.1117/12.844098.

[144] G. Zheng, S. Gollmer, S. Schumann, X. Dong, T. Feilkas, and M. A. González Ballester, “A 2D/3D correspondence building method for reconstruction of a patient-specific 3D bone surface model using point distribution models and calibrated X-ray images,” *Med Image Anal*, vol. 13, no. 6, pp. 883–899, Dec. 2009, doi: 10.1016/j.media.2008.12.003.

[145] M. Kass, A. Witkin, and D. Terzopoulos, “Snakes: Active contour models,” *Int J Comput Vision*, vol. 1, no. 4, pp. 321–331, Jan. 1988, doi: 10.1007/BF00133570.

[146] “The interpretation of structure from motion | Proceedings of the Royal Society of London. Series B. Biological Sciences.” https://royalsocietypublishing.org/doi/10.1098/rspb.1979.0006 (accessed May 12, 2021).

[147] A. R. Widya *et al.*, “3D Reconstruction of Whole Stomach from Endoscope Video Using Structure-from-Motion,” *arXiv:1905.12988 [cs, eess]*, May 2019, Accessed: May 18, 2021. [Online]. Available: http://arxiv.org/abs/1905.12988

[148] N. Micheletti, J. H. Chandler, and S. N. Lane, “Structure from Motion (SfM) Photogrammetry,” *Geomorphological Techniques*, p. 12, 2015.

[149] M. G. Wagner, P. F. Laeseke, T. Schubert, J. M. Slagowski, M. A. Speidel, and C. A. Mistretta, “Feature-based respiratory motion tracking in native fluoroscopic sequences for dynamic roadmaps during minimally invasive procedures in the thorax and abdomen,” in *Medical Imaging 2017: Image-Guided Procedures, Robotic Interventions, and Modeling*, Mar. 2017, vol. 10135, p. 101351H. doi: 10.1117/12.2254148.

[150] M. Klüppel, J. Wang, D. Bernecker, P. Fischer, and J. Hornegger, “On Feature Tracking in X-Ray Images,” in *Bildverarbeitung für die Medizin 2014*, T. M. Deserno, H. Handels, H.-P. Meinzer, and T. Tolxdorff, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2014, pp. 132–137. doi: 10.1007/978-3-642-54111-7_28.

[151] W. P. Anderson, “Dynamic In Vivo Skeletal Feature Tracking Via Fluoroscopy Using a Human Gait Model,” p. 181.

[152] C. L. Baker, “Computed Tomography from Imagery Generated by Fluoroscopy along an Arbitrary Path,” p. 135.

[153] E. H. Adelson, C. H. Anderson, J. R. Bergen, P. J. Burt, and J. M. Ogden, “Pyramid methods in image processing,” p. 9, 1984.

[154] M. B. Stegmann and D. D. Gomez, “A Brief Introduction to Statistical Shape Analysis,” p. 15.

[155] Alexander, “Bone pose estimation Challenge dataset.” Zenodo, Jan. 26, 2017. doi: 10.5281/zenodo.4470102.

[156] P. Fischer *et al.*, “FlowNet: Learning Optical Flow with Convolutional Networks,” *arXiv:1504.06852 [cs]*, May 2015, Accessed: Apr. 17, 2021. [Online]. Available: http://arxiv.org/abs/1504.06852

[157] “Deep Feature Learning for Knee Cartilage Segmentation Using a Triplanar Convolutional Neural Network | SpringerLink.” https://link.springer.com/chapter/10.1007/978-3-642-40763-5_31 (accessed Apr. 22, 2021).

[158] S. Miao, Z. J. Wang, Y. Zheng, and R. Liao, “Real-time 2D/3D registration via CNN regression,” in *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*, Prague, Czech Republic, Apr. 2016, pp. 1430–1434. doi: 10.1109/ISBI.2016.7493536.

[159] C. Spampinato, S. Palazzo, D. Giordano, M. Aldinucci, and R. Leonardi, “Deep learning for automated skeletal bone age assessment in X-ray images,” *Medical Image Analysis*, vol. 36, pp. 41–51, Feb. 2017, doi: 10.1016/j.media.2016.10.010.

[160] “Weights & Biases – Developer tools for ML.” https://wandb.ai/site (accessed Apr. 20, 2021).

[161] *Point Cloud Alignment using ICP (Cyrill Stachniss, 2020; updated)*, (Apr. 03, 2020). Accessed: Sep. 09, 2020. [Online Video]. Available: https://www.youtube.com/watch?v=djnd502836w&ab_channel=CyrillStachniss

[162] T. S. Y. Tang and R. E. Ellis, “2D/3D Deformable Registration Using a Hybrid Atlas,” in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2005*, Berlin, Heidelberg, 2005, pp. 223–230. doi: 10.1007/11566489_28.

[163] C. D. Murawski, A. F. Chen, and F. H. Fu, “Radiographic femoral bicondylar width predicts anterior cruciate ligament insertion site sizes,” *Knee Surg Sports Traumatol Arthrosc*, vol. 25, no. 8, pp. 2424–2427, Aug. 2017, doi: 10.1007/s00167-015-3886-z.

