UTS UNIVERSITY OF TECHNOLOGY SYDNEY

Advanced Approaches for Bone CT Analysis Based on Deep Learning

by Xiaoxu Li

Thesis submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy

under the supervision of A/Prof. Min Xu

University of Technology Sydney
Faculty of Engineering and Information Technology

August 2022

CERTIFICATE OF ORIGINAL AUTHORSHIP

I, Xiaoxu Li declare that this thesis, is submitted in fulfilment of the requirements for the award of Doctor of Philosophy, in the School of Electrical and Data Engineering, Faculty of Engineering and Information Technology at the University of Technology Sydney.

This thesis is wholly my own work unless otherwise referenced or acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

This document has not been submitted for qualifications at any other academic institution.

This research is supported by the Australian Government Research Training Program.

Signature: Production Note:
Signature removed prior to publication.

Date: 15/Aug/2022

Acknowledgements

First and foremost, I would like to express my sincere thanks to my supervisor A/Prof. Min Xu. Thanks for Min's continual support, supervision, and encouragement during the Ph.D. study. She offered this opportunity to do research in the field of deep learning-based bone CT analysis and gave valuable insights during the work. It is my honor and fortune to have a supervisor like her.

I am grateful to my co-supervisors, Haimin Zhang and Yu Peng (StraxCorp, Melbourne), for their kind support. Thanks for their supervision, suggestions, precious discussion, and feedback on my work. I am also grateful to Prof. Jian Zhang and A/Prof. Qiang Wu, for their help and suggestions on my research.

I am also expressing my thanks to my colleagues in A/Prof. Min Xu' research group, in particular, Tianrong Rao, Lei Sang, Qiyu Liao, Lingxiang Wu, Madhu Takalkar, Ruiheng Zhang, Yukun Yang, and Wanneng Wu. Thanks for their selfless help in life and study. It is a precious memory to work with them.

I would also like to thank the colleagues at StraxCorp, in particular, Chao Sui, Sicong Ma, Vasudha Bhat, Tommy Zhao, Haris Dar, and Amit Shewani, for their support, discussion, and data labeling during the research. I am also grateful to my friends, Yang He and Lang Chen, for their warm help.

Finally, I would like to thank my family for their support and encouragement during the Ph.D. study.

Abstract

Computed Tomography (CT), as a 3D imaging technique, has greatly facilitated bone analysis over the past few decades. This thesis aimed to design novel deep learning approaches to analyse human bone CT. Four works have been conducted, i.e., anatomical segmentation of foot weight-bearing cone beam CT (CBCT), instance segmentation of wrist CT, semi-supervised segmentation of bone CT, and bone health analysis via bone fracture prediction.

In the first work, we developed a multi-stage method, FootSeg, for the anatomical segmentation of foot CT. FootSeg consisted of three parts, foot preprocessing, foot region segmentation, and foot bone classification. The multi-stage framework greatly simplified the implementation of the FootSeg method and achieved both qualitatively and quantitatively remarkable results. The mean Intersection-Over-Union on the bone parts was 90.3% on the testing set. To the best of our knowledge, this was the first research of fully automatic foot anatomical segmentation from weight-bearing CBCT via deep learning methods.

The second work focused on the instance segmentation of wrist CT. A novel semi-automatic method was designed to annotate 5K wrist CT slices. The annotation workload and time have been greatly reduced. An end-to-end edge reinforced U-net segmentation model was developed and demonstrated satisfying results. To the best of our knowledge, this was the first work on wrist CT instance segmentation using deep learning methods.

The third work aimed to solve the bone segmentation problem with fewer annotation data via semi-supervised learning. A patch-shuffled data transformation method was developed, and a patch-shuffle-based semi-supervised segmentation method was proposed for bone CT segmentation. Two supervised losses and a consistent unsupervised

loss were employed to utilize both the labeled and unlabeled data. The proposed method was evaluated on various bone CT datasets, and the results demonstrated superior performance.

The last work was about bone health analysis via bone fracture prediction. We collected data from three population-based cohorts and processed the unstructured raw data as a structured database for model training and evaluation. We developed a deep learning-based fracture prediction model to predict the bone fragility fracture in the next five years. Compared with the clinical index of BMD T-score and FRAX, the proposed model could identify the bones with fragility fracture within five years with higher AUC values. This was the first research using the deep learning models to identify individuals with upcoming fragility fractures using wrist CT.

Publications

Journal Papers:

1. Tianrong Rao, Xiaoxu Li, Haimin Zhang, Min Xu, "Multi-level region-based convolutional neural network for image emotion classification," *Neurocomputing*, vol. 333, pp. 429-439, Mar, 2019.
2. Tianronog Rao, Xiaoxu Li, Min Xu, "Learning multi-level deep representations for image emotion classification," *Neural Processing Letters*, vol. 51, no. 3, pp. 2043-2061, June 2020.
3. Xiaoxu Li, Yu Peng, Min Xu, "Patch-Shuffle-Based Semi-Supervised Segmentation of Bone Computed Tomography via Consistent Learning," submitted to *Biomedical Signal Processing and Control*, 2022.
4. Xiaoxu Li, Yu Peng, Min Xu, "Deep Learning in Bone CT: a Systematic Review," submitted to *Artificial Intelligence in Medicine*, 2022.
5. Xiaoxu Li, Roland Chapurlat, Serge Ferrari, Min Bui, Ali Ghazem-Zadeh, Ego Seeman, Min Xu, Yu Peng, "Deep Learning Using Only High-Resolution Forearm Images Predicts Fracture," submitted to *New England Journal of Medicine*, 2022.

Conference Papers:

1. Xiaoxu Li, Yu Peng, Min Xu, "Edge-enhanced Instance Segmentation of Wrist CT via a Semi-Automatic Annotation Database Construction Method," in *Proc. 2021*

DICTA: Digital Image Computing: Techniques and Applications, pp. 590-597, Nov 2021.

2. Xiaoxu Li, Yu Peng, Min Xu, "FootSeg: Automatic Anatomical Segmentation of Foot Bones from Weight-Bearing Cone Beam CT Scans," submitted to *DICTA: Digital Image Computing: Techniques and Applications*, 2022.

Table of contents

List of figures

List of tables

Chapter 1

Introduction

This chapter started with the background and motivation of this thesis, then introduced the objectives and contributions, and finally listed the structure of this thesis.

1.1 Background and Motivation

Computed Tomography (CT), as a 3D imaging technique, has been widely used for the clinical diagnosis of bone in the past decades, such as surgery planning [1, 2], fracture detection [3, 4], and osteoporosis detection [5, 6]. Compared with the other radiology techniques such as X-ray or Magnetic resonance imaging (MRI), the bone and its microstructure are depicted more clearly in the CT scans as depicted in Fig. 1.1. In recent years, different kinds of CT machines, such as high resolution peripheral quantitative (HRpQCT), micro CT, and cone-beam CT (CBCT), have been developed to meet the particular needs for orthopedics clinical usage. Doctors have employed these high-quality CT for the medical diagnosis of various body parts, such as foot, wrist, vertebra, and pelvic. More and more bone CT data are generated. Manually analyzing these CT data is time-consuming and depends on the experience and expert knowledge of the doctor. However, the automatic bone CT analysis tools are still immature and in urgent demand.

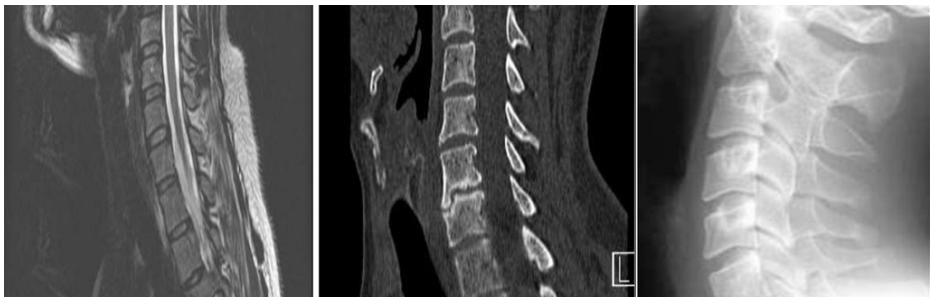

Fig. 1.1 Comparison of CT, MRI and X-ray data on the human spine.

The traditional bone CT analysis studies [7–11] focused on hand-crafted features, and were brittle to complicated occasions, which led to unsatisfactory results. Deep learning (DL) [12], as a subset of Artificial Intelligence (AI), extracted discriminative features and learned the representation of the data using neural networks. Using the convolutional neural networks (CNN), deep learning has achieved state-of-the-art performance in image classification [13, 14] and medical image analysis [15, 16]. Employing the deep learning approaches for bone CT analysis has gained increasing attention, and promising results have been achieved on tasks such as vertebral segmentation [17–21], and disease classification [22–27].

Recently, a new CT technique, weight-bearing cone beam CT (CBCT), has been developed for foot medical analysis. The weight-bearing CBCT could provide high-resolution foot scans in the natural weight-bearing position. Figure 1.2 [28] depicts the scanning procedure of weight-bearing CBCT. The high quality scans from CBCT machines have greatly facilitated the treatment and diagnosis of human foot [29], such as foot align [2] and foot surgery [30, 31]. In these clinical treatments, a fundamental step to analyze the foot CBCT scan is the anatomical segmentation of foot bones, which aims to distribute the correct class to each pixel in the foot CT according to the foot structure. There are thirty-one bones in the human foot, including tibia, fibula, talus, navicular, calcaneus, cuboid, three cuneiform bones, five metatarsal bones, fourteen phalange bones, two sesamoid bones, and accessory bone. The complicated structure of the foot makes the automatic anatomical segmentation of foot CT a challenging work. Besides, there are two challenges that need to be solved. The first one is the foot scan variation. The scanning feet are in different positions and sizes, and either

two feet, left foot or right foot, can be scanned for medical analysis. The foot number, position, and size are different and lead to the challenge of scan variation. The second challenge is the severe data imbalance among the different bone classes. The big bones like the tibia are much larger than the small bones, such as the fifteen phalanges. In our first work, we proposed a multi-stage deep learning-based method, FootSeg, for the anatomical segmentation of foot CT. The FootSeg method consisted of three steps, foot preprocessing, foot region segmentation, and foot bone classification. The multi-stage framework solved the two challenges and achieved both qualitatively and quantitatively remarkable results.

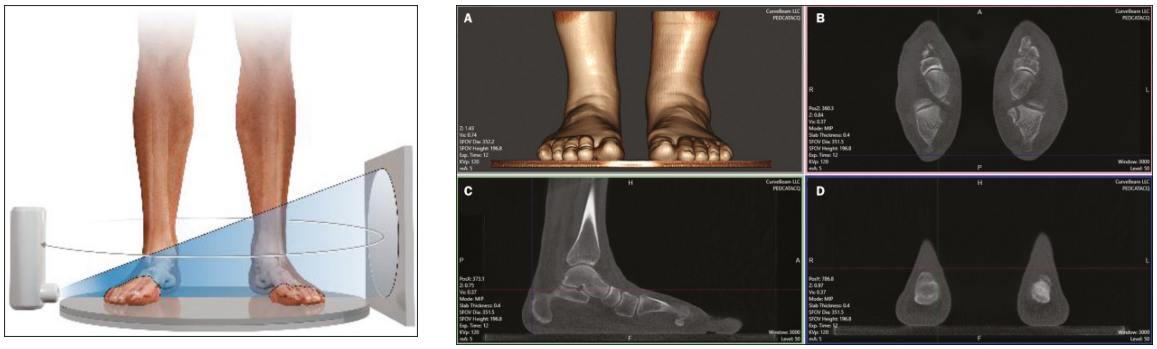

Fig. 1.2 Scanning procedure (left part) and scanning examples (right part) of weight-bearing CBCT.

The Wrist CT performs an essential role in clinical practice and has shown high potentials in various applications such as osteoporosis classification [6], rheumatoid arthritis diagnoses [32], and bone fracture assessment [33]. Similar to foot CT processing, an essential procedure among the above applications is the wrist instance segmentation, i.e., distributing the right class label to each voxel in the wrist CT data. There were some existing works of wrist segmentation in CT images [34–37]. However, most of the existing works still focused on using shallow features, and these methods were brittle to complicated occasions or required human interaction to achieve satisfied segmentation results. What's more, only wrist bones have been segmented while the muscle part has been neglected in these methods. The muscle parts contained health information such as muscle strength, muscle mass, and body mass index. Both muscle and bone were essential in clinical analysis. Traditional methods designed for bone segmentation could not satisfy the need for muscle segmentation. A comprehensive

wrist instance segmentation method that could identify all components, including bones, muscle, and background, in the wrist CT in one go was highly demanded. In our second work, we proposed an edge-enhanced U-net model for the segmentation of wrist CT. For the model training, instead of using manual annotation, we developed a semi-automatic method to annotate the wrist CT via the traditional OTSU method and the U-net-based self-training semi-supervised learning model. The semi-automatic method greatly reduced the annotation workload, and the proposed wrist segmentation model achieved high accuracy.

Sufficient annotation data is essential for training a successful deep learning system. The lack of annotation data often results in the problem of overfitting and underperforming. Several deep learning-based medical segmentation methods [16, 38, 39] have achieved the state of the art performance via massive annotation data. However, the per-pixel manual labeling procedure for bone segmentation tasks is time-consuming and expensive. Using fewer annotation data for bone CT segmentation is in demand. The Hounsfield scale of bone in CT is within a specific range compared with the other parts, such as muscle and air, as shown in Fig. 1.3. This particular imaging feature can be used as prior information for bone segmentation. Besides, the semi-supervised methods, which employed both the labeled and the unlabeled data, have alleviated the workload of data annotation in several tasks [40–42]. In our third work, we explored the semi-supervised learning methods and introduced the prior information of the Hounsfield scale for bone CT segmentation. We proposed a patch-shuffle-based semi-supervised method for bone CT segmentation with a consistent learning strategy and demonstrated the potential for using the semi-supervised method for bone CT annotation.

Bone osteoporosis is a major public health concern, especially for the elderly group. Bone osteoporosis is mainly caused by the decrease of bone mass and leads to bone fragility fracture. The mortality rate is highly increased after the fragility fracture in the first year for the elderly, and the fragility fracture has long-term effects on health and death risk for up to ten years [43, 44]. Analyzing bone health by predicting fractures in the coming years is an important way to improve quality of life. The most common way to measure bone health is bone mineral density (BMD) estimation. A BMD T-score of -2.5 standard deviation (SD) or lower indicated the presence of

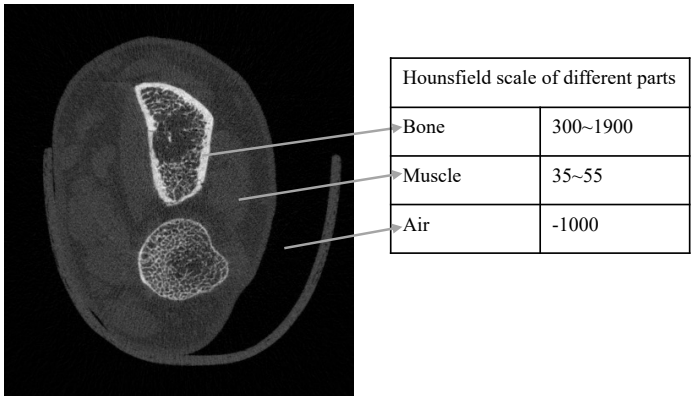

Fig. 1.3 Illustration of the Hounsfield scale for bone, muscle, and air on CT of the wrist.

osteoporosis and a high risk of bone fracture [45]. Another method is the fracture risk assessment tool (FRAX) [46]. The FRAX tool integrates BMD at the femoral neck (FN) and other clinical risk factors to calculate the fracture risk in the next 10 years. However, both methods are insufficient for bone health analysis and fracture prediction. The BMD T-score would vary according to the patient's position during X-ray scanning, and most of the fractures occur at BMD T-score between -1.0 and -2.5 SD. The FRAX tool predicts the bone fracture in the next ten years, which is too long for medical treatment, and the clinical risk factors are ethnically related and are not the same among different countries. In our fourth work, we utilized the deep learning method to develop a more efficient bone health analysis tool for fracture prediction in the mid-term (five years). We collected data from three population-based cohorts to train the deep learning model, and the results demonstrated the effectiveness of the proposed method.

In summary, this thesis focused on developing deep learning methods to analyze the CT data of the foot, wrist, and other bones in clinical treatment. Four tasks have been conducted in this thesis, including:

- Anatomical segmentation of foot weight-bearing CBCT.
- Instance segmentation of wrist CT.
- Semi-supervised segmentation of bone CT.

- Bone health analysis via bone fracture prediction using wrist CT.

The segmentation tasks were fundamental for the bone CT analysis, and predicting bone fractures could identify patients at high risk of fractures and provide treatment earlier.

1.2 Objectives and Contributions

The objectives and contributions are summarized as follows:

- In the first work, we developed a multi-stage method, FootSeg, for the anatomical segmentation of foot CT. FootSeg contained three parts, foot preprocessing, foot region segmentation, and foot bone classification. The multi-stage framework greatly simplified the implementation of the anatomical segmentation model. We introduced a foot standardization method to solve the scan variation problem and an innovative patch-training method to solve the severe data imbalance problem. The model achieved both qualitatively and quantitatively remarkable results, the mean Intersection-Over-Union (mIOU) including the background was 80.3%, and the mIOU on the bone parts was 90.3% on the testing set. To the best of our knowledge, this was the first research of fully automatic foot anatomical segmentation from weight-bearing CBCT via deep learning methods.
- The second work focused on the instance segmentation of wrist CT. A novel semi-automatic method was designed to annotate 5K wrist CT slices. Compared with the time-consuming and laborious manual annotation, the workload has been highly alleviated, and the annotation time was also greatly reduced. An end-to-end edge reinforced U-net segmentation model was developed for the instance segmentation of wrist CT and demonstrated both qualitative and quantitative satisfying results. To the best of our knowledge, this was the first work on wrist CT instance segmentation. All the regions of the wrist CT, including the muscle, radius bone, ulna bone, cast holder, and background, were identified, while the previous methods only worked on the skeletal parts.

- In the third work, we aimed to solve the bone segmentation problem with fewer annotation data via semi-supervised learning. We defined the bone CT segmentation as a local feature-guided task based on the particular Hounsfield scale of different tissues and materials in CT data. A patch-shuffled data transformation method was developed that enabled the segmentation model to segment both original and patch-shuffled CT slices. Then, a patch-shuffle-based semi-supervised segmentation method was proposed for bone CT segmentation while two supervised losses and a consistent unsupervised loss were employed to use both the labeled and unlabeled data. The proposed method was evaluated on a wrist CT dataset, a foot CT dataset, and a bone CT dataset, and the results demonstrated the superior performance of the model.

- The fourth work was about the bone health analysis via bone fracture prediction. We collected data from three population-based cohorts, which were followed for 6.16 years on average. The real unstructured clinical raw data, including the patient clinical information and CT Dicom data, was processed to construct a structured database for model training and evaluation. We developed a deep learning-based fracture prediction model to predict the bone fragility fracture in the next five years. Extensive experiments were conducted on different data selection groups considering the fracture type and patient age. Compared with the clinical index of BMD T-score and FRAX, the proposed model could identify the bones with fragility fracture within five years with higher AUC values. This was the first research using the deep learning models to identify the individuals with upcoming fragility fractures using wrist CT.

1.3 Thesis Organization

This dissertation aimed to design deep learning approaches to assist bone CT analysis in real clinical practice and conducted four related research. The structure of the thesis is as follows:

- **Chapter 1. Introduction**
  This chapter demonstrated the motivation and scope of this dissertation, discussed

the problems of the bone CT analysis in real clinical practice, and listed the conducted research and the structure of this dissertation.

- **Chapter 2. Literature review**
  This chapter reviewed the recent deep learning-based bone CT analysis methods, including the classification, segmentation, regression, and generative tasks.

- **Chapter 3. Anatomical segmentation of human foot CT**
  This chapter presented the FootSeg method for the anatomical segmentation of the foot CT scan and described the three steps, including foot preprocessing, foot bone segmentation, and foot bone classification, to solve the scan variation and data imbalance challenge in the FootSeg method.

- **Chapter 4. Instance segmentation of human wrist CT**
  This chapter worked on the instance segmentation of the wrist CT scan. The semi-automatic method for building the annotation database of wrist CT via the traditional OTSU method and the U-net-based self-training semi-supervised learning model, and the edge-enhanced U-net model for the instance segmentation of wrist CT using the annotation database were described in this chapter.

- **Chapter 5. Semi-supervised segmentation of bone CT**
  This chapter introduced the patch-shuffle-based semi-supervised method for bone CT segmentation by leveraging the unique bone Hounsfield scale in CT data and the consistent learning strategy.

- **Chapter 6. Bone health analysis via bone fracture prediction using wrist CT**
  This chapter depicted the prospective study of the deep learning-based bone health analysis model via predicting the bone fracture in the next five years from wrist CT data. We described the procedure to process the real unstructured clinical data from three cohorts, where more than two thousand patients participated, as structured clinical data and the deep learning model for bone fracture prediction based on the structured dataset in this chapter.

- **Chapter 7. Conclusion and future work**
  This chapter summarized the works of this dissertation and discussed the future research directions for deep learning-based bone CT analysis.

Chapter 2

Literature Review

In this chapter, an overview of the recent deep learning applications in bone CT was presented. According to the characteristics of the research, the deep learning-based bone CT analysis methods could be divided into classification, segmentation, regression, and generative task. The details of different applications, including model designing, data processing, and evaluation, were described in the following sections.

2.1 Classification Tasks for Bone CT Analysis

The classification task aims to assign a class label to the bone CT data according to the property of bone CT applications. Various classification tasks have been explored, like bone class recognition [47], bone fracture classification [48–54], disease classification [22–27] and gender classification [55]. The typical networks structures, such as AlexNet [47, 56, 57], VGGNet [24, 25, 27, 58], GoogLeNet [23, 25, 52, 55, 59], ResNet [26, 27, 53, 60], DenseNet [22, 27, 61], were usually used as the network backbone.

Miki et al. [47] investigated the AlexNet [56] for tooth type classification on dental CBCT scans. The regions of interest (ROIs) of each tooth were firstly manually extracted from the CT slices, and then, the ROIs were classified into seven-tooth types by the AlexNet. 42 CT volumes were used for training, while the remaining 10 CT volumes were used for testing. Since the number of training samples of each tooth

type was unbalanced, a random sampling method was performed on each tooth type to balance the training samples. The data augmentation methods, including random rotation and intensity transformation, were performed on the grayscale ROI regions, and the average classification was 88.8% with the data augmentation methods. Bewes et al. [55] explored the deep learning method for determining gender through skull CT scans. They collected 900 skull CT scans for training (450 male and 450 female) and 100 skull CT scans for testing (50 male and 50 female) from different ancestries, and the skulls were rotated to the left lateral plane to generate 2D color images. GoogLeNet [59] was adapted for gender determination, and the data augmentation techniques like random rotation, translation, flipping were used to prevent over-fitting. The male classification rate was 96%, and the female classification rate was 94%, which indicated the potential of deep learning for gender estimation.

Bone fracture is a major public health problem, and the mortality rates are highly increased in the first year after bone fracture for the elderly population. The deep learning methods [49–54] have revealed its great potential on identifying the bone fracture. Pranata et al. [49] developed a calcaneus fracture classification model by fine-tuning the VGGNet [58] and ResNet [60]. Performances of the two models were tested using 1931 CT images. Both VGGNet and ResNet achieved high accuracy of 98% on the test dataset, and ResNet was chosen as the final model due to its deeper architecture. Meng et al. [48] proposed a rib fracture detection system to classify the four kinds of bone fractures. A cascaded feature pyramid network was used to detect the bone fracture region, and a 3D neural network was used to classify the fracture type. The developed system achieved a classification accuracy of 86.3%, which was higher than two radiologists (81/2% and 85.0%). With the assistance of the deep learning model, the radiologists achieved a higher F1-score of 96.0%, and the reading time was decreased by more than 100 seconds. Li et al. [53] explored the performance of the ResNet-50 model on identifying benign and malignant vertebral fracture. A dataset of 433 spinal CT images was used to train and test the model via a 10-fold cross-validation strategy. The overall accuracy of the ResNet-50 model was 85% on the per-slice diagnosis and 88% on the per-patient diagnosis.