[164] W. M. Kouw and M. Loog, “A review of domain adaptation without target labels,” *arXiv:1901.05335 [cs, stat]*, Jul. 2019, Accessed: May 18, 2021. [Online]. Available: http://arxiv.org/abs/1901.05335

[165] M. Toldo, A. Maracani, U. Michieli, and P. Zanuttigh, “Unsupervised Domain Adaptation in Semantic Segmentation: a Review,” *arXiv:2005.10876 [cs, eess]*, May 2020, Accessed: May 18, 2021. [Online]. Available: http://arxiv.org/abs/2005.10876

[166] M. Wang and W. Deng, “Deep Visual Domain Adaptation: A Survey,” *Neurocomputing*, vol. 312, pp. 135–153, Oct. 2018, doi: 10.1016/j.neucom.2018.05.083.

[167] K. Rajendran, “Mechanism of locking at the knee joint.,” *J Anat*, vol. 143, pp. 189–194, Dec. 1985.

[168] T. F. Cootes, “Chapter 24 - Linear statistical shape models and landmark location,” in *Handbook of Medical Image Computing and Computer Assisted Intervention*, S. K. Zhou, D. Rueckert, and G. Fichtinger, Eds. Academic Press, 2020, pp. 575–598. doi: 10.1016/B978-0-12-816176-0.00029-6.

[169] K. M. Ribeiro, R. A. Braga Jr., G. W. Horgan, D. D. Ferreira, and T. Sáfadi, “Principal component analysis in the spectral analysis of the dynamic laser speckle patterns,” *JEOS:RP*, vol. 9, p. 14009, Feb. 2014, doi: 10.2971/jeos.2014.14009.

[170] “[1603.07254] Gaussian Process Morphable Models.” https://arxiv.org/abs/1603.07254 (accessed Apr. 30, 2021).

[171] B. Schölkopf, A. Smola, and K.-R. Müller, “Kernel principal component analysis,” in *Artificial Neural Networks — ICANN’97*, vol. 1327, W. Gerstner, A. Germond, M. Hasler, and J.-D. Nicoud, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 1997, pp. 583–588. doi: 10.1007/BFb0020217.

[172] H. Rao, “Computational Modelling of Knee Tissue Mechanics During Single-Leg Jump Landing,” Jan. 2020, Accessed: May 15, 2021. [Online]. Available: https://uwspace.uwaterloo.ca/handle/10012/15543

Appendix

Real X-ray images, the corresponding extracted features, and the validated reconstructed shape model of eight subjects

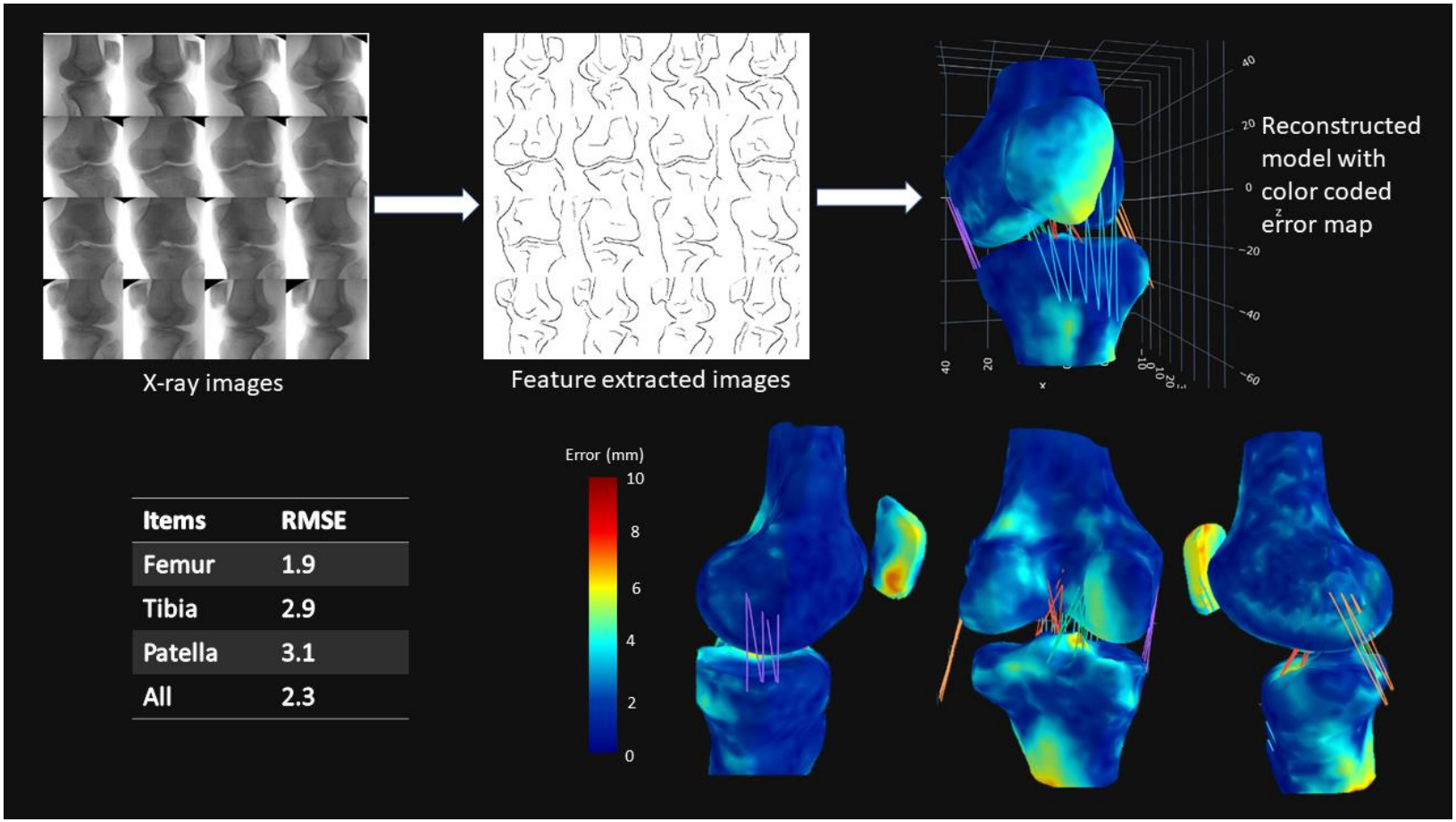