The disease diagnosis [23–25, 27] was another important classification application. Kim et al. [24] used the 3D CBCT craniofacial scan to detect the malocclusion via

VGGNet and Inception-V3 [62]. The 3D CBCT scans were projected as three 2D images from three different views. They explored two different fusion methods to feed the three images into the deep learning model, ensemble and synchronized multi-channel method. The Inception-V3 model has achieved the best accuracy of 93.83% with the synchronized multi-channel method, and a class-selective relevance mapping method [63] was employed to highlight the malocclusion area predicted by the model. Papandrianos et al. [25] proposed a CNN model with three convolutional-pooling layers, a dense layer, and an output layer for the classification of bone metastasis. The other common CNN models, such as VGGNet, ResNet-50, MobileNet [64], Inception-V3, Xception [65] model have also been explored. The proposed model outperformed the popular CNN models and achieved a testing accuracy of 97.38%. You et al. [23] developed a sagittal craniosynostosis classification model via Inception-V3 model and transfer learning. Fifty sagittal scans were used to finetune the model, and the 3D skull scans were projected as 2D binary images by hemispherical projection as input. The prediction accuracy was great than 90% which outperformed their previous hand-crafted features-based method (72%).

The details of the aforementioned classification research and the other researches were listed in Table 2.1 and Table 2.2. Table 2.1 listed the details of head-related classification tasks, and Table 2.2 included other parts, such as the whole body, spine, chest, and foot. Most methods were based on popular models like GooLeNet, ResNet, and VGGNet, while finetuning and data augmentation were usually used during training. The results showed that the deep learning methods achieved high performance on various tasks and could assist the doctor during the medical diagnosis.

2.2 Segmentation Tasks for Bone CT Analysis

The segmentation task aims to distribute the class label to each pixel at the bone CT images according to a specific segmentation protocol. The segmentation usually acted as a fundamental part for the bone CT analysis. The deep learning models have been applied on the two-class or multi-class segmentation of different human bone parts, such as whole body bone segmentation [66, 67], skull segmentation [68–74], temporal

Table 2.1 Classification task details on head-related CT.

Reference	Application	Method;remarks
Miki et al. (2017) [47]	Tooth type classification (head CT, 7-class)	AlexNet as backbone; manually extract the tooth bounding box; data augmentation
Bewes et al. (2019) [55]	Gender classification (head CT, 2-class)	GoogLeNet as backbone; 3D skull projected as a 2D image from lateral view; data augmentation
Zakirov et al. (2019) [22]	Tooth condition classification (head CT, 6-class)	DenseNet as backbone; weighted binary cross entropy loss; multi-label classification
Kim et al. (2020) [24]	Tooth malocclusion classification (head CT, 3-class)	VGGNet and Inception-V3 as backbone; generate the 2D image from 3-view of CT and use prediction ensemble or feature fusion; Class-selective Relevance Mapping to visualize region of interest
You et al. (2020) [23]	Sagittal craniosynostosis classification (head CT, 2-class)	Inception-V3 with transfer learning; segment using Hounsfield Unit threshold; project 3D scan as 2D image

bone segmentation [75–80], tooth segmentation [22, 81–85], shoulder segmentation [86], clavicle bone segmentation [87], vertebra segmentation [17, 18, 20, 21, 88–95], rib segmentation [48], metacarpal bone segmentation [96], pelvic segmentation [40, 97–100], femur segmentation [101], bone metastasis segmentation [102–105], mineralized tissue segmentation [106], and other segmentation tasks[107]. The U-net [108], fully convolutional networks (FCN) [109], Mask-RCNN [110] were the most popular segmentation models among these segmentation tasks.

Identifying the bone region from the CT scans is a crucial task for the clinical diagnosis, and this was tackled as a two-classes bone segmentation task [66–68, 81, 100, 104, 106, 111].

Klein et al. [66] presented a bone segmentation model in whole-body CT scans by the U-net model. The U-Net model was trained using three different methods: 1) training from 2D axial slices, 2) training from axial, sagittal, and coronal slices and average the outputs, 3) training from 2D axial slices from an unsupervised pretraining model. A private dataset (53 CT scans) and a public dataset (27 CT scans) were used

Table 2.2 Classification task details on whole-body, spine, chest, lower body, foot and bone CT.

Reference	Application	Method;remarks
Pranata et al. (2019) [49]	Calcaneus fractures classification (foot CT, 2-class	Finetunine pretrained VGG16 and ResNet50; single slice as input
Farda et al. (2021) [54]	Calcaneus fractures classification (foot CT, 4-class)	PCANet as backbone; performance comparison on different numbers of augmented images
Chettrit et al. (2020) [50]	Vertebrae fractures classification (Spine CT, 2-class)	Multi-stage framework; 3D patch as input; sequence items classification and aggregation
Husseini et al. (2020) [51]	Vertebrae fractures classification (Spine CT, 3-class)	Grading loss to encourage learning the fracture severeness; TSNE visualisation of feature representation
Li et al. (2021) [53]	Vertebral fracture benign/malignant classification (spinal CT, 2-class)	Three consecutive slices as inputs; ResNet50 as backbone; Two evaluations, per-slice classification and per-patient diagnosis
Lee et al. (2020) [52]	Fracture classification (lower body CT, 28-class)	Multi-label classification; integrating the Google inception module; 2D segmentation projection as input
Meng et al. (2021) [48]	Rib fracture classification (chest CT, 4-class)	Multi-stage system; 3D input by resampling; self-designed 3D CNN with four conv layers
Papandrianos et al. (2020) [25]	Bone metastasis classification (whole body CT, 2-class)	Compare performance of VGG16, ResNet50, GoogLeNet, mobile Net with the proposed CNN model (best performance).
Lin et al. (2021) [27]	Bone metastasis classification (whole body CT, 2-class)	Data preprocessing to crop thoracic region; comparing performance of VGG, ResNet and DenseNet
Wu et al. (2020) [26]	Traumatic osteomyelitis classification (bone CT, 2-class)	ResNet as backbone; retrospective study

for the model training and evaluation. The segmentation model training from 2D axial slices directly achieved the best performance, which was 91.0% of intersection-of-union (IOU) in the private dataset and 85.0% in the public dataset. Noguchi et al. [67] also employed the U-Net segmentation model for bone segmentation on the same public dataset as [66]. Three data augmentation methods were used: conventional rotation, zooming, flipping, and shear transformation; mixup; and random image cropping and patching (RICAP). The combination of the conventional method and RICAP method data augmentation achieved the best result, an IOU of 92.6% on the public dataset.

Although U-net has demonstrated high performance, the other segmentation models have also been investigated for bone segmentation. Egger et al. [68] used the FCN to develop a lower jawbone (mandible) segmentation model using ten skull CT for training and ten skull CT for testing. VGG16 was utilized as the backbone, and three different networks, FCN-32s, FCN-16s, and FCN-8s, were used to segment the mandible region. The FCN-8s outperformed the other two models and achieved the best Dice coefficients of 92.03% on the training set and 89.64% on the testing set. Lin et al. [104] built several deep learning models, including U-net, U-net with a residual module (U-Net-Res), Mask R-CNN, Mask R-CNN with spatial attention module (Mask R-CNN-Att), for the hotspots segmentation of bone metastasis in SPECT scans. 112 CT samples were selected, and with the data augmentation methods of image mirror, translation, rotation, 2280 samples were generated. 1830 samples and 450 samples were used as training and testing groups, respectively. The U-Net-Res model outperformed the other three models with an IOU of 61.03%, and the U-net model (IOU of 59.41%) was better than the Mask R-CNN model (IOU of 55.44%) and Mask R-CNN-Att model (IOU of 54.27%).

Except for the single slice-based approaches, some works explored the multiple slices as input for bone segmentation. Li et al. [81] combined the Attention U-net (AttU-Net) and bi-directional convolutional long short-term memory (BDC-LSTM) model for the tooth roots segmentation from CBCT scans. The AttU-Net gave large weight to the tooth region via attention gates between the downsample feature and upsample feature, and the LSTM was used to extract the intra-slice and inter-slice contexts between the tooth root sequence. Twenty-four scans were used to validate the proposed method, and the model achieved an average IOU of 91.4% on five testing scans. Léger et al.

[106] used the 3D U-net for the segmentation of mineralized cartilage in high-resolution micro-CT images. Due to the high resolution of the CT images, two strategies, the downsampling method or the 3D-patch method, were applied to avoid the GPU memory limitations. The 3D U-Net with downsampling strategy outperformed the 3D-patch strategy on six testing CT scans and showed better 3D consistency than 2D U-net.

Dreizin et al. [100, 111] developed a recurrent saliency transformation network for the pelvic hematoma segmentation with the coarse-to-fine strategy. Two identical fully convolutional neural networks (FCNs) were used in the coarse and fine stages, respectively. A coarse segmentation map was generated in the coarse stage to localize the region of interest using a saliency transformation module. The cropped region was fed into the fine-segmentation FCN model to generate the more accurate segmentation results of pelvic hematoma. Both the coarse-FCN and fine-scale-FCN were optimized by the Dice loss function. 253 trauma CT scans were used to train and test the model, and the proposed method achieved a Dice score of 0.71 compared to 0.49 of the 3D U-Net model.

Identifying regions of different bones or disease lesions from the CT scan is an essential step for the medical diagnosis, and this requires the multi-class segmentation methods [40, 75, 76, 84, 92, 98, 107].

Liu et al. [98] implemented a multi-class segmentation model of pelvic CT scans based on the U-Net. The target organs included the left femoral head, right femoral head, spinal cord, bladder, bone marrow, rectum, and small intestine. A weighted cross-entropy loss was used to train the model to overcome the class imbalance problem. The weight was set according to the sizes of different segmented parts. 105 patients CT scans (77 for training, 14 for validation, and 14 for testing) were collected. The Dice similarity coefficient was 90.6%, 90.0%, and 82.7% for the left femoral head, right femoral head, and spinal cord, respectively, which outperformed the standard U-Net training method on the test set. Liu et al. [40] trained a multi-class network for pelvic bones segmentation more scans (1184 CT scans). A cascaded 3D U-Net was utilized. The first U-Net generated the low-resolution 3D segmentation results by training from the down-sampling data. The second U-Net model was trained on full-size CT data and the up-sampling segmentation results from the first U-Net data

model. A postprocessing step based on signed distance function [112] was employed to create a robust result for clinical usage. The average segmentation Dice score is 98.7% on the metal-free CT scans, and the SDF post-processor gained a decrease of 15.1% in Hausdorff distance compared with the maximum connected region post-processor. Uemura et al. [107] applied the Bayesian U-Net model [113] for the segmentation of intensity calibration phantom of Pelvic CT. Forty scans from two CT machines were selected to train the Bayesian U-Net model and data augmentation methods like intensity normalization, translation, rotation, scaling, and shear transformation were used during training. The Dice coefficient was 97.7% on 1000 testing CT scans which demonstrated the excellent and robust performance of the proposed segmentation method.

The multi-stage methods also delivered satisfying results in other multi-class segmentation tasks. Belal et al. [92] proposed a three-step CNN-based approach for the segmentation of 49 bones in the upper body CT scans. The first step used a CNN to detect the 29 landmarks of rib joints and vertebral, employed the active shape model to determine the landmarks identity, and utilized another network to detect the rib centerlines. The second step fed the original CT slices and the identified landmarks into another CNN model to generate a raw voxel-wise segmentation. The last step used several postprocessing methods such as connect component analysis and binary hole filling to modify the spurious voxels. The proposed method achieved satisfying qualitative results, and the Dice coefficient on five selected bones (Th7, L3, sacrum, right 7th rib, and sternum) was between 83.0% $\sim$ 88.0% on five testing cases.

The temporal bone is a complicated structure. Fauser et al. [75] proposed a 2D U-Net-based method for the structure segmentation of temporal bone CT instead of the 3D U-Net approach considering the scarcity of available annotation data. Three U-Net models were trained using the axial-view, sagittal-view, and coronal-view slices of temporal bone CT data, respectively. An initial 3D segmentation of the temporal bone structure was generated by majority voting, and a probabilistic active shape model was employed to refine the segmentation results. The proposed method achieved better organ shapes on 24 testing CT data of real patients than the existing semi-supervised methods. Li et al. [76] designed a 3D Deep Supervised Densely Network (3D-DSD Net) by extending the 3D-U-Net for the small organs segmentation of temporal bone

CT. Three densely connected blocks were used to extract the low-level features in the encoder part, and these features were transferred to the decoder via a multi-pooling feature fusion strategy. The decoder generated the segmentation feature maps on both the last layer and the two hidden layers. A joint loss between each layer in the decoder was designed to train the 3D-DSD net with 56 CT scans. Eight scans were used for testing, and the average Dice coefficient over the nine organs of the temporal bone was 77.18%.

Another challenging task in head-related CT is tooth segmentation. Cui et al. [84] proposed ToothNet, a two-staged network for tooth instance segmentation from CBCT scans. In stage one, an edge map extraction network with a deep supervised scheme was designed to detect the tooth edges. The second stage fed both the edge maps and original CBCT slices into a 3D region proposal network and employed a similarity matrix to discard redundant proposals. The region proposals were used for four tasks, tooth segmentation, classification, 3D box regression, and identification, and the whole model was trained using the joint loss of the four tasks. Twelve scans were used to train the proposed model, and the average Dice coefficient was 92.37% over eight testing scans.

The deep learning-based bone CT segmentation tasks have achieved high performance with enough annotation data. However, manually annotating is laborious and time-consuming. Using both the labeled and unlabelled data for bone segmentation is a potential solution. Malinda et al. [21] constructed a hybrid deep segmentation generative adversarial network (Hybrid-SegGan) for the lumbar vertebrae CT segmentation to utilize both labeled and unlabeled data. The proposed Hybrid-SegGan was based on CycleGAN [114] while a two-cycle consistency strategy was used instead of the one cycle consistency in CycleGAN. The adversarial loss, consistency loss, constraint loss, and identity loss that derived from four discriminators and two generators were used to optimize the model to determine if the CT data and segmentation data were paired or not. The proposed model achieved an IOU of 99.1% compared to 98.7% of U-Net in 120 CT scans.

Overall, the segmentation tasks acted as a fundamental task for the bone CT analysis. Table 2.3, Table 2.4, Table 2.5, and Table 2.6 listed the details of the

segmentation tasks on head, spine, pelvic, and other CT, respectively. The U-net model was the most commonly used model for two-classes segmentation tasks due to the unique skip connection structure of U-net and the special Hounsfield scale of bones in CT. The multi-stage framework usually outperformed the single-stage methods for the multi-classes segmentation tasks.

Table 2.3 Segmentation task details on head-related CT.

Reference	Application	Method;remarks
Zhang et al. et al. (2017) [71]	Craniomaxillofacial segmentation (head CT, 2-class)	Two cascaded FCN; first FCN generates the displacement map; second FCN generates both segmentation map and landmark heatmaps
Egger et al. (2018) [68]	Mandible segmentation (head CT, 2-class)	VGG16 for feature extraction; FCN-32s, FCN-16s and FCN-8s for segmentation
Torosdagli et al. (2019) [74]	Mandible segmentation (head CT, 2-class)	Fully convolutional DenseNET with 103 conv layers for segmentation
Cui et al. (2019) [84]	Tooth segmentation (head CT, 40-class)	Two-stage framework; edge map extraction CNN; using edge maps and slices to generate to generate region proposals for segmentation and classification
Ezhov et al. (2019) [85]	Tooth segmentation (head CT, 33-class)	Multi-stage framework; coarse weakly supervised segmentation; fine-tuning on downsampled masks; V-Net as backbone
Zakirov et al. (2019) [22]	Tooth segmentation (head CT, 33-class)	V-Net for tooth segmentation; instance normalization before each convolution instead of batch normalization
Jaskari et al. (2020) [82]	Tooth segmentation (head CT, 2-class)	3D fully convolutional network with the U-Net structure; training patches extracted from coarsely-annotated CTs

Li et al. (2020) [81]	Tooth segmentation (head CT, 2-class)	Attention Unet (AttU-Net) for large weight on tooth region; bi-directional LSTM to extract intra-slice and inter-slice contexts
Setzer et al. (2020) [83]	Tooth periapical lesion segmentation (head CT, 5-class)	U-Net model as backbone; five slices as input; 3D convolution layer to predict the segmentation of center slice
Fauser et al. (2019) [75]	Temporal bone segmentation (head CT, 11-class)	2D U-Net for data scarcity; three U-Net models from three views of the scan; majority voting and probabilistic active shape model as postprocessing
Li et al. (2020) [76]	Temporal bone segmentation (head CT, 10-class)	3D Deep Supervised Densely Network; 3D-U-Net as backbone; multi-pooling feature fusion
Neves et al. (2021) [77]	Temporal bone segmentation (head CT, 5-class)	Compare Anisotropic Hybrid Network (AH-Net) [115], U-Net and ResNet; evaluation considering testing time
Nikan et al. (2020) [80]	Temporal bone segmentation (head CT, 9-class)	3D fully connection neural network; subsampling strategy for class imbalance;
Nikan et al. (2020) [78]	Temporal bone segmentation (head CT, 9-class)	Patch-wise densely connected three-dimensional network; B-spline upsampling for decoder; balanced-weighted patch sampling
Wang et al. (2021) [79]	Temporal bone segmentation (head CT, 4-class)	W-net with two analysis paths and two synthesis paths; weighted cross entropy loss
Lee et al. (2019) [73]	Orbital bone segmentation (Head CT, 2-class)	Derived from U-Net model; orbital region cropping; thin bone segmentation branch; cortical bone segmentation branch

Huang et al. (2020) [69]	Skull bone segmentation (head CT, 2-class)	Multiphase CT as input; 3D U-Net model and multi-channel atrous CNN for segmentation
Matzkin et al. (2020) [70]	Skull retracting segmentation (Head CT, 2-class)	Three methods proposed; autoencoder or U-net model for first reconstruct then subtract; U-Net directly estimated bone flap (best)
Lian et al. (2020) [72]	Jaw segmentation (head CT, 2-class)	A U-Net shaped structure using adaptive transformer module for feature extraction

Table 2.6 Segmentation task details on whole body, shoulder, chest, upper body, hand, femur and bone CT.

Reference	Application	Method;remarks
Chen et al. et al. (2017) [101]	Femur segmentation (femur CT, 2-class)	3D-U-Net as backbone; edge detection task embedded into the feature extraction; multi-task loss function
Klein et al. (2018) [66]	Whole body bone segmentation (whole body CT, 2-class)	U-Net model using three training methods; training from 2D axial slices; from axial, sagittal, and coronal slices; from 2D axial slices via an unsupervised pretraining model
Noguchi et al. (2020) [67]	Whole body bone segmentation (whole body CT, 2-class)	U-Net as backbone; Three data augmentation methods; conventional method; mixup; random image cropping and patching
Belal et al. (2019) [92]	Upper body bone segmentation (chest CT, 50-class)	Multi-step approach; CNN for landmarks detection; active shape model for labeling landmarks; patch-CNN model for segmentation; postprocessing

Song et al. (2019) [103]	Bone metastasis segmentation (bone CT, 2-class)	VGG16 as backbone; two side connections to extract contour and global feature;
Lin et al. (2020) [104]	Bone metastasis segmentation (whole-body CT, 2-class)	Compare several models, U-Net, U-Net with residual module (U-Net-Res), Mask R-CNN, Mask R-CNN with spatial attention module (Mask R-CNN-Att)
Moreau et al. (2020) [105]	Bone metastasis segmentation (upper-body CT, 3-class)	3D U-Net as backbone; combine cross entropy loss and multi-class Dice loss as loss function
Léger et al. (2019) [106]	Mineralized tissue segmentation (cartilage and bone CT, 2-class)	3D-U-Net as backbone; downsampling or 3D patch as input to save GPU memory
Chen et al. et al. (2021) [102]	Bone segmentation (bone CT, 3-class)	Recurrent CNN with eight conv layers; unsupervised and semi-supervised loss based on fuzzy C-means
Folle et al. (2021) [96]	Metacarpal bone segmentation (hand CT, 2-class)	Compared 2D U-Net with pretraining (best), 2D U-Net without pretraining, 3D U-Net without pretraining
Meng et al. (2021) [48]	Rib segmentation (Chest CT, 2-class)	V-Net for bone segmentation; dynamic programming algorithm for labelling
Zhang et al. (2021) [117]	Rib segmentation (Chest CT, 2-class)	Two cascaded CNN framework, Foveal network for segmentation, Faster R-CNN for detection; three types of evaluation, human only, deep learning only, human with deep learning assistance;
Sanghani et al. (2021) [87]	Clavicle bone segmentation (clavicle CT, 2-class)	U-Net model as backbone; 2D slice as input

Taghizadeh et al. (2021) [86]	Shoulder segmentation (shoulder CT, 2-class)	U-Net for segmentation; data augmentation; five-fold validation
Uemura et al. (2021) [107]	Phantom segmentation (Femur CT, 5-class)	Bayesian U-Net model for segmentation; data augmentation of translation, rotation, scaling and shear transformation

2.3 Regression Tasks for Bone CT Analysis

The purpose of the regression tasks is to map the input CT slices into one or several continuous outputs according to the definition of task. The regression tasks like age estimation [118], bone mineral density estimation [94, 95, 119, 120], skeleton landmarks detection [19, 71, 72, 74, 121–128], bone fracture detection [48, 117], and slice position detection [57] have been explored using the deep learning methods recently. Different models such as the self-designed neural network [119], VGGNet [129], DenseNet [74, 94], LSTM [74, 124], FCN [71, 74, 123, 124] have been used for the regression tasks.

Directly fed the scan into a CNN model to yield the results was used in some regression tasks. González et al. [119] designed a one-stage regression neural network for bone mineral density (BMD) estimation. The model consisted of three convolutions layers, one fully connected layer, and one output layer. The 3D upper body CT scans were projected as 2D slices as the model input. 9925 CT scans were used to train, validate and test the proposed model and the root mean squared error was applied to optimize the model. The correlation coefficient between the predicted BMD and the real BMD was 94.0% on the 1000 testing scans. Nguyen et al. [118] developed an age assessment system using the whole body bone CT scans. The whole body CT scan was reduced to a 2D front view image and was fed into a modified VGGNet, where the features from different layers were concatenated to incorporate both the high-level and low-level features for age assessment. 569 CT scans with the age range from eight months to 87 years were used to train the model. The mean square error (MAE) on 244

Table 2.4 Segmentation task details on spinal CT.

Reference	Application	Method;remarks
Novikov et al. (2017) [17]	Vertebrae segmentation (spinal CT, 2-class)	Integrate bidirectional convolutional LSTM into U-Net-like architecture
Fan et al. (2019) [18]	Vertebrae and nerve segmentation (spinal CT, 3-class)	3D-U-Net as backbone; resampling, cropping, and intensity normalization as preprocessing
Krishnaraj et al. (2019) [95]	Vertebrae bone segmentation (Spinal CT, 4-class)	Two cascaded U-Net models to identify L1 $\sim$ L4 vertebrae bones; first U-Net on sagittal view; second U-Net on sagittal and coronal view
Bae et al. (2020) [90]	Vertebrae segmentation (spinal CT, 2-class)	U-Net model; two-stage post-processing; mislabelling error correction for first stage; separate each vertebra part by identifying the separation points at second stage
Fan et al. (2020) [20]	Vertebrae segmentation (spinal CT, 4-class)	3D-U-Net for segmentation of nerve, bone, disc and background; resampling, cropping, and intensity normalization as preprocessing
Malinda et al. (2020) [21]	Vertebrae segmentation (spinal CT, 2-class)	hybrid deep segmentation based on CycleGan; two-cycle consistency strategy
Pan et al. (2020) [91]	Vertebrae segmentation (Spinal CT, 4-class)	3D-Unet for segmentation; Treat T1 $\sim$ T6, T7 $\sim$ T12, L1 $\sim$ L2 as three categories for segmentation
Rehman et al. (2020) [93]	Vertebrae segmentation (spinal CT, 2-class)	Combine region-based level set and U-Net model for the segmentation; fracture CT scans for segmentation
Fang et al. (2021) [94]	Vertebrae segmentation and BMD estimation (spinal CT, 5-class)	U-Net model for four lumbar vertebral segmentation
Löffler et al. (2021) [88]	Vertebrae segmentation (spinal CT, 4-class)	Multi-stage framework; fully CNN to detect spine;, a butterfly-shaped CNN [116] for labelling; U-Net based model to segment the vertebrae patches
Suri et al. (2021) [89]	Vertebrae segmentation (spinal CT, 2-class)	Multi-stage approach; Feature generation network for feature extraction; Region recognition network for bounding box proposal; segmentation network to refine results

Table 2.5 Segmentation task details on pelvic CT.