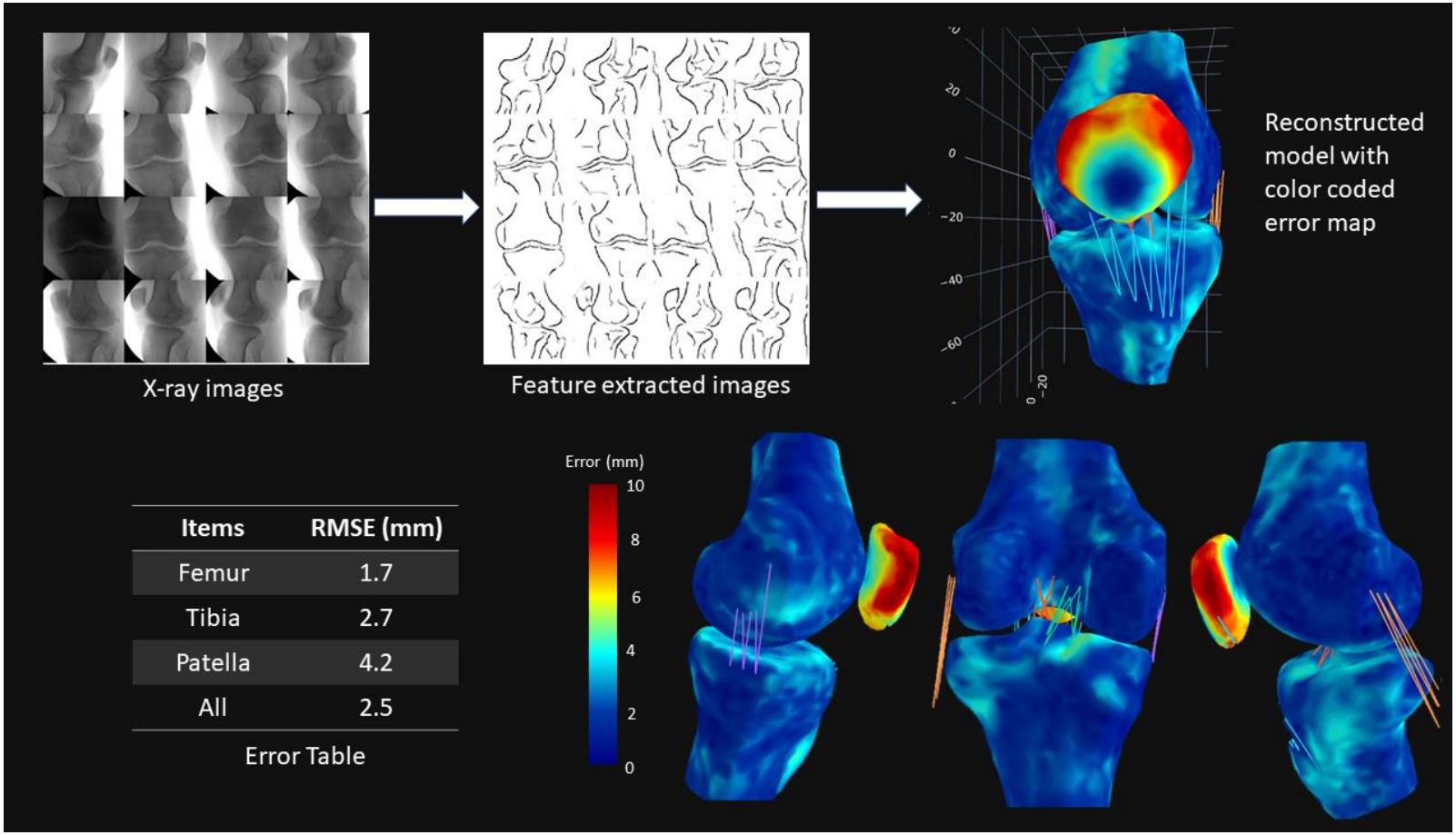