Reference	Application	Method;remarks
Dreizin et al. (2020) [111, 100]	Pelvic hematoma measurement (pelvic CT, 2-class)	Recurrent saliency transformation network; coarse-to-fine segmentation; FCN as backbone
Hemke et al. (2020) [99]	Pelvic segmentation (pelvic CT, 6-class)	U-Net as backbone; data augmentation to enlarge training set
Liu et al. (2021) [40]	Pelvic segmentation (pelvic CT, 5-class)	Cascaded 3D U-Net; first U-Net on low-resolution CT; second U-Net on full-size CT; postprocessing as signed distance function
Liu et al. (2020) [98]	Organs at risk segmentation (pelvic CT, 8-class)	U-Net as backbone; weighted cross-entropy loss to overcome class imbalance
Sánchez et al. (2020) [97]	Bone segmentation (Abdomen and pelvic CT, 2-class)	3D-Unet for segmentation; 5-fold validation; low- and high-energy CT as input

testing scans of the proposed model was 4.856 years which outperformed the VGGNet (9.741 years), GoogLeNet (5.522 years), and ResNet (5.738 years).

The multi-stage or multi-task frameworks were also explored for the regression tasks on complicated body structures, such as the spine and head. Fang et al. [94] proposed a two-stage regression framework for the BMD estimation from spine CT. A U-Net model was used to identify the four lumbar vertebral bones from the sagittal view of the spinal CT. The extracted vertebral bones were then fed into the DenseNet-121 model for BMD estimation. The proposed framework was trained on 586 CT cases and tested on three cohorts with 463 scans, 200 scans, and 200 scans, respectively. The Dice coefficients of the lumbar vertebral bones for the three cohorts were 82.3%, 78.6%, and 78.2%, respectively. The average BMD estimation results were highly correlated ($r>0.98$) with the ground truth. Liao et al. [124] combined the 3D FCN and bi-directional Recurrent neural network (Bi-RNN) for localization and identification of the vertebrae CT scans. A multi-task 3D FCN was trained to extract the features of vertebral samples and encode the short-range contextual information. The Bi-RNN then learned the long-range contextual information of the vertebral anatomic structure from the feature maps of the 3D FCN and outputted the identification and localization

results of the CT scan. The overall identification performance of the proposed model was 88.3%, and the average localization error was 6.47mm.

Zhang et al. [71] also used the multi-stage strategy and developed a context-guided multi-task fully convolutional networks for craniomaxillofacial landmarks detection and bone segmentation simultaneously. The proposed network consisted of two FCNs, and both adopted the U-Net structure. The first FCN generated the displacement maps of each voxel to the 15 landmarks as a guidance of the spatial context information. The second FCN combined the displacement maps and original CT data as input and generated the bone segmentation map, and estimated the position of the landmarks. 107 CT scans from two medical centers were used to train and evaluate the proposed method with a 5-fold cross-validation strategy, and the landmark digitization error was around 1.10 mm, which was superior to the compared methods. Torosdagli et al. [74] proposed a three-step framework to detect the mandible landmarks with three neural networks. The first step employed the fully convolutional DenseNet [130] to generate the linear time distance transform (LTDT) of the mandible bone, and the LTDTs were transformed as a combined geodesic map of five mandibular landmarks via a U-Net model. To detect the remaining four landmarks, a LSTM network was applied according to the detection position of the menton landmark from the U-Net model. A CBCT dataset with 250 patient CBCT was used to evaluate the proposed method qualitatively. Two experts gave scores of the segmentation results while only 5% of the segmentation results were scored as unacceptable by both two experts.

Table 2.7, Table 2.8 and Table 2.9 listed the details of the recent regression tasks, including the applications, methods details, and remarks of head, spine and other CT, respectively. The results demonstrated the effectiveness of the deep learning methods in bio-marker estimation and landmarks detection. During the regression system design, the segmentation model was often utilized within a multi-stage or multi-task framework to deliver better performance.

Table 2.7 Regression task details on head CT.

Reference	Application	Method;remarks
Zhang et al. et al. (2017) [71]	Craniomaxillofacial landmarks detection (skull CT, 15 landmarks)	Two cascaded FCN; first FCN generates the displacement map; second FCN generates both segmentation map and landmark heatmaps
Torosdagli et al. (2019) [74]	Mandible landmarks detection (head CT, 2-class, 9 landmarks)	Multi-stage framework; fully convolutional DenseNet for segmentation; U-Net to detect five landmarks; LSTM to detect another five landmarks
Lian et al. (2020) [72]	Mandible landmarks detection (head CT, 64 landmarks)	Multi-stage framework; global context feature from down-sampled CT; fine-grained features from patch; adaptive transformer module for feature extraction
Yun et al. (2020) [122]	Skull landmarks detection (head CT, 93-landmark)	Multi-stage framework; global and local slice for coarse and fine detection; VAE-based latent representation of landmarks
Xiao et al. (2021) [128]	Skull reference bony shape estimation (head CT)	Encoder-decoder model derived from PointNet++; synthesis data from normal bones; displacement vectors estimation

Table 2.8 Regression task details on spine CT.

Reference	Application	Method;remarks
Suzani et al. (2015) [127]	Vertebrae landmark detection (spine CT, 26 landmarks)	Multi-stage approach; six-conv-layer CNN; centroid estimated and refinement by kernel density estimation method
Netherton et al. (2020) [19]	Vertebrae landmark detection (spine CT, 26 landmarks)	X-Net for landmark detection; sagittal and coronal intensity projection as two inputs and output corresponding detection map
Pisov et al. (2020) [121]	Vertebrae landmark detection (spine CT, 6-landmarks)	Multi-stage approach; spine straightening via 3D U-Net architecture; YOLOv3 based vertebrae-level prediction
Cai et al. (2016) [126]	Vertebrae bounding box detection (spine CT)	Multi-stage approach; MRI and CT as input; feature extraction by CRBM; SVM for bone labelling
Belharbi et al. (2017) [57]	Vertebrae slice detection (lumbar spinel CT)	Transfer learning based on VGG16; project the scan through sagittal view as a 2D maximum intensity projection
Liao et al. (2018) [124]	Vertebrae identification and localization (spinal CT, 27-landmark)	Multi-stage framework; 3D CNN for vertebrae sample; FCN for centroid detection or coarse segmentation; bidirectional RNN for fine centroid detection and labelling
Krishnaraj et al. (2019) [95]	Vertebrae labelling and BMD estimation (Spine CT)	Multi-stage approach; two U-Net model for vertebrae labelling, linear regression for BMD estimation
Fang et al. (2021) [94]	Vertebrae segmentation and BMD estimation (spine CT, 5-class)	Multi-stage framework; U-Net model for four lumbar vertebral segmentation; DenseNet-121 for regression
Jakubicek et al. (2020) [125]	Spine centerline detection (spine CT)	Two-stage method;, spine-ends detection by AlexNet; Faster R-CNN for spine centerline detection
Meng et al. (2021) [48]	Rib fracture region detection (spine CT)	Modified V-Net with two ResNet module and a bottleneck module
Yasaka et al. (2020) [120]	Lumbar BMD estimation (Spine CT)	Self-designed CNN with four conv layers and three fully connection layer; manually cropped lumbar region; data augmentation

Table 2.9 Regression task details on whole body, upper body, and chest CT.

Reference	Application	Method;remarks
González et al. (2018) [119]	BMD estimation (upper body CT)	Self-designed CNN with three convolution layers; 3D scans projected as 2D images
Nguyen et al. (2019) [118]	Age estimation (Whole body CT)	VGGNet as backbone; whole body CT for age estimation; 3D scans projected as 2D images from front view
Thies et al. (2020) [129]	CBCT Source trajectories adjustment (chest CT)	VGGNet as regressor; data augmentation with rotation; experiments on both simulation and real dataset
Zhang et al. (2021) [117]	Rib fracture detection (Chest CT)	Two cascaded CNN framework, Foveal network for segmentation, Faster R-CNN for detection; three types of evaluation, human only, deep learning only, human with deep learning assistance

2.4 Generative Tasks for Bone CT Analysis

Different from the discriminative tasks like classification, segmentation, and regression, the generative tasks in bone CT analysis aim to generate new data instance from the input CT data. Tasks such as noise reduction [131], cross-modality image synthesis [132–137], CT super resolution reconstruction [138–140], scatter correction [141], metal artifact reduction [142–145] have been studied. U-net [133, 134, 136, 138, 141], GAN and its variations [132, 135, 139, 140, 144], and self-designed CNN [131, 142, 143, 145], were used in the generative tasks.

The U-net model has demonstrated its performance in other tasks, and some researchers investigated the U-net structure for generative tasks. Park et al. [138] applied the U-net to learn the mapping criteria between the low and high-resolution CT head scans. The low-resolution CT slice was generated by averaging the existing high-resolution CT slices, and the ground truth was set as the middle slice of the corresponding high-resolution slices. With 52 head scans as the training set, the predicted high-resolution slices were not only virtually equivalent to the ground truth but also achieved low noise with 10% higher peak signal-to-noise ratio and more clear boundaries of the bone structures. Liu et al. [133] developed a deepAC model for the

generation of pseudo-CT scan from the non-attenuation-corrected (NAC) PET images. A convolutional encoder-decoder architecture with short connection was employed to generate the corresponding pseudo-CT from the NAC PET data. A set of 100 PET head scans was used for training, and the model yielded a mean error of 111 HU between the pseudo-CTs and the real CT on 28 evaluation scans.

The GAN-based models were also used for the generative tasks. Hiasa et al. [132] employed the CycleGAN to generate CT images from MRI data by adding a gradient consistency loss to improve the accuracy at boundaries. A CT generator, MRI generator, CT discriminator, MRI discriminator were designed, and four losses, including the adversarial loss between CT data, adversarial loss between MRI data, the cycle consistency loss between MRI and CT data, and the gradient consistency loss, were used to train the whole model. The mean absolute error (MAE) and peak-signal-to-noise ratio (PSNR) between real CT and synthesized CT and mutual information (MI) between synthesized CT and real MRI were used for quantitative evaluation. The U-net segmentation results using the synthesis CT were treated as quantitative evaluation. Both results demonstrated the efficiency of the proposed CycleGAN-based CT synthesis model. You et al. [139] presented a semi-supervised method for the super-resolution reconstruction of Tibia micro-CT and abdominal CT. They adopted the CycleGAN framework with two generators, two discriminators, and a residual CNN-based network to reserve the CT details. The CycleGAN model was jointly trained with the adversarial loss, cycle consistency loss, identity loss, and a joint sparsifying transform loss. The results on the tibia micro-CT and the abdominal CT reflected both qualitative and quantitative superior results compared to the other models.

Some works designed unique CNN models according to the specific characteristics of the generative task and usually achieved good performance. Chen et al. [131] developed three layers convolutional neural network for noise reduction in low-dose CT. The three layers acted as patch coding, non-linear filtering, and reconstruction, respectively. 200 normal-dose as input and corresponding low-dose CT slices as ground truth were used to train the model. The low-dose images were generated by imposing Poisson noise on the normal-dose images. The PSNR, root mean square error (RMSE), and structural

similarity index (SSIM) on 100 testing slices were better than the non-deep learning methods, such as K-SVD [146], and BM3D [147].

Zhang et al. [143] developed a CNN-MAR for metal artifact removal on the tooth CT and abdominal CT. They established a pseudo metal artifact database by simulating the metal artifacts of metal using the sinogram. In the first stage, the pseudo-metal-inserted, metal-free, and pre-corrected CT images from linear interpolation (LI) and beam hardening correction (BHC) [148] were used to train a CNN model with five convolutional layers. Then, the water and equivalent tissues in the output of the CNN model were further assigned with a uniform value to obtain a CNN prior image. At last, the forward projections of the CNN prior image, the original CT, and the metal-only image were jointly used to remove the metal artifact. The model demonstrated both remarkable performances of RMSE and SSIM index on the simulated data and qualitatively satisfying results on 16 real CT data. Lin et al. [142] proposed a Dual Domain Network (DuDoNet) with a sinogram enhancement module, an image enhancement module, and a radon inversion layer for metal artifact removal on the abdominal CT. The sinogram enhancement module restored the sinogram of metal-free CT from the linear interpolation results via a mask pyramid U-Net. To further reduce the second artifact in the metal-free CT, a radon inversion layer with radon consistency loss was designed between the generated metal-free CT from sinogram and ground truth. The image enhancement module further refined the metal-free image by applying a U-net architecture. The average PSNR of the DuDoNet on the large metal parts, small metal parts, and all metal parts of 2000 simulated abdominal metal CT was 29.02, 36.72, and 33.51, respectively.

The generative tasks were widely used for CT data generation and preprocessing on bone CT analysis. Table. 2.10 and Table. 2.11 listed the details of recent generative tasks in head and other CT, respectively. Both supervised and unsupervised methods were used for the generative tasks, and the synthesized data were often applied for model training due to the lack of ground truth. The results showed that the deep learning methods were effective on the generative tasks for bone CT analysis.

Table 2.10 Generative task details on head CT.

Reference	Application	Method;remarks
Liu et al. (2018) [133]	PET to CT (head CT)	Encoder-decoder architecture with short connections that similar to U-Net model
Koike et al. (2020) [137]	Virtual non-contrast CT generation (Head CT)	CNN with densely connection; virtual non-contrast CT training data generated from contrast-enhanced CT
Yong et al. (2021) [136]	QCBCT generation (Head CT)	Cycle-Gan model to generate QCT from CBCT; multi-channel U-Net for QCBCT generation

2.5 Discussion and Conclusion

An extensive analysis of the recent research progress of deep learning-based bone CT analysis was performed in this chapter. Various kinds of tasks, such as segmentation, classification, regression, and generative tasks, have been discussed. We summarized the bone CT applications (Fig. 2.1a), the research trends until 2021 (Fig. 2.1b), the proportion of different tasks (Fig. 2.1c) and the proportion of different scanning parts (Fig. 2.1d) in Fig. 2.1. Nearly all body parts have related deep learning applications (Fig. 2.1a) while head and spine contributed more than half of the research (Fig. 2.1d). Half of the recent researches were segmentation tasks (Fig. 2.1c) on account of segmentation being the foundation of bone CT analysis.

The results revealed that deep learning methods have demonstrated excellent performance in various bone CT analysis tasks and have shown great potential to assist clinical practice. Both single-stage deep learning methods and multi-stage deep learning methods have reported good performance on specific tasks. However, for complicated tasks like vertebra labeling and segmentation, temporal bone segmentation, skull landmark detection, fracture classification, and BMD estimation, a multi-stage hybrid deep learning framework could achieve better results. The preprocessing, data augmentation, and postprocessing could also improve the performance and robustness of the deep learning approach.

Enough annotation data is required to train, validate and evaluate the deep learning-based bone CT analysis system. The data annotation is usually a laborious and

Table 2.11 Generative task details on chest, spinal, abdominal, pelvic, hip, lower body, femur and bone CT.

Reference	Application	Method;remarks
Chen et al. (2017) [131]	Noise reduction (Chest and abdominal CT)	Self-designed CNN with three conv layers; low-dose images generated by Poisson noise and ray-driven algorithm
Park et al. (2018) [138]	Super resolution reconstruction (chest and abdominal CT)	U-Net model as backbone; training data generated by average the high resolution slices
You et al. (2019) [139]	Super resolution reconstruction (tibia and abdominal CT)	CycleGAN model as backbone; joint constraints in loss functions; cycle-consistency in Wasserstein distance
Guha et al. (2020) [140]	Super resolution reconstruction (ankle CT)	CycleGAN model as backbone; real low-resolution and corresponding high-resolution data for model training and testing from two different CT scanners
Hiasa et al. (2018) [132]	MRI to CT (femur CT)	CycleGAN model as backbone; gradient consistency loss to improve the accuracy at boundaries; U-Net segmentation results for quantitative evaluation
Leynes et al. (2018) [134]	MRI to CT (spinal and pelvic CT)	U-Net as backbone; model trained with gradient difference loss; combining difference loss and L_1 loss
Nomura et al. (2019) [141]	Scatter correction (bone phantom CBCT)	U-Net based 25-layer CNN; Monte Carlo simulation of dataset; mean absolute error and mean squared error as loss function
Zhang et al. (2018) [143]	Metal artifact reduction (abdominal CT)	Multi-stage approach; integrate the LI and BHC result for metal artifact reduction; simulated data; self-designed CNN with five conv layers
Liao et al. (2019) [144]	Metal artifact reduction (abdominal CT)	Unsupervised learning; disentangle the metal artifact in latent space; testing on both simulated dataset and clinical dataset
Lin et al. (2019) [142]	Metal artifact reduction (abdominal CT)	Metal artifact in both sinogram image and CT image; radon inversion layer; simulated data for model training
Qi et al. (2020) [145]	Metal artifact reduction (hip CT)	CNN with six conv layers to remove metal artifact from sinogram space; simulated data using real hip image with and without metal
Kawahara et al. (2021) [135]	Dual-energy CT to kilovoltage CT (pelvic CT)	Conditional GAN model; L1 norm loss and discriminator loss

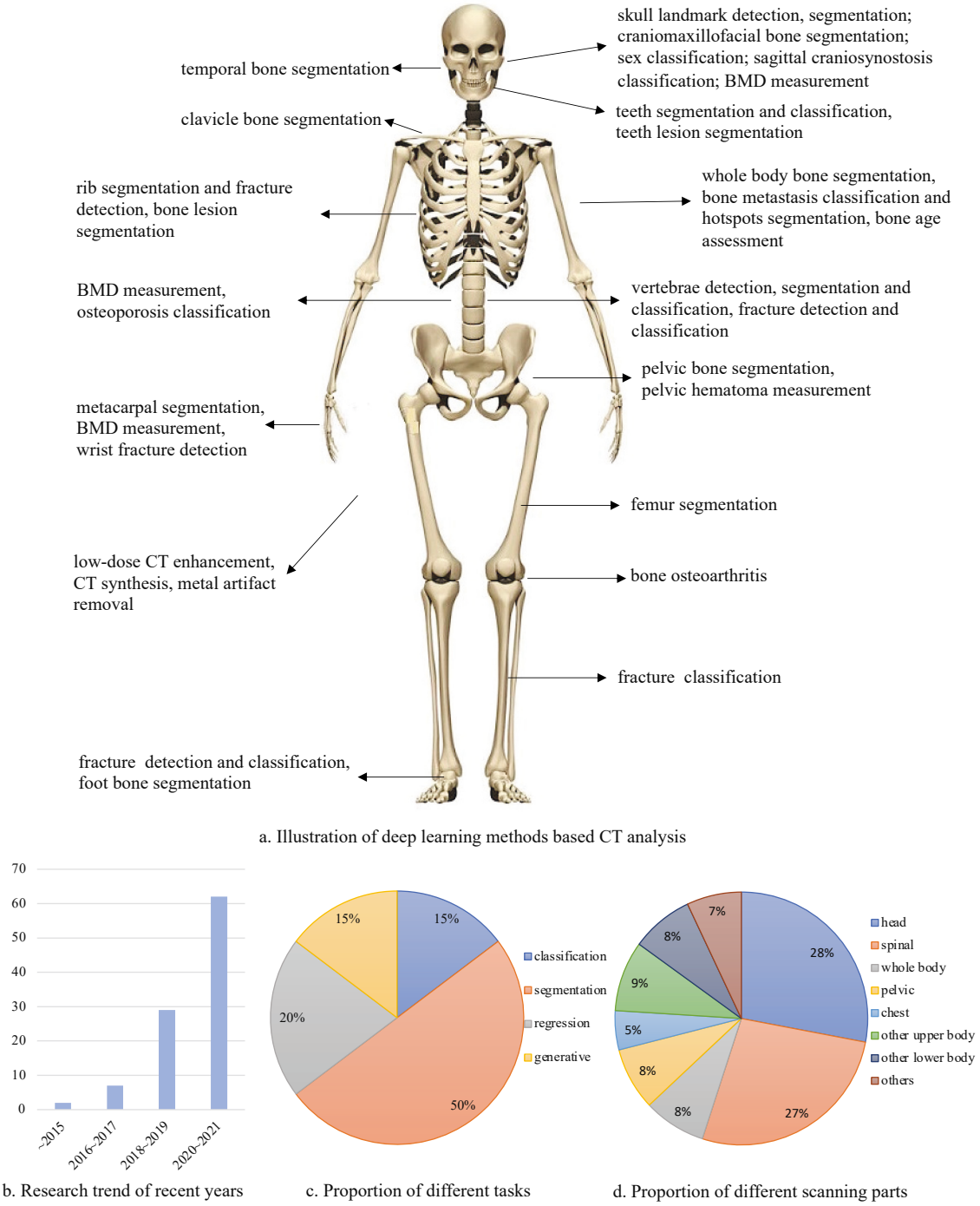

Fig. 2.1 Illustration of applications, research trends, and tasks proportion of deep learning based bone CT analysis.

time-consuming process, especially for the segmentation tasks, where the pixel-wise annotation is needed. The semi-supervised method could be a potential way to accelerate the annotation procedure. The semi-supervised methods in [40–42] have reduced the complexity for data labeling of wrist, spine, and pelvic CT. More research could be conducted on the multi-classes segmentation tasks of other bone CT scans to facilitate the annotation procedure. Besides, since collecting the medical data is usually more difficult than nature images due to privacy, the public medical datasets containing bone CT, such as [149–151], could be a potential source to establish the bone database and raise new research tasks. Since the current public datasets mainly focused on segmentation and detection-related tasks, the public CT dataset about disease detection and classification was still unseen. This could be a potential direction for the construction of future public datasets.

Most of the existing studies utilized the existing deep learning models like VGGNet, ResNet, DenseNet, U-Net, and FCN, as the backbone. Recently, other effective models have emerged, such as the transformer [152, 153]. Introducing transformers for the bone CT analysis could be a direction to extract more discriminative features and improve the performance. Besides, in the classification, segmentation, or regression tasks, an issue that cannot be ignored during the model training is the data imbalance. For example, the bone sizes are different in segmentation tasks, the majority of cases in the collected cohorts are healthy cases for the classification tasks, and the BMD values are usually clustered in a range for regression tasks. Resampling and data augmentation were common methods to tackle this issue. However, more effective methods could be explored for future research, such as loss design [154] or the global-local model framework [40].

Another existing 'elephant in the room' problem for the bone CT analysis is that the habits, such as eating and exercising, or medical condition of the patient such as age, cancer, or other diseases, also have an impact on the bone diagnosis. These parts do not reflect on the bone CT but are essential for clinical diagnosis. The recently electronic health records [155, 156] presented a potential for bone CT diagnosis. A future direction could be to explore the multi-modality fusion methods using both bone CT data and patient medical records for the medical analysis.