Figure A.1: Results of subject 1



Items	RMSE (mm)
Femur	1.7
Tibia	2.7
Patella	4.2
All	2.5

Error Table

Figure A.2: Results of subject 2

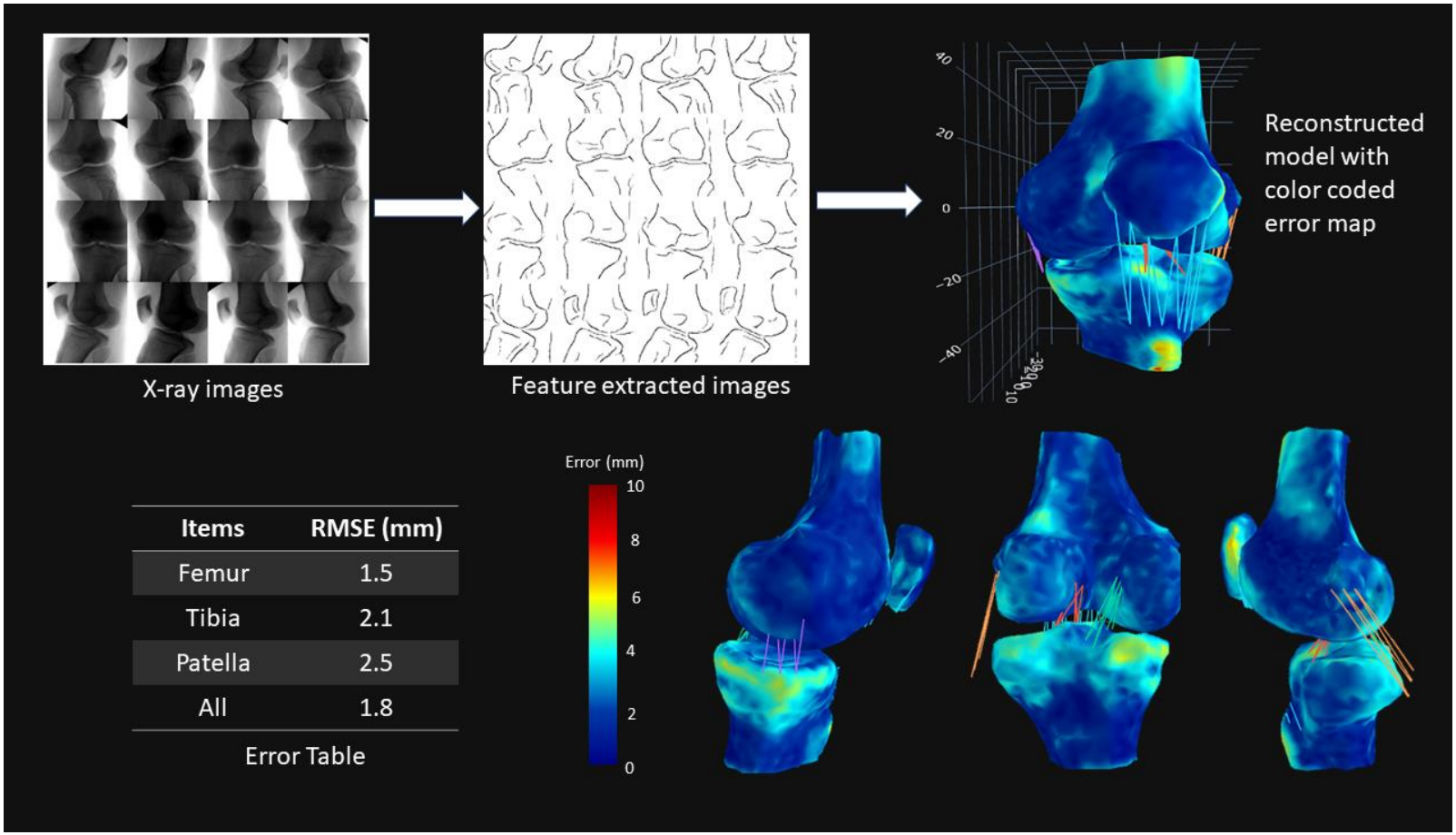


Figure A.3: Results of subject 3

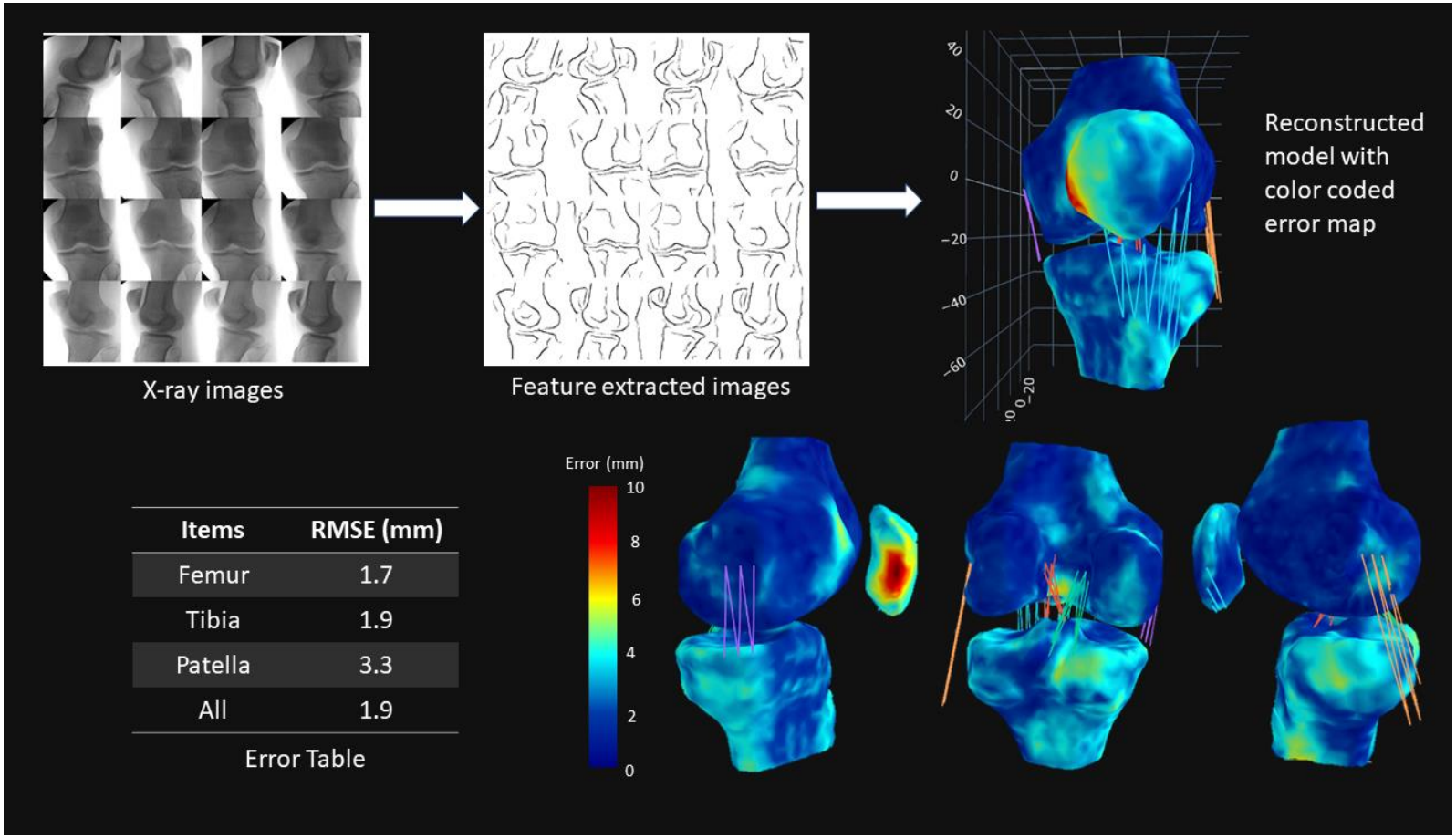
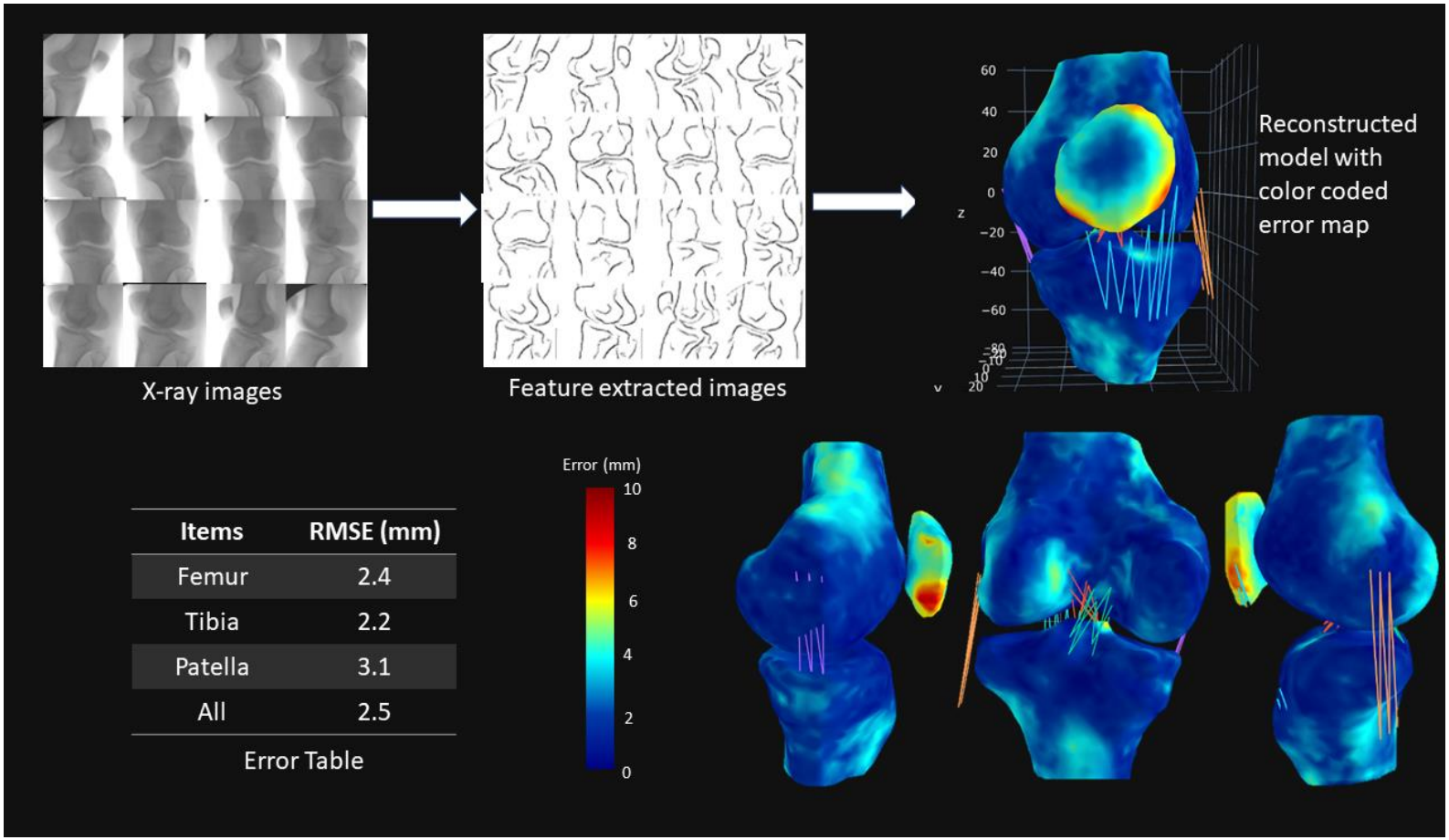