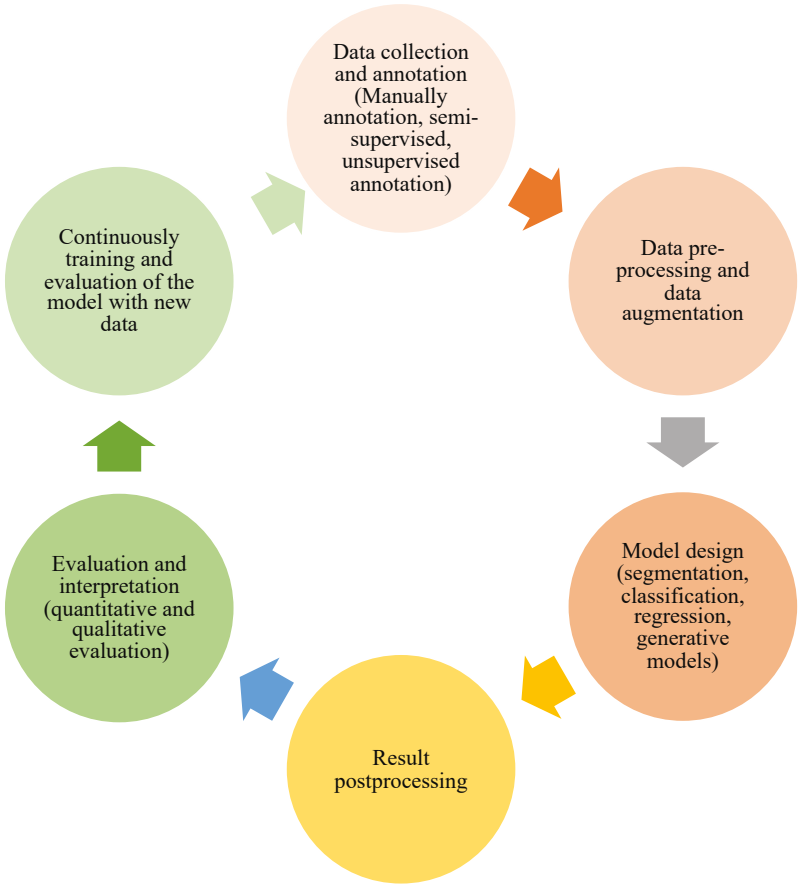

Fig. 2.2 Several parts to be considered for deep-learning based bone CT analysis.

Besides, understanding the decision-making mechanism of the deep learning-based bone CT analysis system is another crucial issue in clinical practice. Two ways can be employed for the model explanation, uncertainty estimation map [157, 158] and model decision heatmap [159]. The uncertainty measures the confidence level of the model's prediction, and the heatmap indicates the most useful part of the bone CT for the model to generate the output. Exploring more advanced uncertainty estimation and heatmap generation methods could be an interesting future work to help the doctors understand the model results and work as a diagnostic assistant for a mature deep learning system.

A successful deep learning solution of bone analysis task should consider several parts that illustrated in Fig. 2.2. In order to train and evaluate the deep learning approach, data collection and annotation is an essential part. Data preprocessing could aim to transfer the raw data into a usable format and simplify the complexity of the

task. The data augmentation improves the robustness and generalization ability of the designed model. Either choosing a validated model such as U-Net, FCN, ResNet, and DenseNet, or designing a new model could be a solution for the model selection. Considering the nature of the task should be an important part during the model designing, and splitting the task as several sub-tasks would improve the performance. The postprocessing part aims to eliminate noise and deliver better results. Both quantitative evaluation and qualitative evaluation could be considered to test the model's performance. The interpretation of the model like heatmap is also suggested to explain the decision strategy of the model. After the model is deployed, it should be continuously evaluated and updated with the newly collected data.

Chapter 3

Anatomical Segmentation of Human Foot CT

3.1 Introduction

The weight-bearing CBCT, which allows the patient to stand in the natural weight-bearing position during scanning, is a novel imaging technique for the medical treatment of foot and ankle. The high-resolution scanning from weight-bearing CBCT has tremendously aided the treatment and diagnosis of human foot, such as foot alignment and surgery [2, 29–31]. In these clinical procedures, the anatomical segmentation of foot bones, which offers an overall understanding of the patient's condition, is an important step in analyzing the CBCT foot scan.

The anatomical segmentation of the human foot is illustrated in Fig. 3.1. In total, there are thirty-one human foot bones, including tibia, fibula, talus, navicular, calcaneus, cuboid, three cuneiform bones, five metatarsal bones, fourteen phalange bones, two sesamoid bones, and accessory bone. Manual annotation of foot CT is a tedious and time-consuming process due to the complicated structure of the foot, and specialist knowledge is necessary during annotation. A fully automatic and accurate foot anatomical segmentation approach will greatly increase doctors' efficiency and is urgently needed.

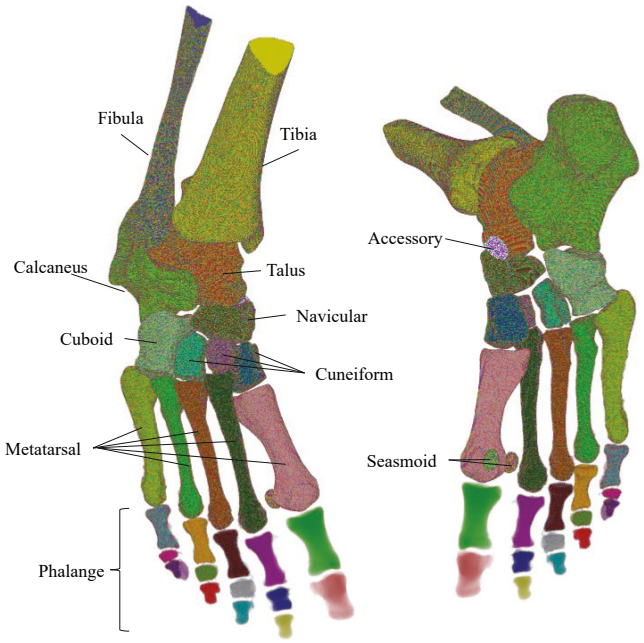

Fig. 3.1 Illustration of the anatomical segmentation of human foot bones.

Automatic anatomical segmentation of the 3D human foot CT scan requires assigning each pixel to the correct class based on the foot structure. Brehler et al. [160] developed a coupled active shape model to segment foot bones from CBCT scan. However, only the calcaneus, talus, navicular and cuboid bones have been segmented, and the segmentation results highly relied on the model initialization status. We considered dividing the human foot weight-bearing CBCT scans into thirty-one anatomical regions based on the foot structure to assist the clinical diagnosis [161, 162], which was more complicated than the previous works.

The automated anatomical segmentation of foot CT faces two challenges. The foot scan variation is the first challenge. Several examples of the average image through the axial-view of the foot CT scan are depicted in Fig. 3.2. There are two-feet scans, left-foot scans, and right-foot scans, with the feet in different positions and sizes in Fig. 3.2, raising the scan variation problem. The severe data imbalance between different bone classes is the second challenge. The respective average bone point numbers of the thirty-one foot bones over 38 feet scans are listed in Fig. 3.3. The tiny bones like

sesamoid and phalange are significantly smaller than the big bones of tibia and fibula. The two challenges severely affect the development of the foot CT segmentation model.

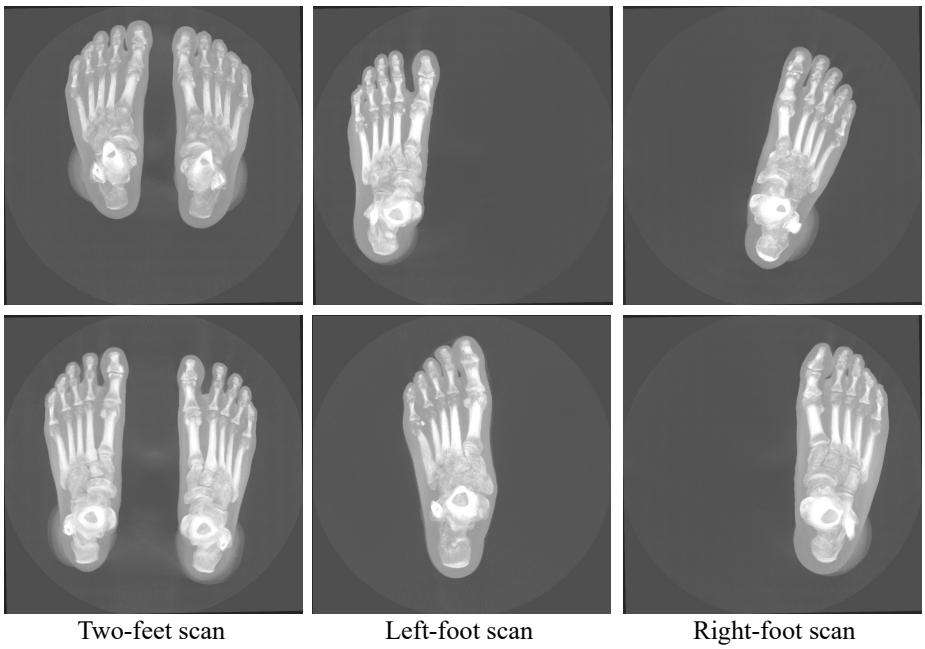

Fig. 3.2 Illustration of different foot scans.

We proposed the FootSeg method, which employed the deep neural network for the automatically anatomical segmentation of the human foot weight-bearing CBCT scan. The FootSeg is based on a three-stage framework, i.e., preprocessing, bone region segmentation, and bone pixel classification, to deal with the two challenges mentioned above. The model framework was depicted in Fig. 3.4. In the preprocessing stage, a foot standardization method was proposed to solve the scan variation problem. The bone region segmentation part aimed to identify the bone pixels while the classification model distributed the correct label to each bone pixel. The bone pixel classification model solved the data imbalance problem via a data sampling strategy. To the best of our knowledge, this was the first research of anatomical segmentation of all foot bones from weight-bearing CBCT using deep learning methods.

The innovation features of this work were:

- Framework for foot anatomical segmentation: The FootSeg method proposed a foot anatomical segmentation framework which consisted of three parts, foot

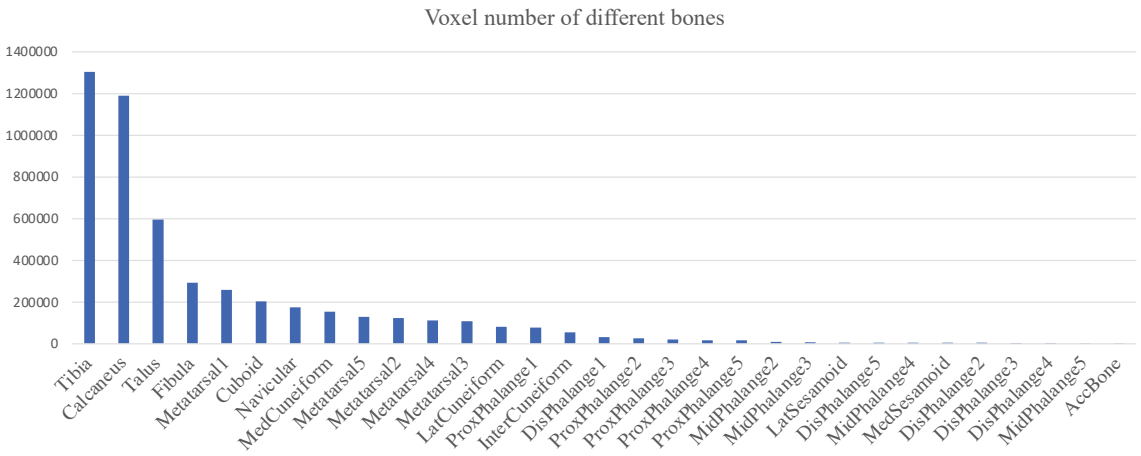

Fig. 3.3 Bone point number of different foot bones.

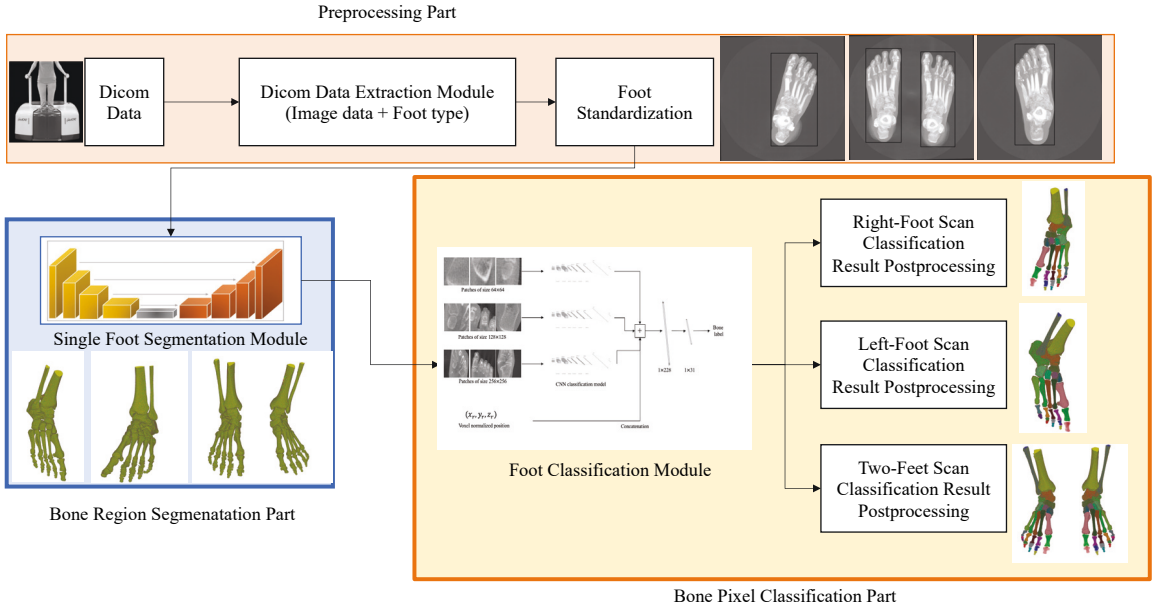

Fig. 3.4 Framework of the FootSeg anatomical segmentation method. The upper pink part was the preprocessing part, which used a foot standardization module to solve the foot scan variation problem; the lower-left blue part was the bone region segmentation module, which employed a Unet-based single foot segmentation model to extract the bone pixels; the lower-right yellow part was the bone pixel classification part, which first employed a CNN model to distribute the label of each bone pixel and then used a postprocessing module to generate the segmentation mask of each CT slice.

preprocessing, foot region segmentation, and foot bone classification. The three-step approach greatly reduced the complexity of the anatomical segmentation task.

- Innovative patch-training method: The serious data imbalance problem constituted a great challenge in the stage of foot bone classification. A data up-sampling and down-sampling training strategy and a patch-based CNN model have been proposed to solve this problem.
- Remarkable performance: The FootSeg method achieved remarkable qualitative and quantitative results in foot anatomical segmentation. The mean Intersection-over-Union (mIoU) including the background part was 80.3%, and the mIoU on the bone parts was 90.3% on the testing set containing eight feet.
- To the best of our knowledge, this was the first research of anatomical segmentation of all foot bones from weight-bearing CBCT using deep learning methods.

3.2 The Multi-Stage FootSeg Method for Foot Anatomical Segmentation

The FootSeg method for anatomical foot segmentation was divided into three stages, preprocessing, bone region segmentation, and bone pixel classification. Fig. 3.4 illustrated the FootSeg framework. The input of the FootSeg model was the Dicom data from the weight-bearing CBCT machine. In the preprocessing step, the Dicom data were firstly processed as tiff images, and a foot standardization step converted the different kinds of foot CT data (two-feet scan, right-foot scan, and left-foot scan) as the right-foot-like scan to simplify the model designing. The foot bone region segmentation step utilized a U-net model to extract the bone pixels from the CT data. The foot classification stage developed a patch-based CNN model to extract bone image feature, and utilized both bone image feature and bone pixel position feature to identify the label of the bone pixels. A postprocessing module was designed to generate the final results according to the foot type.

3.2.1 Foot Standardization

The FootSeg method aimed to handle different kinds of foot scans. However, the different foot scan types, e.g., left-foot scan, right-foot scan, or two-feet scan, introduced more variations to the anatomical segmentation of foot bones. A foot standardization method was proposed to deal with this problem.

The foot Dicom data contained the type of the foot scan, and the Dicom data extraction module generated the tiff image and extracted the foot type information. In the foot standardization step, for the two-feet scan, the Otsu segmentation method [163] was used to calculate the threshold between the foot data and the background, and generate the foot muscle and bone mask on each foot slice. The foot masks were then processed by morphological operations like the opening and closing to remove the noise in the foot masks. After that, only the masks with two regions were used to calculate the foot bounding box coordinates, and the coordinates over all the selected masks were merged to generate the two foot bounding boxes of the two-feet scan. The left-foot and right-foot were separated from the two-feet scan according to the two foot bounding boxes. For the single foot scans, the bounding box was calculated using the same method, while only one bounding box was generated. Then, in the two-feet scan or the left-foot scan, the left foot and left foot bounding box were flipped as a right foot and right foot bounding box. All foot scans were treated as right foot scans in the following steps. The foot standardization step dramatically reduced the complexity of the following model designing, and the bounding box position was used as a bone position normalization parameter in the following step.

3.2.2 Foot Bone Region Segmentation

Directly processing the foot data was difficult since the thirty-one foot bones were of different shapes and sizes. To simplify the problem, we proposed a foot bone region segmentation stage before the foot bone classification. The goal of this step was to extract the region of interest (ROI) from the CT scan. The foot scan consisted of background, muscle, and foot bones. The first two introduced more difficulty to directly classify the foot bones from the CT data. The foot bone region extraction stage aimed

to assign each pixel in the CT slice as a bone pixel or a non-bone pixel. A U-net model was used for the foot bone region segmentation. The input was the single foot slice data from the foot standardization step, and the segmentation map of the foot bone region was utilized as the input to the next foot bone classification stage.

3.2.3 Foot Bone Classification

The goal in this stage was to distribute each bone pixel with its corresponding anatomical bone label. The input features, required to capture enough information for classification, were crucial to the classification results. A patch-based CNN model was designed to extract the bone image features and an auxiliary bone position feature was generated using the foot bounding box. Both the bone image features and the auxiliary bone position feature were employed to classify the label of each bone pixel.

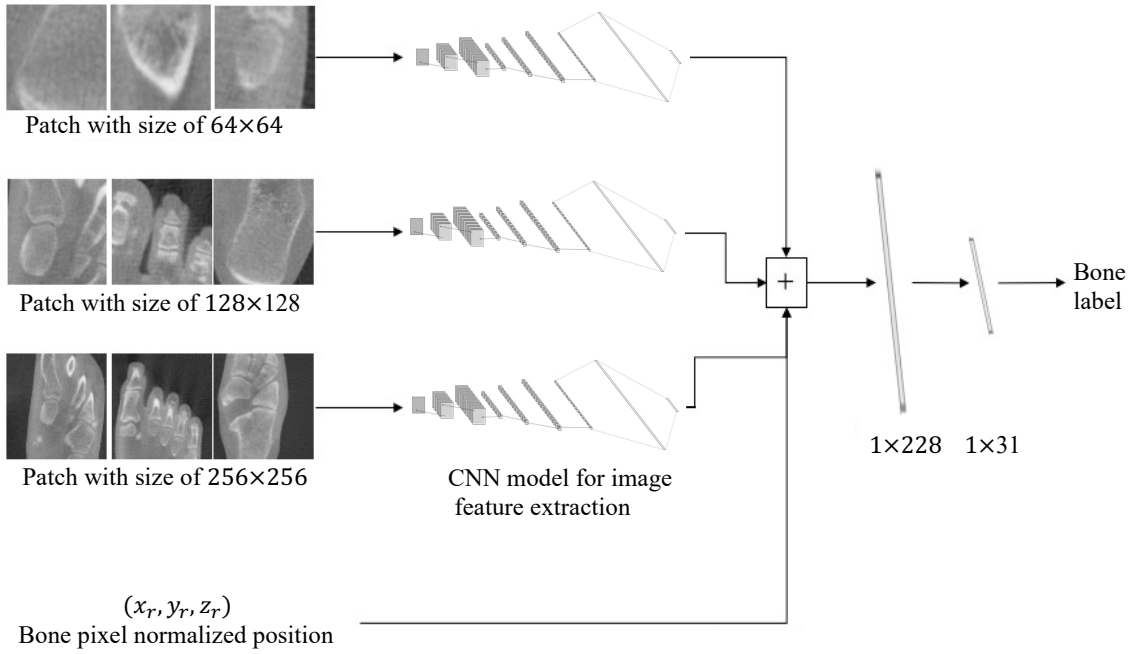

Fig. 3.5 The classification model framework. Three different sizes of patches (as illustrated in Fig. 3.6) were used to extract the bone image features via the CNN model (as illustrated in Fig. 3.7), and they were concatenated with the auxiliary feature for the bone pixel label classification.

Fig. 3.5 demonstrated the proposed classification model framework. Three different sizes of patches (as illustrated in Fig. 3.6) were used to extract the bone image features via the CNN model (as illustrated in Fig. 3.7), and they were concatenated with the auxiliary feature for the bone pixel label classification.

Input Features

As mentioned above, we employed two types of inputs, the bone patches and the bone pixel position, to extract the image features and auxiliary bone position feature for classification. For each bone pixel, three 2D patches of size 64×64, 128×128, and 256×256, which were centered on the bone pixel, were cropped to capture the local and global information of the bone pixel. The three patches were all resized as 64×64 for the computational trade-off. The smaller patch as 64×64 encoded the local information with a high level of bone details, and the medium patch as 128×128 focused on a broader local view around each pixel. The 256×256 patch contained global spatial information of each bone pixel. Fig. 3.6 was the illustration of the different sizes of patch images. The first row was the patches with the size of 64×64, the second row listed the 128×128 patches, and the third row demonstrated the patches with the size of 256×256. The three patches contained the necessary local and global information for the bone label classification.

Besides the image features, the position of each bone pixel on the foot CT scan was also considered to assist classification. Instead of utilizing the absolute coordinate of each bone pixel, a normalized coordinate of each pixel has been employed. For a pixel with position of (x_i, y_i) and slice number z_i, the total number of CT scan N, and the coordinates of upper-left corner of the foot bounding box (x_0, y_0), the coordinates of the lower-right corner of the foot bounding box (x_1, y_1) were used to define the normalized position (x_r, y_r, z_r) of each pixel.

$$\begin{cases} x_r = (x_i - x_0)/(x_1 - x_0) \\ y_r = (y_i - y_0)/(y_1 - y_0) \\ z_r = z_i/N \end{cases} \tag{3.1}$$

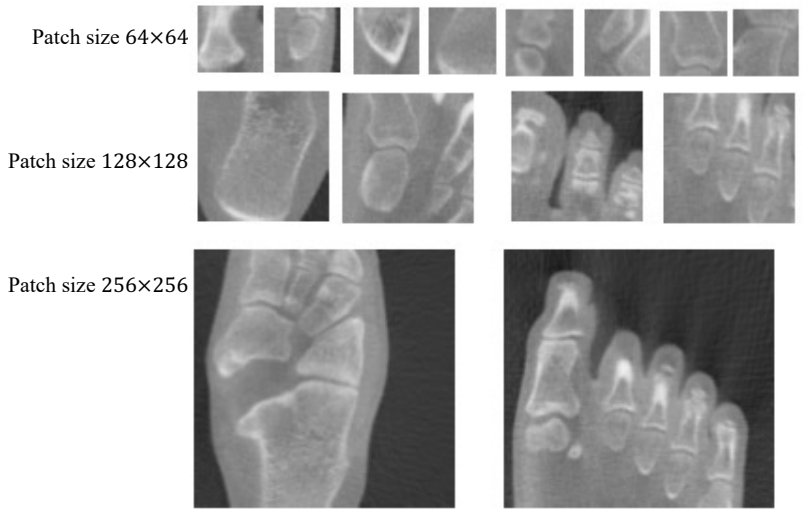

Fig. 3.6 Examples of different size of the bone patch images.

Unlike the absolute coordinates, the normalized coordinates represented the relative position of the bone pixels in the foot bone structure.

Model Design and Training

The proposed classification model architecture was depicted in Fig. 3.5. The classification model had four inputs, including three different size image patches and one auxiliary position information. For each bone pixel, the three types of patches around the bone pixel were fed into a CNN model to extract image features. The architecture of the proposed CNN model was shown in Fig. 3.7. The CNN model consisted of two cascade convolutional blocks followed by two fully connected layers. Each block contained two convolutional layers and one max-pooling layer. The convolutional kernel size of each layer was 3×3. The first max-pooling layer used 4×4 filters with a stride of four, while the second max-pooling layer applied 2×2 filters with a stride of two. The two fully connected layers contained nodes of 500 and 75, respectively. The dimension of the image feature of each bone patch was 75, as well.