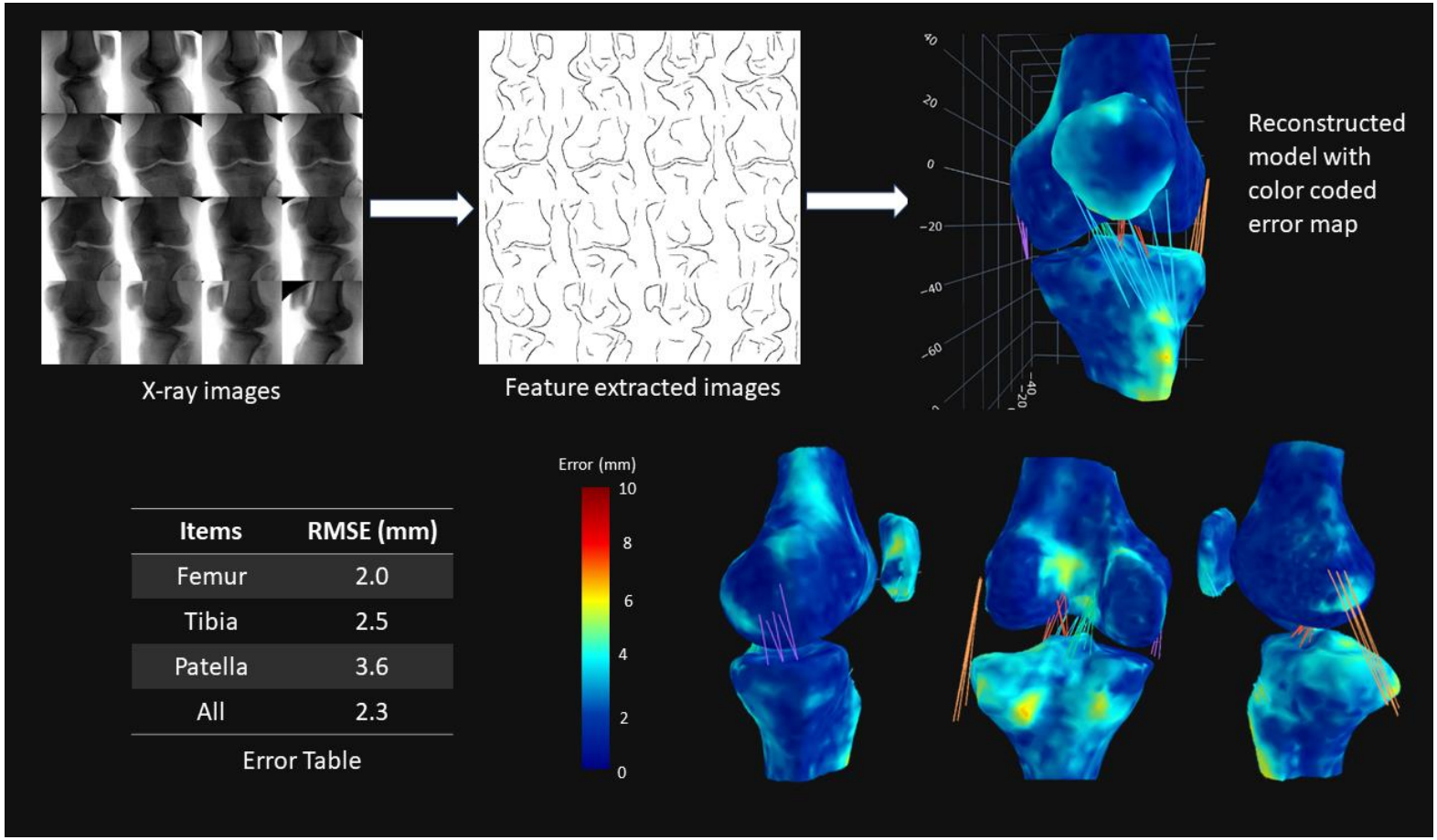
Figure A.4: Results of subject 4



Items	RMSE (mm)
Femur	2.4
Tibia	2.2
Patella	3.1
All	2.5

Error Table

Figure A.5: Results of subject 5



Items	RMSE (mm)
Femur	2.0
Tibia	2.5
Patella	3.6
All	2.3

Error Table