The CNN features of the three image patches and the auxiliary position information were concatenated together as a 228-dimension feature and passed through a fully

connected layer with thirty-one nodes followed by a soft-max classifier. All convolutional layers were equipped with rectification (ReLU) non-linearity.

During training, the severe data imbalance problem where larger bones were more than a hundred times bigger than small bones would lead to an unsatisfied model. Since some of the bone patches were similar inside the same bone class, putting all the bone points into training was not necessary. A data resampling method was proposed to solve the data imbalance problem. We randomly resampled N patches (N=20000 in training) from each foot data in the training set for each bone. For the bones whose total bone points were less than N, up-sampling was performed, while for the others, down-sampling was operated. The cross-entropy loss and the ADAM optimization method were used to train the model.

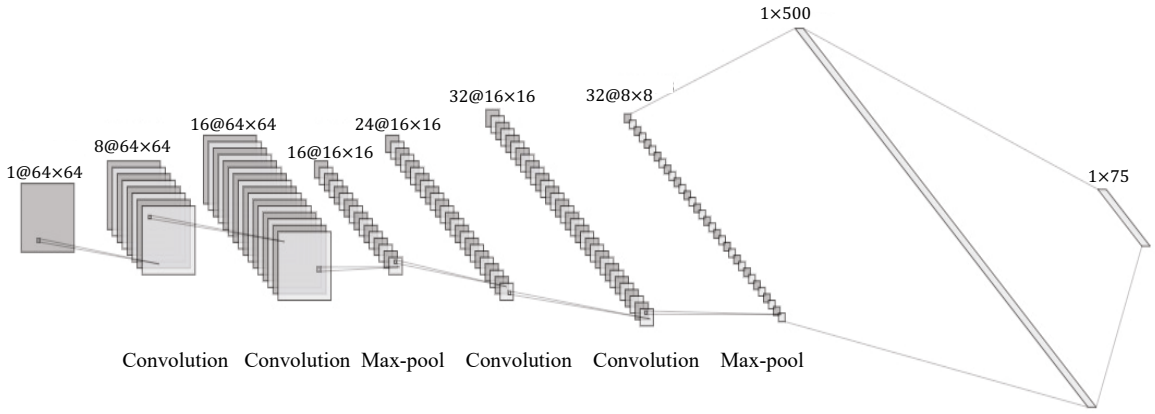

Fig. 3.7 The CNN model for the feature extraction of patch image.

3.3 Experimental Results

3.3.1 Data Collection, Annotation and Database Construction

We collected foot CT data using CurveBeam weight-bearing CBCT machine. Eleven scans of the left foot, eleven scans of the right foot, and eight scans of two feet were collected. All scans were made anonymized by removing all the patient-sensitive information. We manually annotated the thirty-one bones of the foot CT scan data, and

each scan required nearly five $\sim$ eight hours to finish the annotation. The annotation quality were checked by a professional.

In total, we have collected thirty-eight single foot scan data with corresponding bone annotation to develop and evaluate the proposed method. The foot scans were in different shapes, positions and sizes. Each scan contained 533 slices with a resolution of 950×950, and the voxel size was $0.37mm$. Thirty foot scan data (nine left foot scans, nine right foot scans, and six two-feet scans) have been used for training, and the remaining eight foot scan data, including two left-foot scans, two right-foot scans, and two two-feet scans, have been utilized to evaluate the model performance in both the bone region segmentation stage and the bone classification stage.

During the model training of the classification model, for each bone in each scan, we extracted 20000 image patches for the model training using down-sampling or up-sampling methods in section 3.2.3. In total, for each bone, although we only have used thirty scans in the training set, 600000 image patches have been extracted for the model training which were sufficient for the classification model. For the model evaluation, we have used two left-foot scans, two right-foot scans, and two two-feet scans for model testing. These scans covered different foot sizes, shapes and positions in the real situation, which were sufficient to evaluate the model performance.

3.3.2 Implementation Details and Evaluation Matrix

The U-net model for the bone region segmentation and the foot bone classification model were trained on NVIDIA GeForce GTX 1080Ti GPU with a learning rate of 0.001 and 0.0001, respectively. The bone region segmentation model was trained for twenty epochs using Dice loss, and the foot bone classification was trained for ten epochs with cross-entropy loss. The models were implemented using PyTorch.

The bone region segmentation and bone classification performances were assessed by the mIoU index. The IoU was defined as follows:

$$IoU = \frac{A \cap B}{A \cup B} \tag{3.2}$$

A and B were the predicted segmentation and the ground truth, respectively. The IoU ranged from $0 \sim 1$, and a higher IoU value indicated a better segmentation result.

3.3.3 Bone Region Segmentation Results

The bone region segmentation model was developed using U-net, and the results using scans after foot standardization and without foot standardization have been compared. The IoU result of the U-net trained from the foot scans without foot standardization was 96.2%, while the U-net trained from the foot data after foot standardization achieved an IoU of 96.7%. The bone region segmentation model provided an accurate bone region mask for bone classification. The foot standardization step gained an improvement of 0.5% of the IoU result, which indicated that the foot standardization step not only simplified the model designing, but also improved the segmentation performance.

3.3.4 Bone Classification Results

The bone classification model was designed based on the patch images and the pixel position information. We compared the performance of the proposed model with the model using different patches. In clinical treatment, the results of the bones were more important than the background. We conducted two evaluations on the mIoU index, mIoU including background and bones, and mIoU value only considering bones. Table 3.1 listed the results of different methods.

The U-net and its variants, such as patch-based U-net, 3D U-net, and hybrid U-net, have also been tried for foot anatomical segmentation. The dice loss, cross-entropy loss, and focal loss have been used for the U-net based model training. However, because of the severe data imbalance of different bones and the large number of the total classes, the U-net based methods could not be well-trained and all pixels were classified as background. Therefore, we didn't list the results of the U-net based methods in Table 3.1 for simplicity.

3PatchPNModel indicated the proposed model. 3PatchWithoutPNModel indicated the model using the three patches of size 64, 128, and 256 as input without the pixel po-

Table 3.1 The performance comparison of different models.

Model	mIoU (only, bones %)	mIoU (bones and background, %)
3PatchPNModel	90.34	80.30
3PatchWithoutPNModel	90.21	80.21
Patch64Model	85.27	76.31
Patch128Model	89.38	79.57
Patch256Model	88.16	78.65

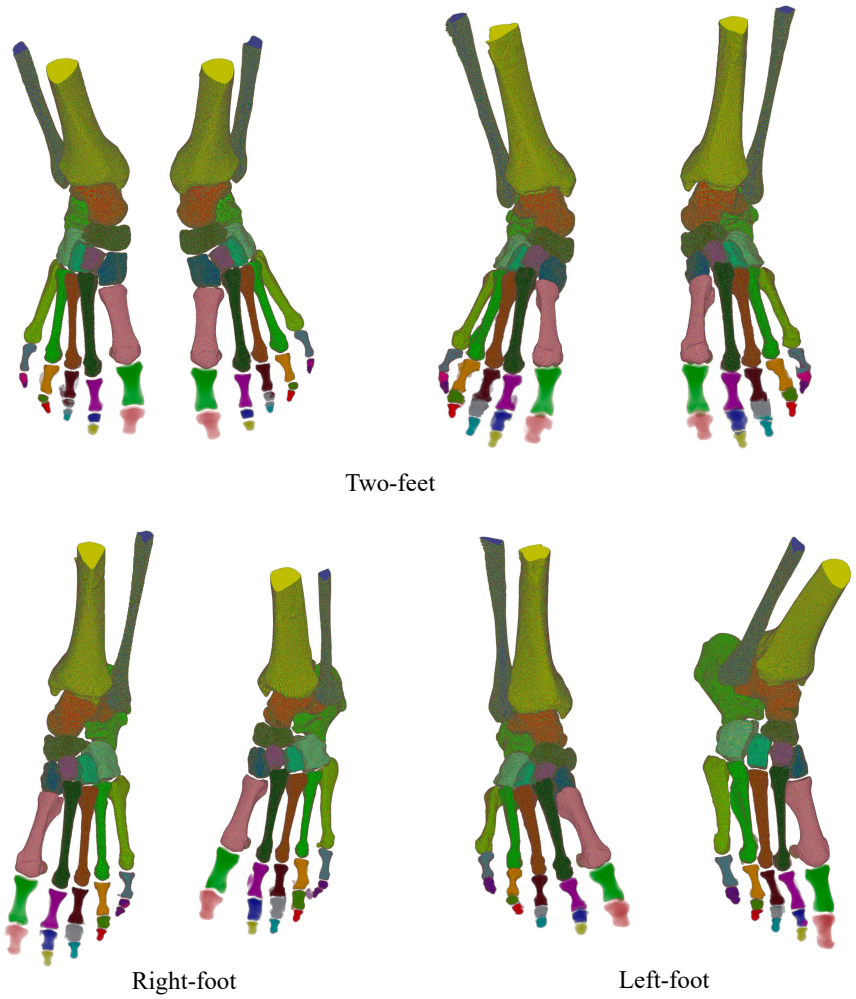

Fig. 3.8 The anatomical segmentation results on the test dataset.

sition normalization feature. The Patch64Model, Patch128Model, and Patch256Model indicated the model using the images with the patch size of 64, 128, and 256, respectively. The proposed model achieved the highest performance compared with the other models. The mIoU of the proposed model was 90.34% when only bone parts have been considered, and the mIoU was 80.3% when both bones and background were included for evaluation. The decrease between the two mIoU numbers was due to the errors in the bone region segmentation step. The qualitative results of the test dataset were depicted in Fig. 3.8 and showed visually satisfied segmentation results on both the single foot scans and the two-feet scans. Both the quantitative and qualitative results on the test dataset demonstrated the effectiveness of the proposed method.

3.4 Conclusion

We developed FootSeg, a deep learning-based method for automatically foot anatomical segmentation from weight-bearing CBCT scans. The proposed method was implemented based on a three-stage procedure, preprocessing, foot bone region segmentation, and foot bone classification. The three steps simplified the model design, and the experimental results demonstrated that the proposed method produced accurate and visually satisfying results for foot anatomical segmentation. To the best of our knowledge, this was the first work in automatically foot anatomical structure segmentation from weight-bearing CBCT scan.

Chapter 4

Instance Segmentation of Human Wrist CT

4.1 Introduction

In recent years, wrist Computed Tomography (CT) has performed an essential role in clinical practice. Due to the advantage of high image quality and low radiation of CT, the wrist CT has shown high potential in various applications such as osteoporosis classification [6], rheumatoid arthritis diagnoses [32], and bone fracture assessment [33]. A vital procedure among the above applications is the wrist instance segmentation, i.e., distributing the right class label to each voxel in the CT data. An illustration of wrist instance segmentation is shown in Fig. 4.1.

There were some existing works of wrist segmentation in CT images [34–37, 164, 165]. These methods leveraged the intensity difference between the bone boundary and the other parts and utilized the shallow features (e.g., edge and line, intensity, region growing, and active contours) for the wrist bone segmentation. Sebastian et al. [164] proposed a skeletally coupled deformable model (SCDM) for wrist bone segmentation by combining the advantage of active contour, seeded region growing, and region competition. The manually initialized region seeds would grow under a curve evolution approach, and the growth was modulated by skeletally-mediated competition between neighboring regions in the SCDM method. Chen et al. [165] employed a rigid

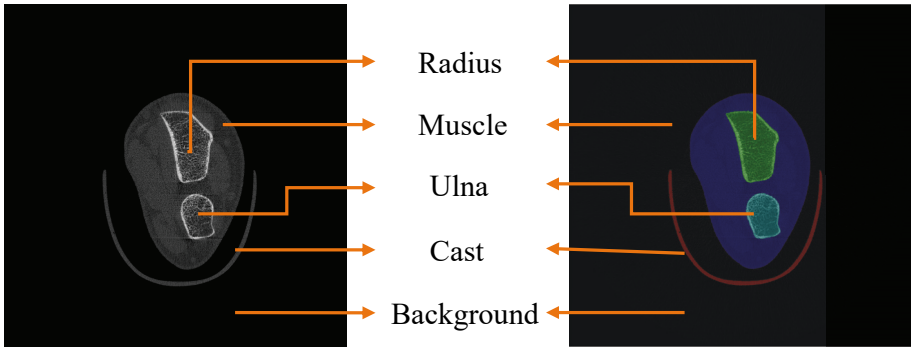

Fig. 4.1 Illustration of wrist instance segmentation (The wrist CT contains radius bone, ulna bone, muscle, cast, and background. Cast is the carbon fiber holder in the CT machine. The pixel values are similar between cast and muscle in wrist CT).

image registration method to boost the grow cut segmentation model for the sake of propagating the segmentation of bones to new poses or new individuals. However, both [164] and [165] required manually seed selection before segmentation. Anas et al. [166] designed an automatic wrist bone segmentation technique by using a group-wise registration framework based on a Gaussian Mixture Model. The segmentation results highly relied on the initial coarse segmentation masks from the OTSU's thresholding technique [163]. However, these methods were brittle to complicated occasions or needed human interaction to achieve satisfied segmentation results.

Some works explored the deep learning methods for wrist segmentation. Xue et al. [167] used the U-net model for the segmentation of capitate bone from CT images and showed the potential of the U-net model in wrist segmentation. However, only one bone has been identified in their model because of the lack of annotation data.

Moreover, only wrist bones have been segmented while the muscle part has been neglected in these methods. The muscle part contains health information such as muscle strength, muscle mass, and body mass index. They are essential in clinical analysis. Traditional bone segmentation methods can not satisfy the need for muscle segmentation. A comprehensive wrist instance segmentation method that can segment all components, including bones, muscle, cast, and background, in the wrist CT in one go is highly demanded.

Usually, a massive number of annotated images are essential to train a reasonable deep learning model. However, manually annotating numerous wrist CT slices is a

tedious and time-consuming task. Lacking annotation of wrist CT slices heavily limits the application of deep learning-based methods in wrist CT segmentation.

Some works explored the self-training semi-supervised methods to develop an image segmentation model to alleviate the annotation work load. The self-training semi-supervised segmentation methods [168, 169] trained an initial model from the labeled data and generated the pseudo segmentation maps on unlabeled data using the initial model. The initial model was retrained with the pseudo segmentation maps. Bai et al. [168] proposed a semi-supervised approach for cardiac MR image segmentation. Their approach alternately updated the segmentation model and the pseudo model. The segmentation network predicted the pseudo label of the unlabeled data by using the softmax probability prediction map. The labeled and unlabeled data with the pseudo label were used to update the segmentation network. Sedai et al. [169] adopted a student-teacher method to segment the retinal layers in OCT images. The teacher model generated the soft labels and uncertainty map for the unlabeled set using Monte Carlo (MC) dropout. The student was updated by the estimated soft label and the corresponding label confidence. Zhao et al. [170] adopted the self-training method for finger bone segmentation of hand X-ray. They used a U-net model with a CRF module to generate the pseudo label of unlabeled X-ray images. The U-net model was trained using the labeled data in the first step, and the pseudo labels of the unlabeled data were predicted by the U-net model and the CRF module. The model parameters were iteratively updated using the two steps.

Traditional shallow segmentation methods avoid the data annotation procedure with limited successful results. Deep learning methods achieve high performance but need laborious data annotation. Employing the limited successful results from the traditional segmentation methods or self-training semi-supervised segmentation methods could be a possible way to train the deep learning model. This chapter proposed a semi-automatic method to construct a wrist annotation database via the shallow OTSU-based [163] model and the U-net [108] based deep learning model. The different regions on wrist CT were annotated as radius, ulna, muscle, cast, and background as depicted in Fig. 4.1. The proposed method highly alleviated the manual annotation workload.

To further design an efficient wrist instance segmentation model, this chapter proposed an end-to-end edge reinforced U-net segmentation model. In most cases, edge information can provide strong hints for segmentation [171–173]. Therefore, introducing edge restriction would highly stabilize the procedure of model training and improve the segmentation results.

The contributions of this work were summarized as follows:

- We proposed a novel semi-automatic method to annotate 5k wrist CT slices. The proposed method highly alleviated the workload and reduced the annotation time compared with the laborious and time-consuming manual annotation.
- We designed an end-to-end edge enhanced U-net segmentation model for wrist CT instance segmentation. The edge-enhanced segmentation model achieved both qualitative and quantitative results.
- The proposed model could segment all components in the wrist CT while the existing methods segment bone parts only. To the best of our knowledge, this was the first work for the instance segmentation of wrist CT.

4.2 Overview of the Proposed Methods

We first described the overview of the proposed methods. Given a CT slices collection of the wrist, two goals should be achieved to develop a wrist instance segmentation model. The first goal was to construct a wrist annotation database. The second goal aimed to develop a deep neural network that used a single slice as input and generated the instance segmentation result, including the radius, ulna, muscle, cast, and background. A segmentation example was illustrated in Fig. 4.1. The overview of the wrist annotation database construction method and wrist segmentation model designing were described below.

4.2.1 Overview of the Semi-Automatic Construction Method of the Wrist Instance Segmentation Database

The overall framework of the semi-automatic construction method of the annotated database was depicted in Fig. 4.2. The semi-automatic construction method of the annotated database consisted of an OTSU-based radius, ulna, muscle&cast, and background annotation part and a U-net based cast annotation part.

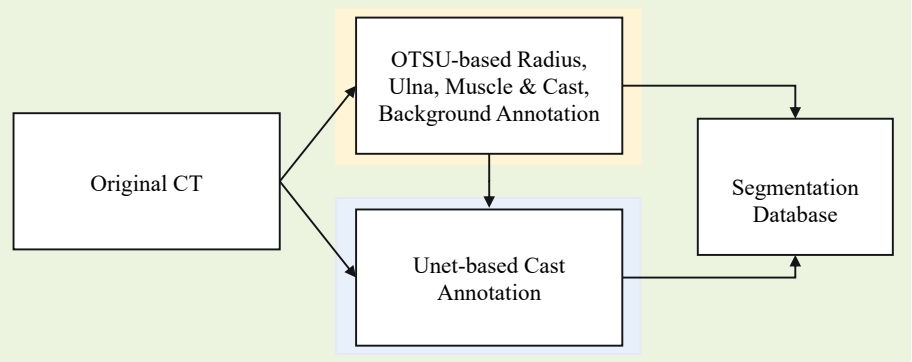

Fig. 4.2 The framework of semi-automatic construction of the wrist instance segmentation database.

The detailed framework of the proposed semi-automatic construction method was depicted in Fig. 4.3. The middle green part was the main procedure of the semi-automatic annotation method. The upper yellow part was the framework of the OTSU-based annotation, and the lower blue part was the framework of the U-net-based cast annotation.

From the unlabeled CT slice database, an OTSU-based segmentation method was designed to generate the segmentation mask of the radius bone, ulna bone, muscle-cast, and background. The cast was made of carbon fiber, and the pixel value was similar to the pixel value of muscle in wrist CT. The shallow features from the OTSU method have failed for the segmentation of muscle and cast because of their similar data distribution.

We employed the self-training idea from the semi-supervised segmentation method to extract the cast part from the CT slices. Two U-net models have been trained in

this step. The first U-net was trained using 100 cast-annotated slices, and the second U-net was trained based on the pseudo segmentation results of the first U-net model.

We selected the fine segmentation masks from the OTSU segmentation results and the U-net cast segmentation results. The OTSU segmentation results and the cast segmentation results were then merged as the wrist instance segmentation database. Two examples of the annotated slices were shown at the right of the middle part at Fig. 4.3.

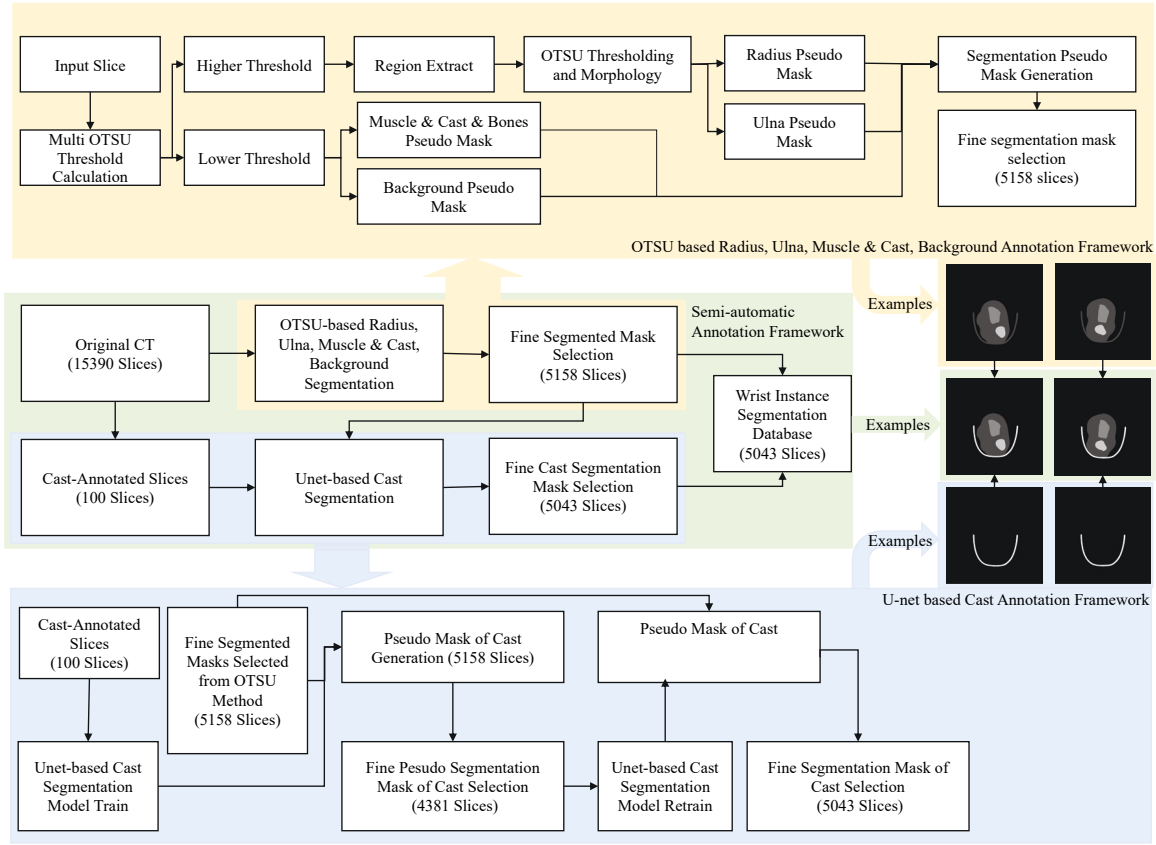

Fig. 4.3 The detailed framework of semi-automatic construction of the wrist instance segmentation database (The upper yellow part was the framework of the OTSU-based radius, ulna, muscle and cast, and background annotation, the middle green part was the framework of the semi-automatic annotation method, and the lower blue part was the framework of the U-net based cast annotation).

4.2.2 Overview of the Edge-Enhanced Wrist Instance Segmentation Model

Edge information provided essential information for segmentation. We proposed an edge loss module to reinforce the U-net model for wrist instance segmentation. The model framework was illustrated in Fig. 4.4. The U-net segmentation module used an encoder-decoder with skip connections to generate the feature maps of the wrist slice. The feature maps were fed into a 1×1 convolution layer to generate the segmentation prediction. Then, we used another 1×1 convolution layer to force the segmentation result to produce an accurate edge prediction of the original wrist slice. The whole model was jointly optimized by the segmentation loss L_{seg} and the edge prediction loss L_{edge}.

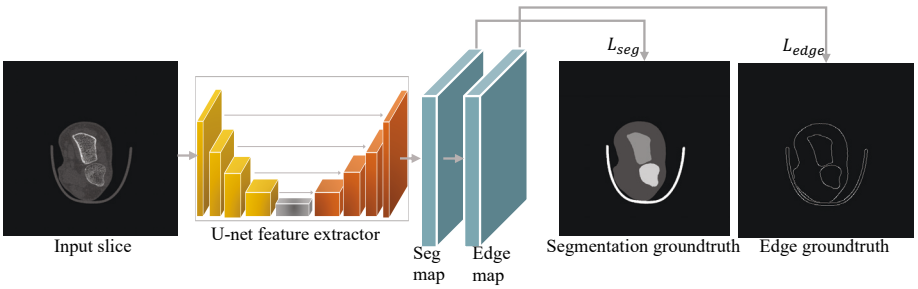

Fig. 4.4 Edge-enhanced wrist segmentation framework.