Figure A.6: Results of subject 6

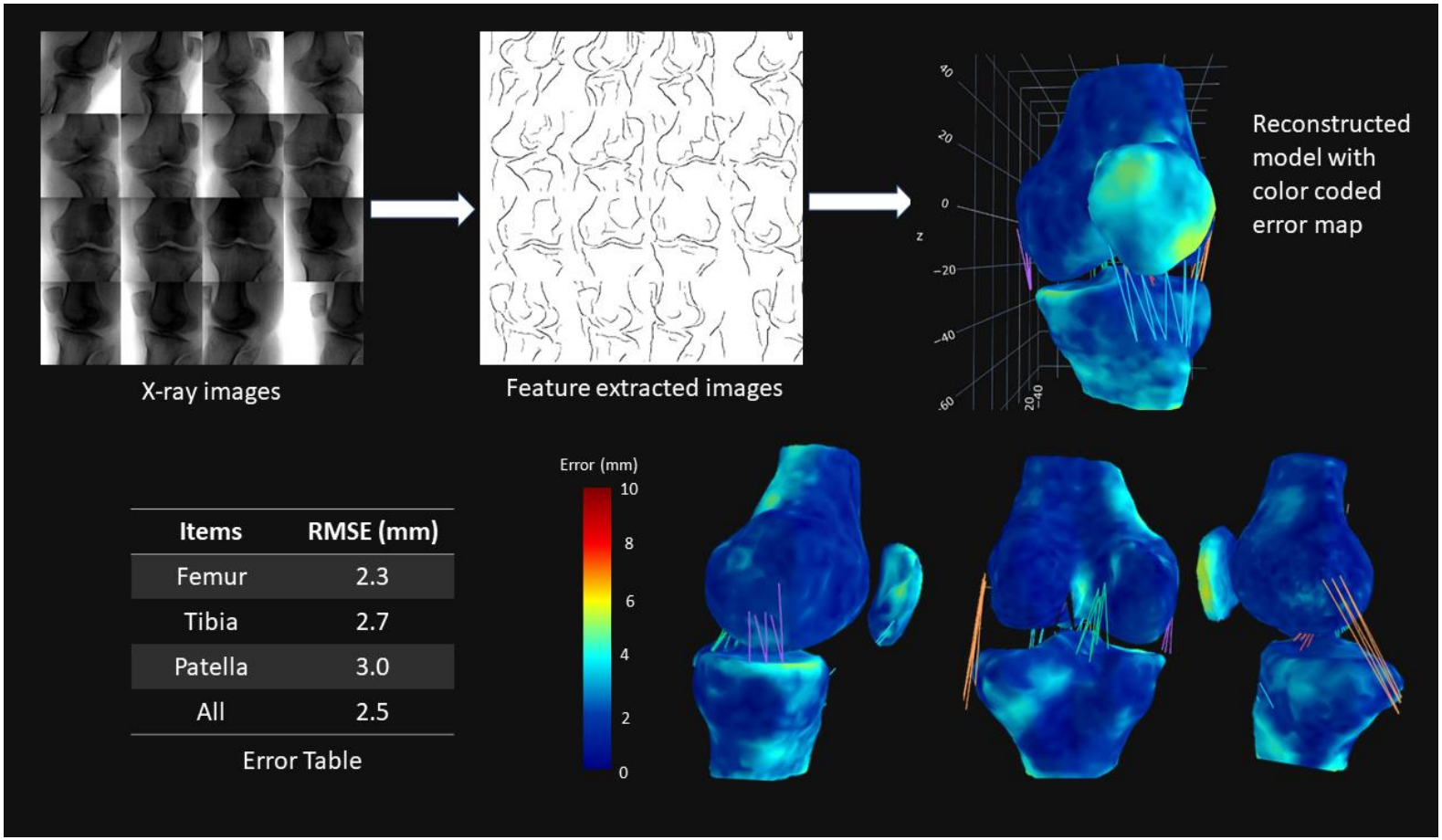
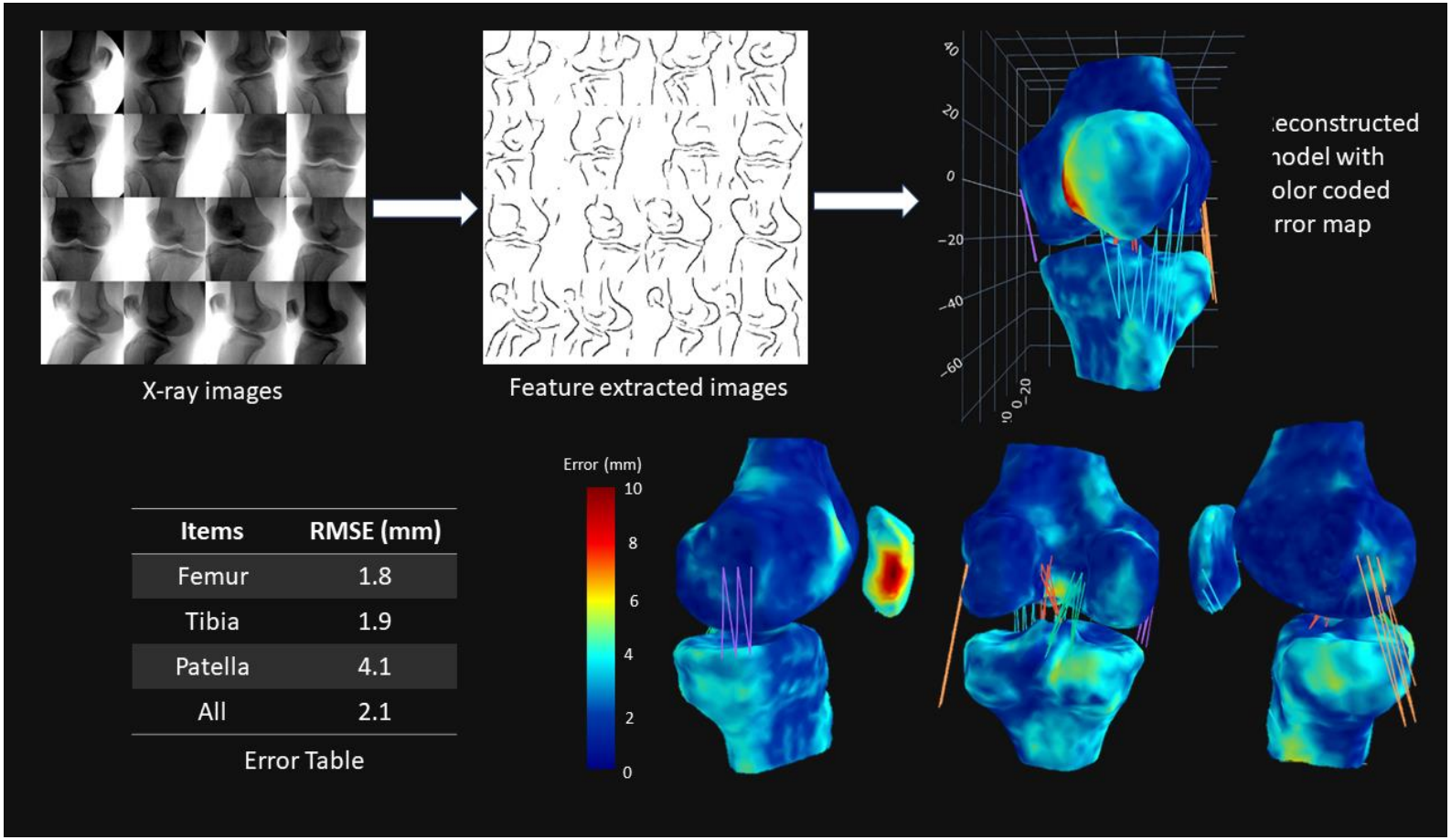


Figure A.7: Results of subject 7



Items	RMSE (mm)
Femur	1.8
Tibia	1.9
Patella	4.1
All	2.1

Error Table

Figure A.8: Results of subject 10