4.3 Semi-Automatic Construction Method of the Wrist Annotation Database

Given M unlabeled wrist CT slices, $\{x_{u_0}, x_{u_1}, \ldots, x_{u_M}\}$, where x_{u_i} was the i_{th} unlabeled slice data, the proposed wrist annotation method would successfully generate N annotated masks ($N \leq M$) from the M slices as $\{y_{l_0}, y_{l_1}, \ldots, y_{l_N}\}$ where y_{l_j} was the j_{th} annotated mask data. As the depicted framework in Fig. 4.3, the annotated data was generated by combining the result from the OTSU-based radius, ulna, muscle-cast, and background segmentation method and the result from the U-net based cast segmentation method. The OTSU-based segmentation method and U-net based cast segmentation method were described below.

4.3.1 OTSU-Based Radius, Ulna, Muscle-Cast, and Background Segmentation Method

The pixel values of muscle and cast, background, and bones were within different ranges. We leverage this feature for the segmentation of the wrist CT. An OTSU-based segmentation model was designed. The framework of the OTSU-based segmentation model was depicted in the upper yellow part of Fig. 4.3.

The basic theory of the single-OTSU method [163] and the multi-OTSU method [174] were summarized below. Given a slice x with resolution of $W \times H$, where W was the width, H was the height, the single-OTSU method used all pixel values $p_{mn}, 0 \leq m \leq W, 0 \leq n \leq H$ to calculate a threshold th. The threshold th minimized the intra-class intensity variance. The multi-OTSU method generated multiple thresholds and still satisfied the requirement of minimizing the intra-class intensity variance.

In the proposed semi-automatic annotation method, the input slice generated two thresholds th_{low} and th_{high} using the multi-OTSU method [174] firstly. The lower threshold th_{low} was used to divide the background and the muscle-cast-bones part. The higher threshold th_{high} was used to extract the bone region $[w_l - 20 : w_h + 20, h_l - 20 : h_h + 20]$ where w_l, w_h, h_l, h_h were the lowest and highest coordinates value along the width-axis and height-axis of the pixels whose values were larger than th_{high}, respectively. The region was expanded by 20 pixels along each direction to reduce the segmentation errors. The image two and image three of Fig. 4.5 were examples after the multi-OTSU threshold processing where image two was the segmentation illustration of background and the muscle-cast-bones using the lower threshold, image three was the bone region extraction illustration using the higher threshold.

The next step was to generate the radius and ulna bone mask from the bone region. A single-OTSU model was applied to the extracted region to generate the threshold th_{bone} to divide bone and muscle. The regions of radius bone and ulna bone were roughly segmented as where the pixel value was larger than th_{bone}. However, the pixel values on the cortical bone part were often larger than the th_{bone}, and the pixel values of parts of the trabecular bone part were smaller than the th_{bone}. This caused the bone

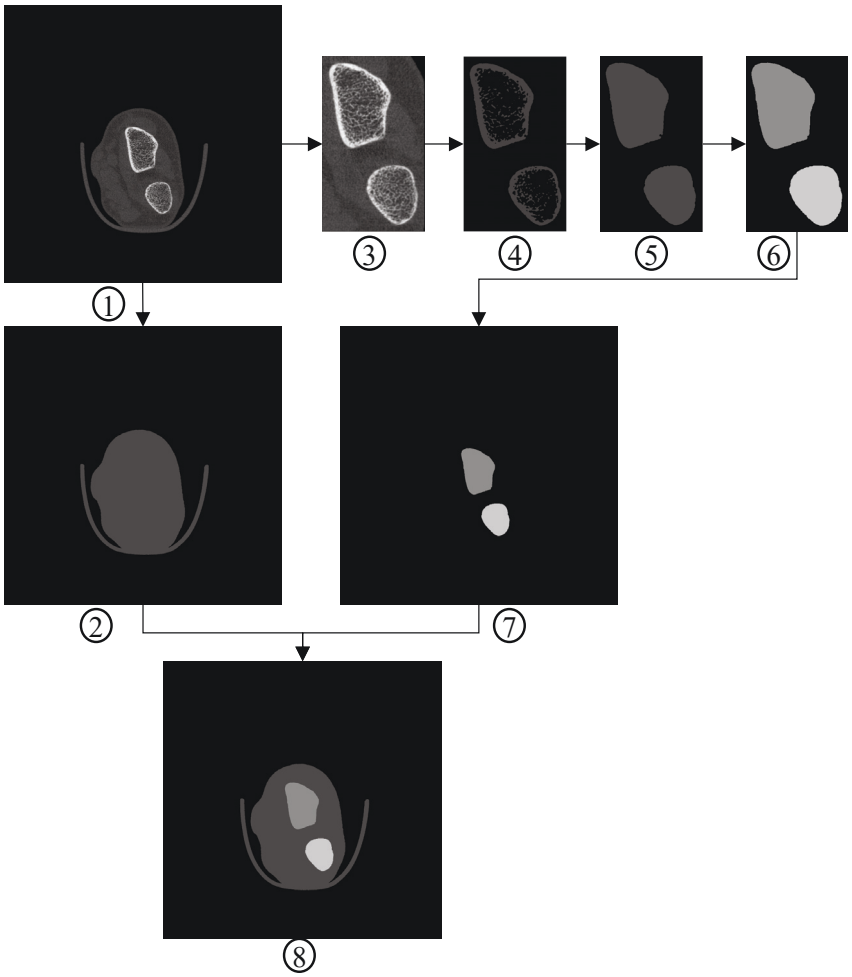

Fig. 4.5 Examples illustration of OTSU-based radius, ulna, muscle-cast, and background segmentation. Image one was the input slice, image two was the lower threshold segmentation map, image three was the bone region extraction patch using the higher threshold, image four was the segmentation map from OTSU, image five was the segmentation map after morphology processing, image six was the radius and ulna segmentation map on the bone region patch, image seven was the radius and ulna segmentation map in the original image, image eight was the merged segmentation result.

holes in the segmentation mask. The next step was to fill the holes in the bone region segmentation mask and identify the two bones.

A morphology-based method has been proposed to fill the holes. The bone regions were dilated for k times until two bone regions were connected or the k had reached the maximum dilation number K. $K = 4$ in experiment settings. After each dilation step, the bone holes were filled in the bone masks. Then, the two k_{th} dilated bone masks were eroded back for k times if the two k_{th} dilated bone masks were not connected. If the two k_{th} dilated bone masks were connected, then the $(k-1)_{th}$ dilated masks were eroded back for $(k-1)_{th}$ times. Image four and image five of Fig. 4.5 were the segmentation examples of the extracted bone region after the OTSU thresholding and morphology processing, respectively.

We leveraged the fact that radius bone was larger than ulna bone in the CT slices for the segmentation of the two bones. The larger bone mask was identified as the radius bone, and the smaller bone mask was identified as the ulna bone. Image six and image seven of Fig. 4.5 depicted the examples of the radius and ulna segmentation map.

The radius mask, ulna mask, muscle-cast mask, and background mask were merged as the generated segmentation mask. After the mask generation of the total dataset, a fine segmentation mask selection was performed to pick up well-annotated masks. Image eight of Fig. 4.5 was the example of the fine segmentation mask.

4.3.2 U-net-Based Cast Segmentation Method

The cast was a carbon fiber holder inside the CT machine, and the pixel value distribution was similar to the muscle. It was difficult to segment the cast and muscle using shallow features. We employed the self-training idea from the semi-supervised segmentation method to extract the cast part from the CT slices. The framework was depicted in the lower blue part of Fig. 4.3.

We firstly manually annotated the cast part of 100 wrist CT slices, and a U-net model $Unet_{initial}$ was trained to segment the cast. The pseudo cast label of the dataset was generated using the trained $Unet_{initial}$ model. Similar to the fine mask selection in

the OTSU-based segmentation stage, we selected fine cast segmentation slices for the parameter update of the U-net model. The U-net model was retrained from the selected slices. The retrained U-net model $Unet_{retrain}$ was used to regenerate the pseudo cast mask of the dataset. Then, we reselected the fine cast mask after the pseudo cast label generation. Both the U-net models were trained for 100 epochs.

The cast segmentation result and the OTSU-based segmentation results were merged to construct the final wrist instance segmentation database. Two examples were depicted at the right of the middle part of Fig. 4.3.

4.3.3 Results of the Semi-Automatic Construction Method of the Wrist Instance Annotation Database

We constructed a wrist CT database using the Scanco XTremeCT-I machine. 15390 slices have been collected, and the image resolution of each slice was 1536×1536.

OTSU-based Radius, Ulna, Muscle-Cast, and Background Segmentation Results

We used the OTSU-based methods to segment the radius, ulna, muscle-cast, and background from the 15390 slices. Since the OTSU-based methods highly relied on threshold selection, the segmentation results faced the problem of over-segmentation and under-segmentation sometimes. Fig. 4.6 showed examples of successful and unsuccessful segmentation results from the OTSU-based method.

After the processing of the OTSU-based method, we manually selected the fine segmentation masks. As shown in Table 4.1, 5158 slices have been selected from the 15390 slices, which was 33.52% of the whole candidate slices.

U-net-based Cast Segmentation Results

The 5158 slices selected from the OTSU-based method were used for cast segmentation. The U-net model was trained on NVIDIA GeForce GTX 1080Ti GPU using the ADAM optimization method and learning rate of 0.0001, respectively. Two U-net models have

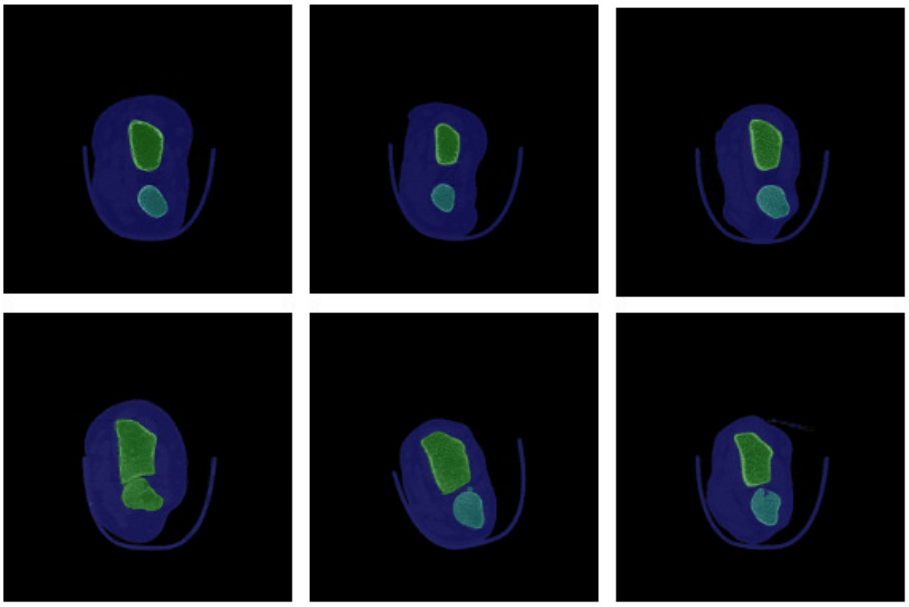

Fig. 4.6 Example of successful (first row) and unsuccessful (second row) segmentation results from the OTSU-based method.

been trained using the same experiment settings. The performance of segmentation was assessed by the IOU index.

The first $Unet_{initial}$ model was trained using 100 manually annotated slices. Ninety slices were used for training, and ten slices were used for validation. The input slice was resized to 480×480, and batch size was set as five. The model has been trained for 100 epochs, and the best IOU on the ten validation slices was 99.08%.

Then, the $Unet_{initial}$ model generated the pseudo cast masks of the 5158 slices picked from the OTSU methods, and 4381 slices have been selected for the retrain of the second U-net model $Unet_{retrain}$. The segmentation acceptance rate of the $Unet_{initial}$ model was 84.94%.

Based on the 4381 slices picked from the $Unet_{initial}$ model, the $Unet_{retrain}$ model used 3943 slices for training, 219 slices for validation, and 219 slices for testing. The model has been trained for 100 epochs. The experimental settings were the same as the first $Unet_{initial}$ model. The best IOU was 99.55% on the validation set, and the IOU on the test set was 99.54%.

Table 4.1 Segmentation results and acceptance rate of OTSU-based segmentation model and U-net-based cast segmentation model.

Model	Processed slices number	Selected slices number	Acceptance rate (%)
OTSU-based segmentation	15390	5158	33.52
$Unet_{initial}$ model for cast segmentation	5158	4381	84.94
$Unet_{retrain}$ model for cast segmentation	5158	5043	97.77

The pseudo cast masks of the 5158 selected slices from the OTSU-based method were generated by the $Unet_{retrain}$ model. 5043 slices were selected to construct the wrist annotation database. The acceptance rate of the $Unet_{retrain}$ model was 97.77%, an improvement of 12.83% from the $Unet_{initial}$ model.

Finally, 5043 slices with different shapes, positions and sizes of the wrist part were selected to construct the wrist segmentation database by merging the segmentation mask of the OTSU-based method and the cast mask of the $Unet_{retrain}$ model.

4.3.4 Data Annotation Time Analysis

We compared the database annotation time of the proposed method with the manual annotation method. In the proposed method, the time of manual annotation of the cast of 100 CT slices was two hours, around 1.2 minutes per slice on cast annotation. The time of the fine segmentation result selection from the OTSU-based method was around six hours. The time of the fine segmentation result selection from the $Unet_{initial}$ model and the time of the fine segmentation result selection from the $Unet_{retrain}$ model were both around two hours. The total time of manual work of annotating the 5043 wrist slices was around ten hours.

If we manually annotated the 5043 slices, the manual instance annotation time of one wrist slice was around four minutes. The total estimated manual annotation time of the same wrist database was roughly around 336 hours. The proposed data annotation method saved more time with much less manual work. What's more, since

the annotator only required to click yes or no in the selection of fine segmentation results, the workload was highly alleviated than manually annotation.

4.4 Edge-Enhanced Wrist Instance Segmentation Model

4.4.1 Method of the Edge-Enhanced Wrist Instance Segmentation Model

Given an wrist CT database containing N annotated slices ($\{x_j, y_j, e_j\}_{j=1}^{N}$), where $x_j \in \mathbb{R}^{H\times W}$ (H and W were the height and width) is the slice data, and $y_j \in \{0,1,2,3,4\}^{H\times W}$ was the ground truth annotated mask, ($\{0,1,2,3,4\}$ was the class label of radius, ulna, muscle, cast and background) and $y_j^e \in \{0,1\}^{H\times W}$ was the edge ground truth of the annotation mask ($\{0,1\}$ is the label of edge and non-edge), we proposed a U-net model reinforced by an edge prediction module for the wrist instance segmentation since the edge of the muscle, cast, radius bone, and ulna bone provided strong hints for the segmentation task and the edge prediction module could be used as a regularization item for the segmentation model training. The model framework was illustrated in Fig. 4.4. The model parameters were denoted as θ_{seg}.

The U-net feature extractor generated the feature maps of the input wrist slice. The feature maps were fed into a 1×1 convolution layer to generate the segmentation prediction possibility map p^s. A dice loss [175] was employed using the prediction possibility map p^s and the segmentation ground truth y.

$$L_{seg}(y, p^s, \theta_{seg}) = 1 - \frac{yp^s + 1}{y + p^s + 1} \tag{4.1}$$

Then, a 1×1 convolution layer was designed to force the segmentation possibility map p^s to produce accurate edge prediction map p^e of the ground truth segmentation mask y. An edge loss was designed using the edge prediction map p^e and the edge ground truth y^e. The edge prediction loss L_{edge} was a weighted binary cross entropy loss with a weight factor w on edge pixels. w was 120 in the experimental setting.

$$L_{edge}(y^e, p^e, \theta_{seg}) = -y^e log p^e - w(1-y^e) log(1-p^e) \tag{4.2}$$

The whole model was jointly optimized by the segmentation loss L_{seg} and the edge prediction loss L_{edge} as below.

$$L(y, y^e, p^s, p^e, \theta_{seg}) = L_{seg}(y, p^s, \theta_{seg}) + \lambda L_{edge}(y^e, p^e, \theta_{seg}) \tag{4.3}$$

λ was a balancing factor between the segmentation loss L_{seg} and the edge loss L_{edge} and λ was five in the experimental setting.

4.4.2 Wrist Instance Segmentation Results

The experiments were conducted on the constructed wrist annotation database. We used 4287 slices to train the proposed wrist segmentation model, 252 slices for validation, and 504 for testing model performance. The input slices were resized to 480×480, and batch size was set as five. The model has been trained for 100 epochs and optimized using the ADAM algorithm. The initial learning rate was set as 0.0005 and was decreased by half after every fifteen epochs. We compared the proposed model with the U-net [108] model with the dice loss. The U-net training settings were the same as the proposed method.

The performance of segmentation was assessed by the mean Intersection-Over-Union ($mIOU$) index. The $mIOU$ ranges from $0 \sim 1$ and a higher $mIOU$ value indicated a better segmentation result.

Table 4.2 Result comparison of IOU on wrist segmentation of U-net and the proposed model.

IOU (%)	Whole Part	Radius	Ulna	Muscle	Cast
U-net[108]	98.13%	98.91%	97.92%	99.58%	98.88%
Proposed model	98.68%	99.17%	98.24%	99.77%	99.48%

Table 4.2 reported the wrist segmentation performance on the test set of our model and the U-net model. The baseline U-net model achieved a high $mIOU$ of 98.13%

based on the large annotation database. However, the proposed model achieved a higher $mIOU$ of 98.68% and got an improvement of 0.55% compared to the baseline model. We also reported the IOU on each wrist component, including radius, ulna, muscle, and cast. The performance of the proposed method was 99.17% (radius), 98.24% (ulna), 99.77% (muscle) and 99.48% (cast), respectively. Compared with the performance of the U-net model (98.91% (radius), 97.92% (ulna), 99.58% (muscle), and 98.88% (cast)), the proposed model also achieved improvement on the segmentation of each component of wrist CT.

Since the baseline U-net model has achieved an IOU of more than 98%, the improvement of 0.55% was very remarkable of our proposed model (98.68%). Besides, since the cast and muscle were connected in some slices, the proposed edge loss could be helpful for these cases by preserving the edge shape of the muscle and the cast part, and achieved an IOU of 99.48% on the cast part and 99.77% on the muscle part, which were nearly 100% correct and better than the U-net baseline.

Moreover, the proposed model was more stable than the U-net model during training. The segmentation performance comparison on the validation set at each epoch of the proposed model and the U-net model was depicted in Fig 4.7. The results revealed that the proposed model was not vulnerable to the over-fitting problem compared with the U-net model since the edge loss performed as a regularization item.

Fig. 4.8 demonstrated the comparison of the qualitative results between the proposed model and the U-net model. We enlarged the comparison region for better visualization. Compared with the proposed model, the U-net model suffered the over-segmentation problem in case one and case 3, and suffered under-segmentation in case 2. In case 4, the segmentation results of radius and ulna bone from the U-net model were connected while the results of the proposed model were not. The comparison of the qualitative results also proved that the introduction of the edge loss improved the segmentation results.

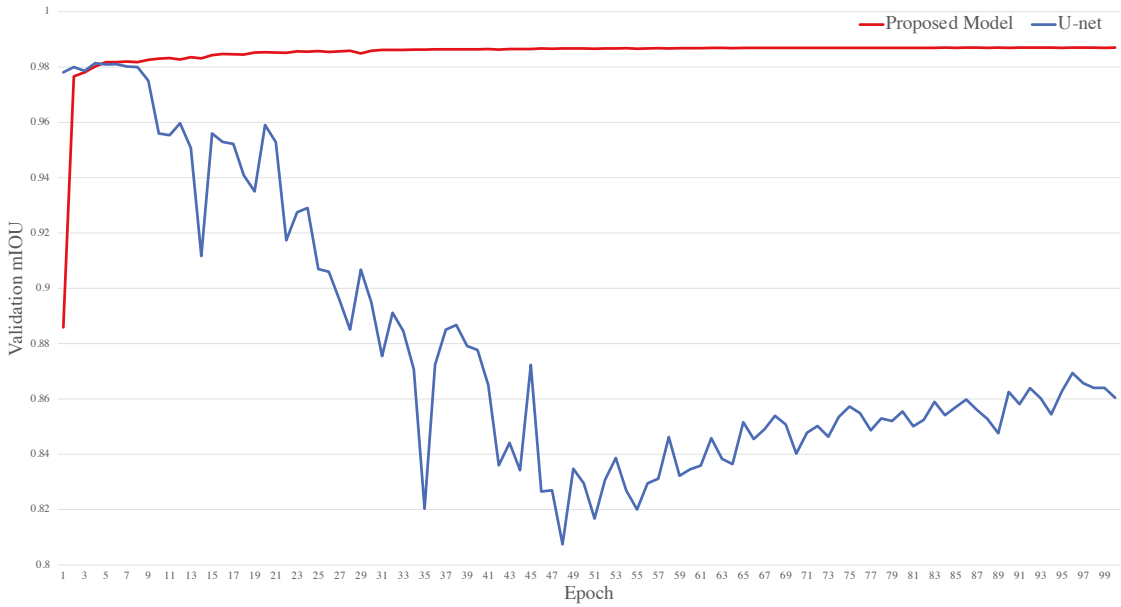

Fig. 4.7 Comparison of mIOU on validation set of U-net and the proposed model at each epoch during training. The proposed model (red line) was not vulnerable to the over-fitting compared with the U-net model (blue line).

4.5 Conclusion

We developed a semi-automatic method to annotate 5k wrist CT slices by employing the OTSU-based method and the U-net-based method. Our method only required fewer manual annotations, saved much time, and alleviated the annotation workload greatly. We also proposed an edge-enhanced segmentation model for the instance segmentation of wrist CT slice. The proposed model achieved better performance compared with the U-net model. The training procedure was more stable and was not vulnerable to over-fitting.

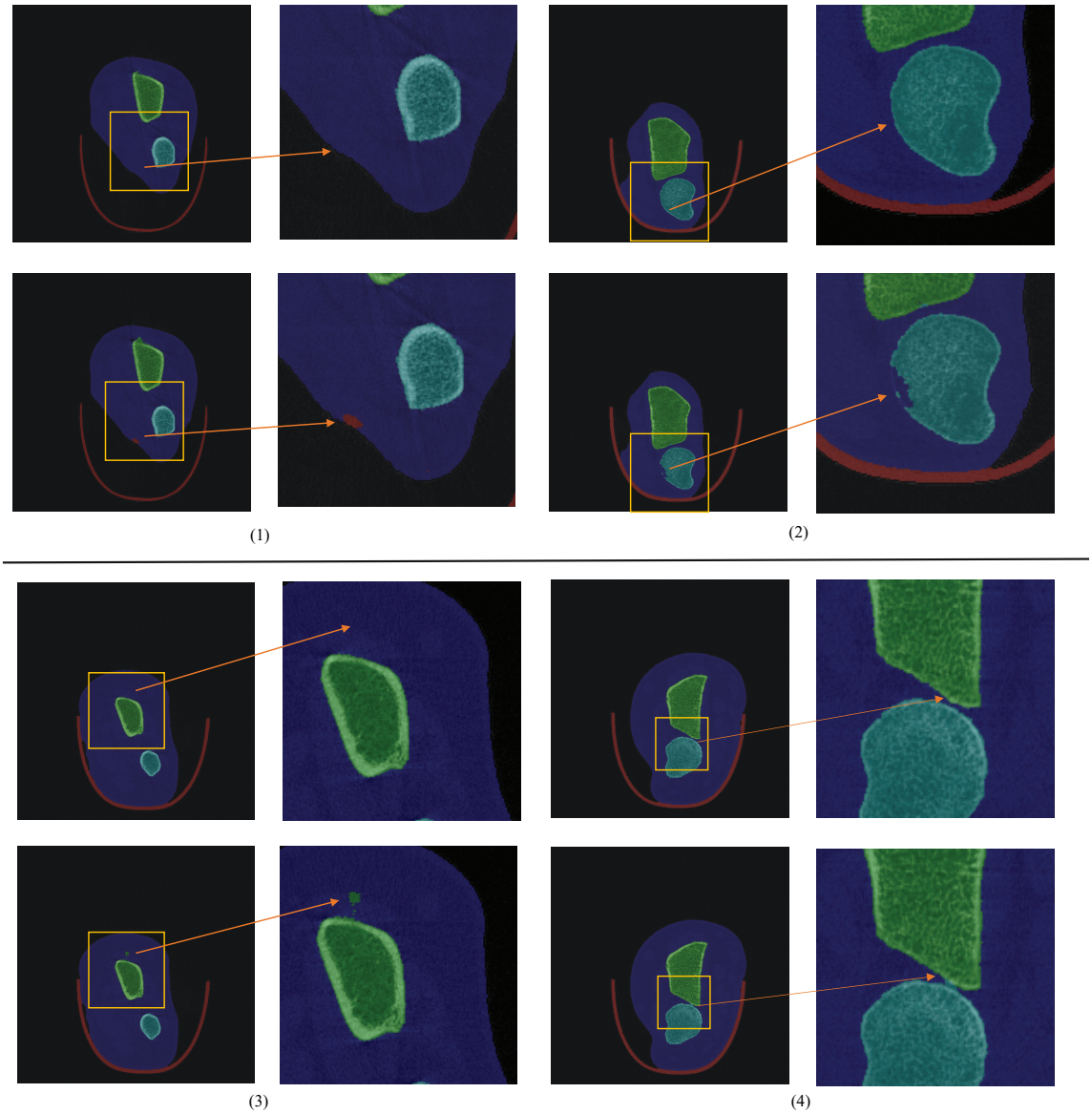

Fig. 4.8 Qualitative results comparison between U-net model and the proposed model (Row one and row three were the results of the proposed model, row two and row four were the results of the U-net model. The yellow bounding boxes denoted the comparison area of the proposed model and the U-net model. The comparison areas were enlarged in column two and column four, respectively).

Chapter 5

Semi-Supervised Segmentation of Bone CT

5.1 Introduction

We first described the motivation for this study of semi-supervised segmentation of bone CT, and then introduced the particular Hounsfield unit (HU) scale for bone CT segmentation in this section.

5.1.1 Motivation for the Semi-Supervised Segmentation of Bone CT

The bone segmentation is the fundamental procedure to assist the doctor's diagnosis for various bone CT medical applications, like osteoporosis analysis[6], orthopedic surgery [176], and bone fracture detection [177]. A recent trend is to utilize deep neural networks for bone segmentation [40, 66, 67, 97, 127, 178, 179]. Liu et al. [40] developed a two-stage pelvic bone segmentation model via a pelvic dataset of 1184 CT scans (over 320,000 CT slices), and the proposed method achieved a mean Dice coefficient of 98.7%. Noguchi et al. [67] also trained a U-net model for bone segmentation of whole-body scan from 16218 slices and achieved a mean Dice coefficient of 98.3%. Relying on

massive well-annotated data, deep learning models have achieved the state of the art performance in various bone segmentation tasks.

We have designed a deep learning based anatomical foot segmentation model in chapter 3 and a deep-learning based wrist segmentation model in chapter 4. Both of the two models achieved satisfied performance based on large training database. However, the pixel-level manual labeling, which was laborious and expensive, was performed in the two tasks. We have manually annotated 38 feet scans in the foot anatomical segmentation task. The annotation of one foot scan usually cost five $\sim$ eight hours which was a very time-consuming procedure. We have improved the annotation procedure in the wrist segmentation by utilizing the wrist structure and semi-supervised learning. Only the cast holder part of 100 wrist slices have been annotated and we have constructed a wrist annotation database with more than 5000 annotated slices within a short time. The semi-supervised learning strategy has significantly alleviated the workload for wrist segmentation model design.

Consistency regularization and self-training were the two main stream for the semi-supervised segmentation of various kinds of medical images, such as bone [180, 170, 181–183], cardiac segmentation [168, 184], liver segmentation [185] and vessel segmentation [169].

Consistency regularization methods [184, 186–188] were based on the smoothness assumption that data samples with the same label were close in the feature space. Li et al. [187, 188] introduced a transformation-consistent strategy for the semi-supervised medical image segmentation. Their method forced the model to generate consistent features for different transformations (rotation, noise, and scaling) of the input data. A regularization loss to encourage the consistency of pixel-level features was used for both labeled and unlabeled data. Yu et al. [184] proposed an uncertainty-aware mean-teacher framework for semi-supervised segmentation of left atrium MRI. The teacher model generated the estimation of the uncertainty value of each target prediction. The reliable ones were preserved, and unreliable predictions were filtered out. The student model was guided to produce consistent predictions according to the estimated uncertainty of the teacher model.

The self-training semi-supervised segmentation methods [168, 169] trained an initial model from the labeled data and generated the pseudo segmentation maps on unlabeled data using the initial model. We have used the self-training semi-supervised segmentation method in Chapter 4 to segment the cast part from the wrist slices and the annotation workload has been greatly reduced.

In spired by performance of the semi-supervised learning methods in our wrist segmentation database construction and other medical segmentation methods, we assume the semi-supervised learning methods, which leverage the particular attributes of bone CT, could be a potential way to reduce the annotation workload for bone CT segmentation model designing.

5.1.2 The Particular CT Attribute for Semi-Supervised Segmentation of Bone CT

We have used the pixel value to coarsely segment the radius and ulna in the chapter 4. In fact, the particular pixel value range of radius and ulna is due to the HU value of CT. The HU scale is used to express CT numbers in a standardized form. HU values of different objects are calculated from a linear transformation compared with attenuation coefficients of air and pure water. Different human body tissues and materials have different HU scale ranges. Table 5.1 lists the HU values of several typical body tissues and materials. The denser the tissue, the higher HU values in the CT scan. Higher HU values are displayed as brighter in the CT scan.

Though the bones of the human body exhibit a variety of sizes and shapes, their structure is the same. Bone is a rigid tissue that consists of the cortical bone and trabecular bone, as depicted in Fig. 5.1. The cortical bone is the hard outer layer of bones. The trabecular bone is the internal tissue of bones and is an open cell porous network formed by trabecular bone tissues and bone marrow. In the CT scan, the cortical part and trabecular part of the bone are displayed much brighter than the surrounding tissues (fat and muscle) due to the high attenuation of the high dense material in the bone tissues.

Table 5.1 HU values of different body tissues and materials.

Body tissues and other materials	HU value
Bone - Cancellous	300∼400
Bone - Cortical	500∼1900
Air	-1000
Water	0
Muscle	35∼55
Fat	-120∼-90
Lung	-700∼-600
Kidney	20∼45
Thymus	20∼120

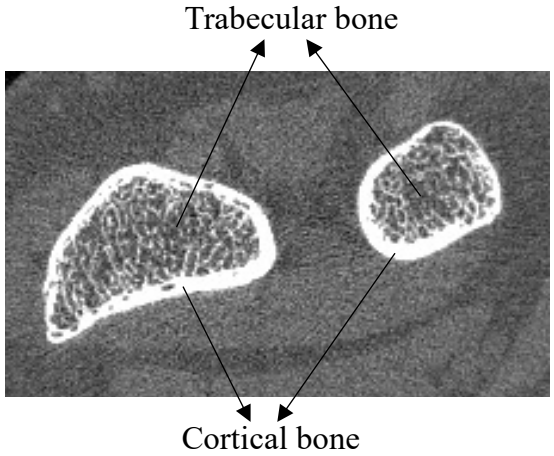

Fig. 5.1 Illustration of the cortical bone and trabecular bone. Cortical bone is the dense outer surface of bone, while trabecular bone is inside bone.

The bright cortex and trabecular bone tissues in the CT scan provide a strong local feature that distinguishes the bone from the other tissues. As illustrated in Fig. 5.2, we randomly extracted several patches from a CT scan of a wrist. Each patch contained only part of the bones. However, the bone structures could be easily identified even from the patches of CT scan due to distinguishing local features of bone, regardless of the bone size and shape. We further assumed that the local features of bone tissues were crucial to bone segmentation in CT scans, and proposed a semi-supervised learning method to leverage the local features.

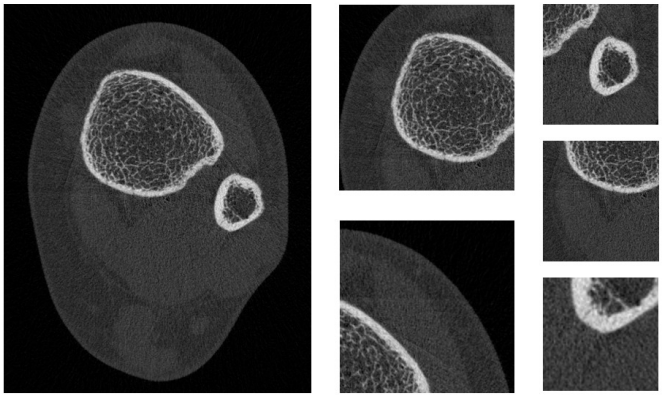

Fig. 5.2 Example of a wrist CT slice and the random extracted patches. The left image is the wrist slice, and the right five patches are the random extracted patches. It is easy to identify the bone region in the random extracted patches.

The model framework was depicted in Fig. 5.3. The CT image was divided into small patches, and the patches constituted a new patch-shuffled slice with a random patch order. Fig. 5.4 depicted examples of the original and patch-shuffled slices. Due to the strong local features of bone structures, the segmentation model should have the ability to distribute correct labels to both the original slice and patch-shuffled slice. We employed supervised losses to encourage the model to produce correct segmentation maps from both the original slice and the patch-shuffled slice on the labeled data. To utilize the unlabeled data, an unsupervised consistent loss was designed to force the output feature of the corresponding pixels in the patch-shuffled slice and the original slice to be the same. The proposed model was optimized jointly by the supervised losses and the unsupervised loss.

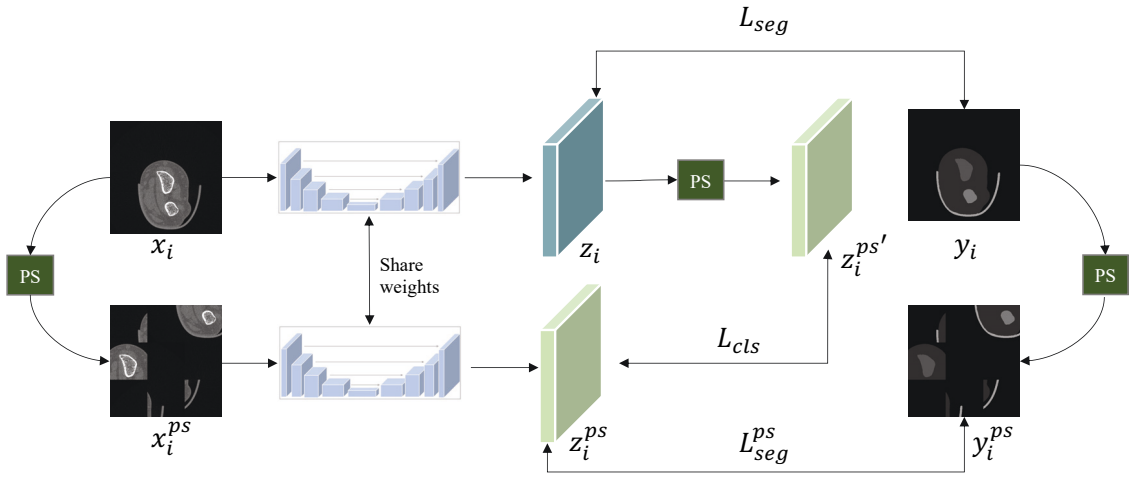

Fig. 5.3 The framework of the proposed semi-supervised learning method. x_i and x_i^{ps} were the original slice and patch-shuffled slice from the patch-shuffle transformation PS, respectively. z_i and z_i^{ps} were the model outputs of the x_i and x_i^{ps}, respectively. z_i^{ps} was the patch-shuffled feature map of z_i. y_i and y_i^{ps} were the ground truth and the corresponding patch-shuffled ground truth, respectively. The supervised loss L_{seg} between z_i and y_i, L_{seg}^{ps} between z_i^{ps} and y_i^{ps} and an unsupervised loss L_{cls} between z_i^{ps} and $z_i^{ps'}$ were used to optimize the segmentation model.

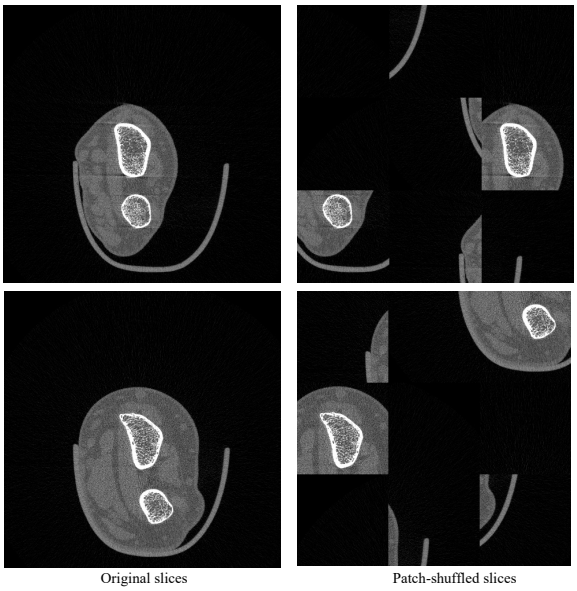

Fig. 5.4 Examples of original slices (left) and patch-shuffled slices (right) of wrist CT.

The contributions of our work were summarized as follows:

- Based on bone structures and CT imaging characteristics, we defined the bone CT segmentation as a local-feature-guided task. We proposed a patch-shuffled data transformation method to enable the segmentation model to segment both original and patch-shuffled CT slices.
- We developed a patch-shuffle-based semi-supervised segmentation method for bone CT segmentation. The proposed method could employ the unlabeled data and largely alleviate the workload of annotation. Two supervised segmentation losses and an unsupervised consistent loss were employed to optimize the segmentation model. The unsupervised consistent loss performed as a regularization item and enabled the model to utilize both the labeled and unlabeled data.
- We evaluated the proposed model on three CT scan datasets of different image qualities and different anatomies (wrist, foot, chest, head, abdomen, and limb). The results demonstrated the outperformance of the proposed model in bone CT segmentation.

5.2 Method of Patch-Shuffle-Based Semi-Supervised Segmentation of Bone CT

5.2.1 Overview of the Proposed Method

Fig. 5.3 overviewed the framework of the proposed patch-shuffle-based bone CT semi-supervised segmentation method. The framework consisted of two streams, an original slice training stream, and a patch-shuffled slice training stream.

The training dataset consisted of N bone CT slices, including M labeled slices $\{x_1, x_2, \ldots, x_M\}$ with M segmentation groundtruth $\{y_1, y_2, \ldots, y_M\}$ and $N - M$ unlabeled slices $\{x_{M+1}, x_{M+2}, \ldots, x_N\}$. The patch shuffle transformation divided each training slice x_i into nine patches and generated a new slice x_i^{ps} by randomly reconstituting these patches. The segmentation model $f(\cdot)$ with parameter θ was trained

jointly by three losses, two supervised constraint losses L_{seg}, L_{seg}^{ps} and a unsupervised constraint loss L_{cls}.

$$\min_{\theta} L_{seg} + L_{seg}^{ps} + \lambda L_{cls} \tag{5.1}$$

The supervised loss L_{seg} was used for the original slice training stream, which enabled the model to segment the original slice. The supervised loss L_{seg}^{ps} for the patch-shuffled slice training stream was inferred from the patch-shuffled slice x_i^{ps} and the corresponding ground truth y_i^{ps}. It endowed the model with the segmentation ability of the reconstituting slice. The unsupervised consistent loss L_{cls} was used for both labeled and unlabeled data to force the model to generate the same outputs for the corresponding pixels in x_i and x_i^{ps}, and λ was the weight parameter between the supervised losses and the unsupervised loss.

5.2.2 Patch-Shuffle-Based Semi-Supervised Segmentation

A well-trained segmentation model from the original slice could not directly be used for the patch-shuffled slices. Fig. 5.5 depicted the segmentation results of the original slice and patch-shuffled slice of a segmentation model trained from the original slice. Image one and image two were the original slice and corresponding segmentation result. Image three and image four were the segmentation results of the patch-shuffled slice and the corresponding segmentation result. Image five was the patch-reversed segmentation result of image four. As shown, image five was not the same as image two, which denoted that the segmentation model trained from the original slice was not robust for the patch-shuffled image. The inherent reason was that the neural network segmentation model using the original slice was not patch invariant.

The convolution operation of the segmentation model usually generated different features of the corresponding pixel in the original slice and the patch shuffled slice. Fig. 5.6 depicted the difference of the feature map of the original slice and patch-shuffled slice. Image two and image four were the feature maps of an original slice and the corresponding patch-shuffled slice. After the patch-reverse transformation of image four, it was obvious to observe the difference between the two feature maps. The

non-consistent feature of the corresponding pixel in the two kinds of slices has led to inconsistent segmentation results of Fig. 5.5.

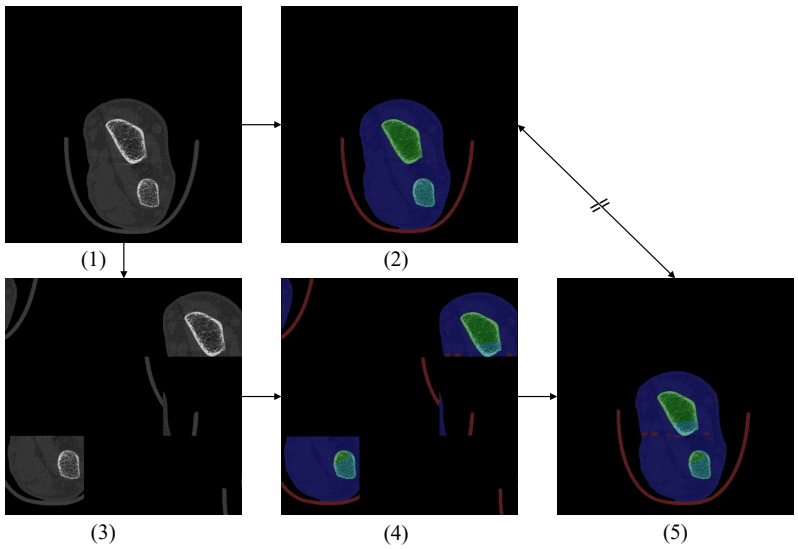

Fig. 5.5 Segmentation result illustration of original slice and patch-shuffled slice. Image one and image three were the original slice and patch-shuffled slice, respectively. Image two and image four were the corresponding segmentation result. Image five was the inverse patch-shuffled image of image four. Image five was not equal with image two.

To endow the model with the ability to learn the local features of bone structures that differentiated it from other tissues, we designed a two-stream training framework. The model made full use of both labeled and unlabeled data by an unsupervised consistency learning method. The framework of the proposed model was depicted in Fig. 5.3 and the pseudocode was shown in Algorithm 1.

For each input x_i, a patch-shuffled input x_i^{ps} was generated by the patch-shuffle operation. Both the slice x_i and the patch-shuffled slice x_i^{ps} were fed into the segmentation model and generated two outputs z_i and z_i^{ps}, respectively. For the labeled data, the segmentation ground truth map y_i was also required to generate a patch-shuffled ground truth map y_i^{ps} under the same patch-shuffle criteria. A supervised dice loss L_{seg} between the output z_i and the ground truth y_i was used to handle the standard segmentation process. A supervised loss L_{seg}^{ps} between the z_i^{ps} and the patch-shuffled ground truth map y_i^{ps} was used to strengthen the generalization ability of the segmentation

Algorithm 1 Pseudo code of the patch-shuffle-based semi-supervised segmentation method of bone CT.

Input:
Labeled data: $x_i \in D_L$, $y_i \in D_L$
Unlabeled data: $x_j \in D_U$
Model weights: θ
Weight parameter: λ
Function:
$f(x, \theta)$: neural network forward function
$update(.)$: backpropagation for model weights update
$PS(.)$: patch-shuffle transformation
L_{seg}, L_{seg}^{ps}: supervised dice loss calculation with prediction and groundtruth of the original slice and patch-shuffled slice, respectively
L_{cls}: unsupervised consistent loss calculation between the features of original and patch-shuffled slice
Procedure:
for $t \in [1, numepochs]$ **do**
  for each minibatch B **do**
    random update $PS(.)$
    $x_i^{ps} \leftarrow PS(x_i); y_i^{ps} \leftarrow PS(y_i), x_j^{ps} \leftarrow PS(x_j)$
    $z_i \leftarrow f(x_i, \theta); z_i^{ps} \leftarrow f(x_i^{ps}, \theta)$
    $z_j \leftarrow f(x_j, \theta); z_j^{ps} \leftarrow f(x_j^{ps}, \theta);$
    $z_i^{ps'} \leftarrow PS(z_i); z_j^{ps'} \leftarrow PS(z_j)$
    $loss \leftarrow L_{seg}(y_i, z_i) + L_{seg}^{ps}(y_i^{ps}, z_i^{ps}) + \lambda(L_{cls}(z_i^{ps}, z_i^{ps'}) + L_{cls}(z_j^{ps}, z_j^{ps'}))$
    $\theta \leftarrow update(loss)$
  end for
end for
return θ

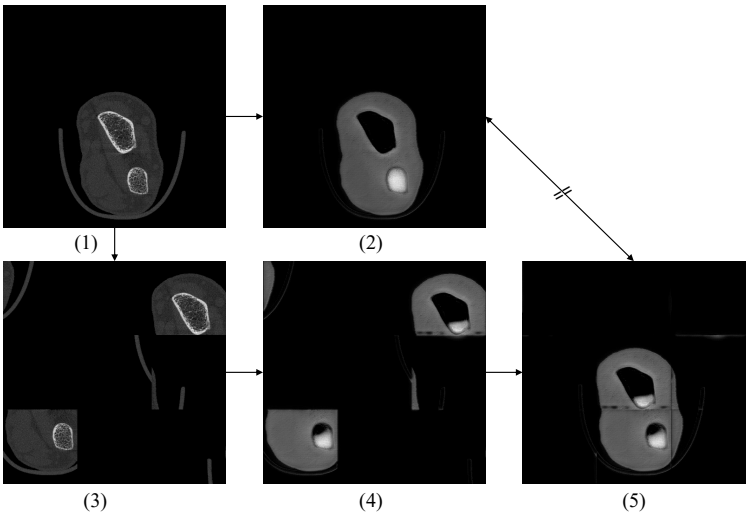

Fig. 5.6 Feature map illustration of original slice and patch-shuffled slice. Image one and image three were the original slice and patch-shuffled slice, respectively. Image two and image four were the corresponding segmentation feature map. Image five was the inverse patch-shuffled feature map of image four. Image five was not equal with image two.

model. The two losses forced the model to explore the exact feature that represented the bone structure. The dice loss acted as the supervised loss for the labeled data.

$$L_{seg}(y_i, z_i, \theta) = 1 - \frac{y_i z_i + 1}{y_i + z_i + 1} \tag{5.2}$$

$$L_{seg}^{ps}(y_i^{ps}, z_i^{ps}, \theta) = 1 - \frac{y_i^{ps} z_i^{ps} + 1}{y_i^{ps} + z_i^{ps} + 1} \tag{5.3}$$

For both unlabeled data and labeled data, the output z_i was transformed into a patch-shuffled output feature of $z_i^{ps'}$ by the same patch-shuffle criteria. The feature z_i^{ps} was the output feature of the patch-shuffled slice x_i^{ps}. $z_i^{ps'}$ was the patch-shuffled output feature of the original input x_i. Based on the assumption that the model should have the ability to learn the local features of bone structures regardless of the different sizes and shapes, the two features z_i^{ps} and $z_i^{ps'}$ should be the same. An unsupervised consistent loss L_{cls} between z_i^{ps} and $z_i^{ps'}$ was used to maximize the similarity of the

two features. The mean square error loss was used as the unsupervised consistent loss on both the unlabeled and labeled input.

$$L_{cls}(z_i^{ps}, z_i^{ps'}, \theta) = \| z_i^{ps} - z_i^{ps'} \|_2 \tag{5.4}$$

A weighting factor λ was used to adjust the impact of the supervised loss and the unsupervised loss. We set the weighting factor λ as five in the experiment settings. The model was optimized by the loss as Eq. 5.5.

$$\begin{aligned} L(x, \theta) = \sum_{i=1}^{M} (L_{seg}(y_i, z_i, \theta) + L_{seg}^{ps}(y_i^{ps}, z_i^{ps}, \theta)) \\ + \lambda \sum_{i=1}^{N} L_{cls}(z_i^{ps}, z_i^{ps'}, \theta) \end{aligned} \tag{5.5}$$

5.3 Experiments and Analysis

5.3.1 Datasets for Model Evaluation

Three datasets have been used to evaluate the proposed method. The first dataset was a wrist CT scan dataset, the second dataset was a foot bone CT dataset, and the third dataset was a multi-organ bone dataset (USEvillaBone) [189]. The three datasets were selected to cover a broad range of imaging modalities, image quality, and anatomies to demonstrate the robustness and performance of the proposed method. The details of the datasets, including the data amount, scanning device, scan area, and slice resolution, were described below.

Wrist CT Scan Dataset

The wrist CT scan dataset consisted of 5043 slices, including a training set with 4287 slices, a validation set with 252 slices, and a testing set with 504 slices. The scanning area was the radius and ulna part of the wrist. The CT slices were acquired by the ScancoXtreme machine (SCANCO Medical, Brüttisellen, Switzerland), and the slice resolution was 1536×1536. The 5043 slices were selected randomly from 1539

wrist scans, and the wrist parts were in different shapes, positions and sizes in these slices. Therefore, there were no temporal relationships among them. Fig. 5.7 was the illustration of the scanning slice and the segmentation ground truth. The radius bone, ulna bone, muscle, background, and cast holder were annotated in this dataset. We resized the slice resolution as 480×480 in experiments.

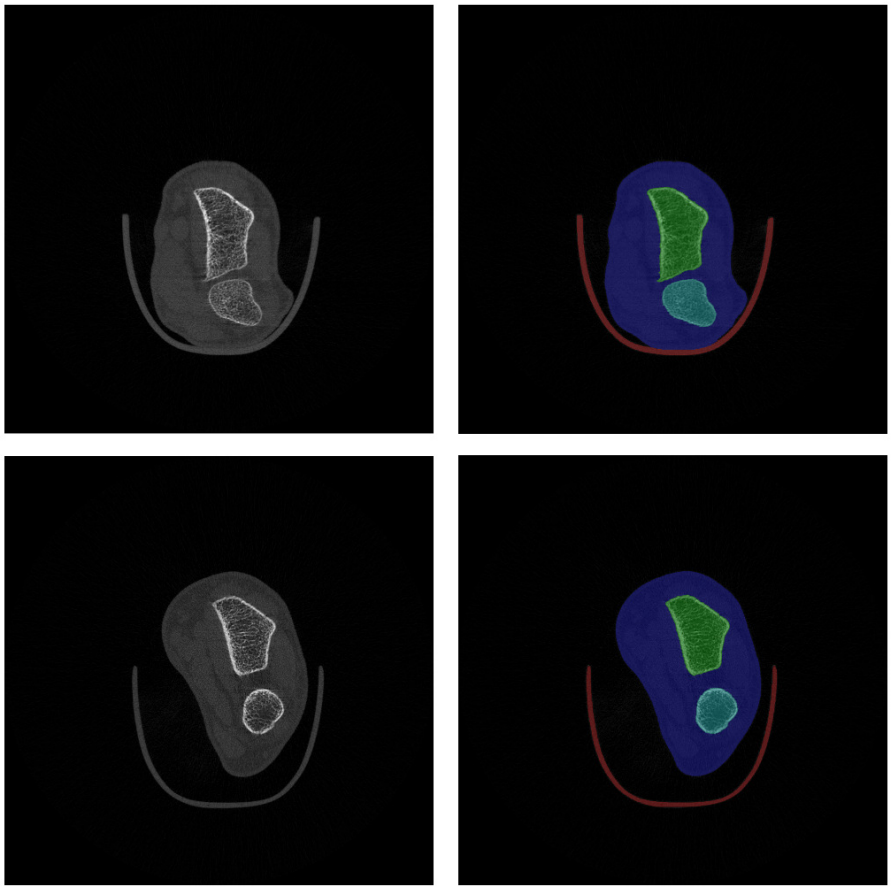

Fig. 5.7 Examples of the wrist CT scan dataset (The left parts were the CT slice data and the right parts were the overlap image of the segmentation mask on the CT slice. The green part was radius bone, the cyan part was ulna bone, the blue part was muscle, the red part was cast holder and the black part was background).

Foot CT Scan Dataset

The foot CT scan dataset included 1500 slices of foot CT images from thirty scans. Eleven scans were the left foot CT scan, eleven scans were the right foot CT scan, and eight scans were the two-feet scan. The foot CT scan dataset contained a train set with 1050 slices, a validation set with 150 slices, and a testing set with 300 slices. The CT slices were acquired by the CurveBeam weight-bearing machine (Curvebeam, Hatfield, United States). The slice resolution was 950×950 and was resized as 480×480 in experiments. The bone part has been annotated in this dataset. Fig. 5.8 illustrated the examples of the slices and the segmentation maps.

The USEvillaBone dataset

The USEvillaBone dataset [189] contained 270 slices of CT images. The scanning parts were the head, chest, abdomen, and limbs. The helical CT scanner (Philips, Amsterdam, The Netherlands) was used as the scanning device, and the slice resolution was 512×512. 170 slices, 50 slices, 50 slices were used as the training set, validation set, and testing set, respectively. The slices were resized as 480×480 for consistency with the other two datasets. Fig. 5.9 listed several examples of the slice and the corresponding ground truth of this dataset.

5.3.2 Experiment Settings and Evaluation Metrics

The U-net model was used as the network backbone. The input layer was a 1-channel convolutional layer, and the output layer was a 1×1 convolutional layer with an output channel number of five for the instance segmentation of the wrist database, an output channel number of two for bone segmentation of the foot database and the USEvillaBone database. The Adam optimizer was used to optimize the model. The learning rate was set as 0.00002. The experiments were implemented using PyTorch. Random flipping was used as the base data augmentation method before the patch shuffle transformation for all experiments. In the inference stage, only the original slice was used as the input. A softmax layer was used to generate the probability map of

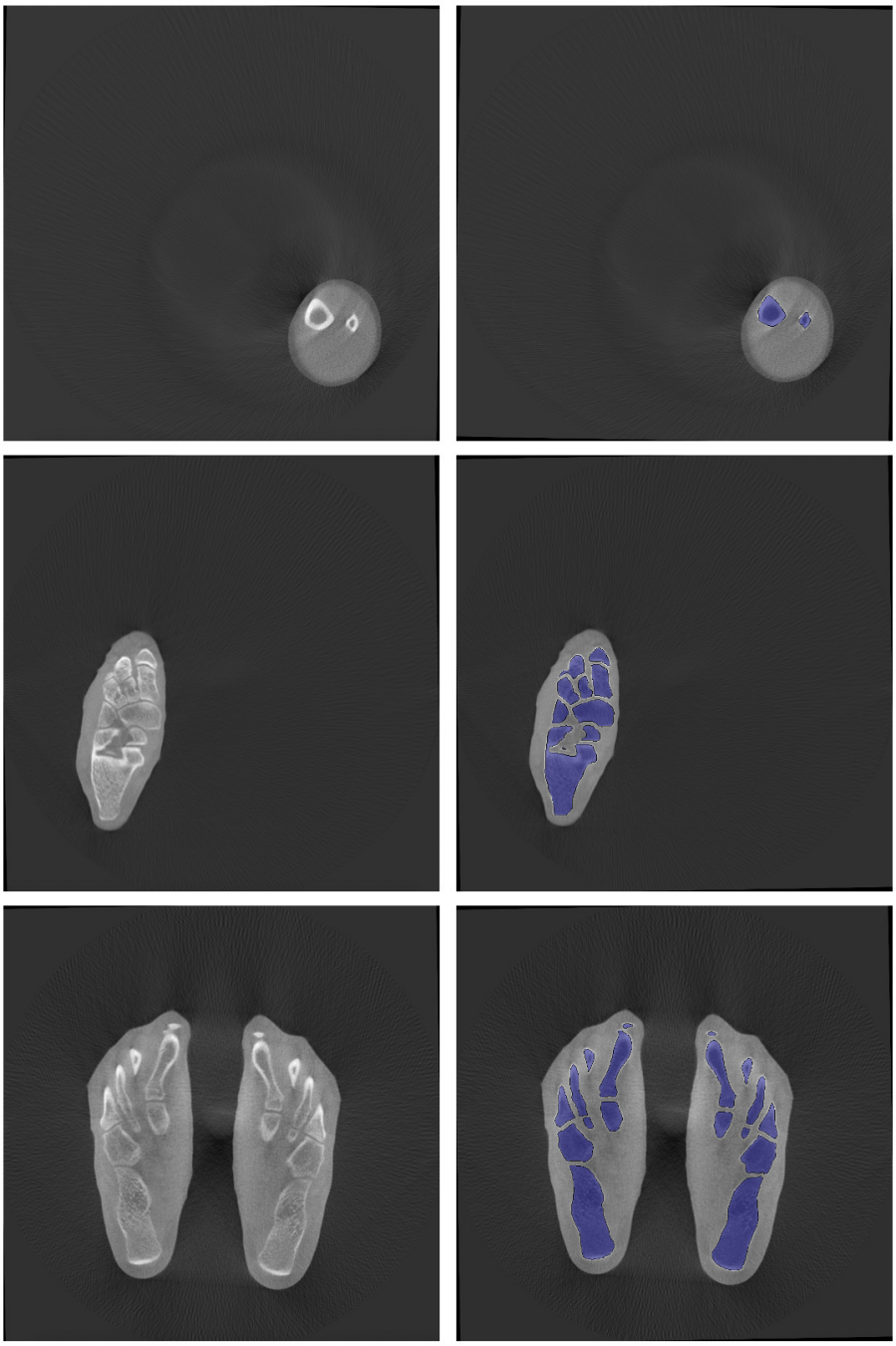

Fig. 5.8 Examples of the foot CT scan dataset (The left parts were the CT slice data and the right parts were the overlap image of the segmentation mask on the CT slice. The bones were illustrated as the blue color in the overlap images. First row was right foot example, second row was left foot example and third row was two-feet example).

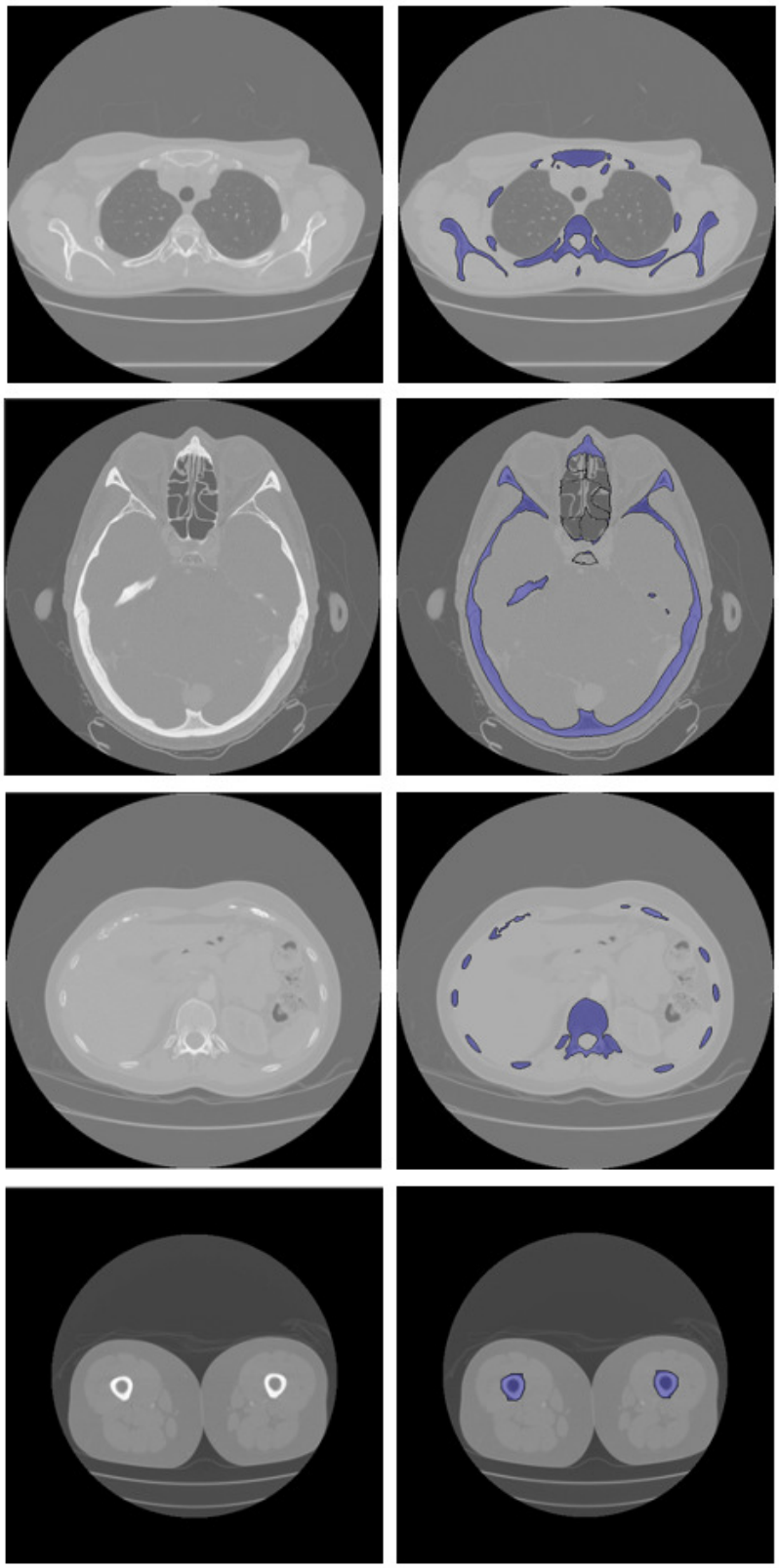

Fig. 5.9 Examples of the USEvillaBone dataset (The left parts were the CT slice data, and the right parts depicted the overlap image of the segmentation mask on the CT slice. The bones were illustrated as the blue color in the overlap images. The first row was an abdomen example, the second row was a brain example, the third row was a chest example, and the fourth row was a limb example).

each class, and the class with the maximum probability of each pixel was selected as the prediction result.

As the illustration in Fig. 5.7, the wrist bone CT slices were annotated into five classes: radius bone (the green part in the overlap image), ulna bone (the cyan-blue part in the overlap image), muscle (the blue part in the overlap image), cast holder (the red part in the overlap image) and background (the black part in the overlap image). The radius bone was larger than the ulna bone in the CT slice. We designed a five-class instance segmentation task for the wrist bone CT dataset.

We designed a foot bone segmentation task in the Foot bone CT dataset. As illustrated in Fig. 5.8, there were different kinds of bone at the foot scan, such as tibia, fibula, cuneiforms, metatarsals, and phalanges. Since these bones were in different shapes, the foot bone segmentation task was more complicated than the wrist segmentation. However, these bones were still within a similar bone structure, making the semi-supervised method a potential solution for foot bone segmentation.

The USEvillaBone dataset consisted of four kinds of organs, brain, abdomen, chest, and limb. This dataset was more complicated than the wrist and foot dataset since the bones were in different shapes and organs. A bone segmentation task was designed on the USEvillaBone dataset to prove the outperformance of the proposed model.

We compared our method with the recent semi-supervised segmentation methods: mean-teacher (MT) [190] and the Transformation-consistent Self-ensembling Model (TCSM) [187] on the three datasets. All models were trained with the same model architecture and data splitting for fairness. The mean Intersection-Over-Union ($mIOU$) index was used to evaluate the performance of the segmentation model. The higher the $mIOU$ value, the better the segmentation model.

We also conducted more experiments on the wrist dataset to verify the effectiveness of the proposed method, including experiments with only two labeled training data, which was extremely low annotated data amount, and experiments with different number of labeled training data to prove that our semi-supervised method with less labeled data could be compatible with the supervised method with more labeled data. The results were reported in the following sections.

Table 5.2 Comparison of mIOU of different semi-supervised methods under different number of training data. (Unit: %)

Method	Two labeled / 100 unlabeled data	Seven labeled / 100 unlabeled data	All labeled (4287) data
Supervised-only	39.37	44.75	98.13
MT	85.24	94.10	—
TCSM	84.95	93.97	—
Ours	93.44	95.95	—

5.3.3 Results on the Wrist CT Scan Dataset

Comparison with Other Semi-Supervised Methods

Table 5.2 showed the performance comparison of our method and the other semi-supervised methods under two labeled data/100 unlabeled data and seven labeled data/100 unlabeled data. From table 5.2, we could observe that our method outperformed the other methods. Especially when only two labeled data were used, the proposed method gained a more remarkable improvement, obtaining a $mIOU$ of 93.44%. The results demonstrated the efficiency of the proposed method compared with the other semi-supervised method.

Results with Two Labeled Training Data

According to the position of the wrist, the wrist data could be divided into two categories. The first type was where the wrist was on the cast holder, which was depicted at the first row in Fig. 5.7, and the second type was where the wrist was above the cast holder as depicted at the second row in Fig. 5.7. Therefore, we tested the model performance with only two labeled data (one for each kind) in the training set, which was an extremely low data amount.

Table 5.3 demonstrated the performance of the supervised method, supervised method with patch-shuffle augmentation, supervised method with patch-shuffle and consistent loss, the proposed method with 100 unlabeled data, the proposed method with all unlabeled data (4285 slices) on the test set with 504 slices. All the experiments used the same training settings for a fair comparison.

Table 5.3 Performance of mIOU on the test set (504 slices) of the supervised and the proposed semi-supervised methods using two labeled training data. (Unit: %)

Method	Two labeled data	Improvement
Supervised-only	39.37	
Supervised-patch shuffle	47.92	8.55
Supervised-patch shuffle and consistent loss	74.31	34.94
Our method with 100 unlabeled data	93.44	54.07
Our method with all unlabeled data (4285 slices)	94.21	54.84

While only two labeled data were used for training the model, our method with only 100 unlabeled data has achieved a $mIOU$ of 93.44%. Compared with the supervised method, the proposed semi-supervised method has achieved a prominent improvement of 54.07% with 100 unlabeled data and improvements of 54.84% with all 4285 unlabeled data. Besides, compared with the supervised-only model, the supervised method with patch shuffle augmentation gained an improvement of 8.55%, and the supervised method with patch shuffle and consistent loss regularization gained a further improvement of 26.39%. The introduction of the patch-shuffle transformation and the consistent loss improved the performance of the supervised model. The results comparison demonstrated that our proposed method could achieve high segmentation accuracy and proved the efficiency of the proposed method compared with the supervised method.

Results under Different Number of Labeled Training Data

Table 5.4 was the results of the proposed method and the supervised-only method under the different number of labeled training data. Compared with the supervised method, the proposed semi-supervised method achieved much better performance. When only two labeled slices were used for training, our method achieved the $mIOU$ of 93.44% with 100 unlabeled data, which was an improvement of 54.07% compared with the supervised-only method. When seven labeled slices were used for training, our method also achieved an improvement of 51.20% and 52.41% with 100 unlabeled data and all unlabeled data compared with the supervised-only method, respectively. The

Table 5.4 Results of mIOU of the proposed method under different number of labeled training data. (Unit: %)

Method	Two labeled data	Seven labeled data	All labeled (4287) data
Supervised-only	39.37	44.75	98.13
Our method with 100 unlabeled data	93.44	95.95	—
Our method with all unlabeled data	94.21	97.16	—

improvements demonstrated the efficiency of our proposed semi-supervised method. With only a few labeled data, the patch-shuffled transformation and the consistent loss regularization improved the performance of the segmentation model.

With more labeled training data, both the supervised method and the semi-supervised method could achieve better performance. The performance of the supervised model trained from all labeled data was 98.13%. The proposed semi-supervised method gained a result of 97.16% from only seven labeled training data. The proposed method only has a decrease of 0.97% compared to the supervised model. This indicated the potential of the proposed method in real clinical practice where only a few data were labeled, and a large number of data were unlabeled.

5.3.4 Results on the Foot Bone CT Dataset

We reported the results of using only twenty labeled slices as training data, fifty labeled slices as training data, and all labeled slices as training data on the foot bone CT dataset. The results were summarized in Table 5.5. With only twenty labeled slices, our semi-supervised method could achieve an improvement of 23.19% compared with the supervised method. Compared with the segmentation model using all 1050 labeled data, the proposed semi-supervised model still achieved a high result of 93.63% with only 2% of the labeled data (twenty labeled data) and 96.49% with only 5% of the labeled data (fifty labeled data). We also compared our method with the other semi-supervised methods. The proposed method achieved the best result compared

Table 5.5 Results of mIOU of different semi-supervised methods on the foot bone CT dataset. (Unit: %)

Method	Twenty labeled / 1030 unlabeled	Fifty labeled / 1000 unlabeled	All (1050) labeled data
Supervised-only	70.44	93.14	98.40
MT	90.69	93.86	—
TCSM	91.14	94.39	—
Ours	93.63	96.49	—

Table 5.6 Results of mIOU of different semi-supervised methods on the USEvillaBone dataset. (Unit: %)

Method	Twenty labeled / 150 unlabeled	Fifty labeled / 120 unlabeled	All (170) labeled data
Supervised-only	78.84	88.94	92.91
MT	76.33	83.89	—
TCSM	81.46	85.90	—
Ours	82.35	91.64	—

with the others. The result revealed the potential of our approach for reducing the manual annotation workload in foot bone segmentation.

5.3.5 Results on the USEvillaBone Dataset

We conducted the bone segmentation task with different semi-supervised methods of using twenty labeled training data, and fifty labeled training data on the USEvillaBone dataset. We also compared the result of using all training data for supervised segmentation.

Table 5.6 reported the results of different semi-supervised methods on the public bone dataset. From the result, we could observe that the proposed method has outperformed the other semi-supervised methods. With only fifty labeled data (29.41% of all labeled data), the proposed method achieved a similar result with the supervised model of using all labeled data. The proposed method has achieved an improvement of 2.7% and 3.51% compared with the supervised method of using fifty labeled data and twenty labeled data, respectively.

5.3.6 Qualitative Analysis on the Segmentation Results and Feature Maps

Fig. 5.10 compared the results of the proposed method, MT method, TCSM method, and the supervised method with the same labeled slices. The first column depicted the wrist CT segmentation results via the proposed method, MT method, TCSM method, the supervised method using 2 labeled slices, and the ground truth, respectively. The second and third columns illustrated the segmentation results of foot CT and the USEvillaBone CT from the proposed method, MT method, TCSM method, the supervised method using 20 labeled slices, and the ground truth, respectively. Compared with the CT ground truth, the proposed method delivered more accurate and usable results than the supervised method and other semi-supervised methods with extremely low data amount.

We also compared the segmentation results and feature maps of the proposed method on the patch-shuffled slice and the original slice in Fig. 5.11. Compared with Fig. 5.5 and Fig. 5.6, the proposed method could generate correct segmentation results for both patch-shuffled slice and the original slice, and also generate consistent feature maps between them, due to the introduction of the supervised losses in Eq. 5.2 and Eq. 5.3, and the unsupervised consistent loss in Eq. 5.4.

5.4 Conclusion

We presented a patch-shuffle-based semi-supervised method for the bone CT segmentation in this chapter. Based on the bone structures and CT imaging characteristics, we defined the patch-shuffle transformation as a strong data augmentation technique for bone CT segmentation. The supervised losses which acted on the original data and the patch-shuffled data enhanced the generalization ability of the segmentation model. Further, we employed an unsupervised consistent loss on the output feature of the corresponding pixel in the original slice and the patch-shuffled slice. The proposed method has been evaluated on three different datasets. The experiment results demonstrated the high effectiveness of the proposed method on bone CT segmentation.

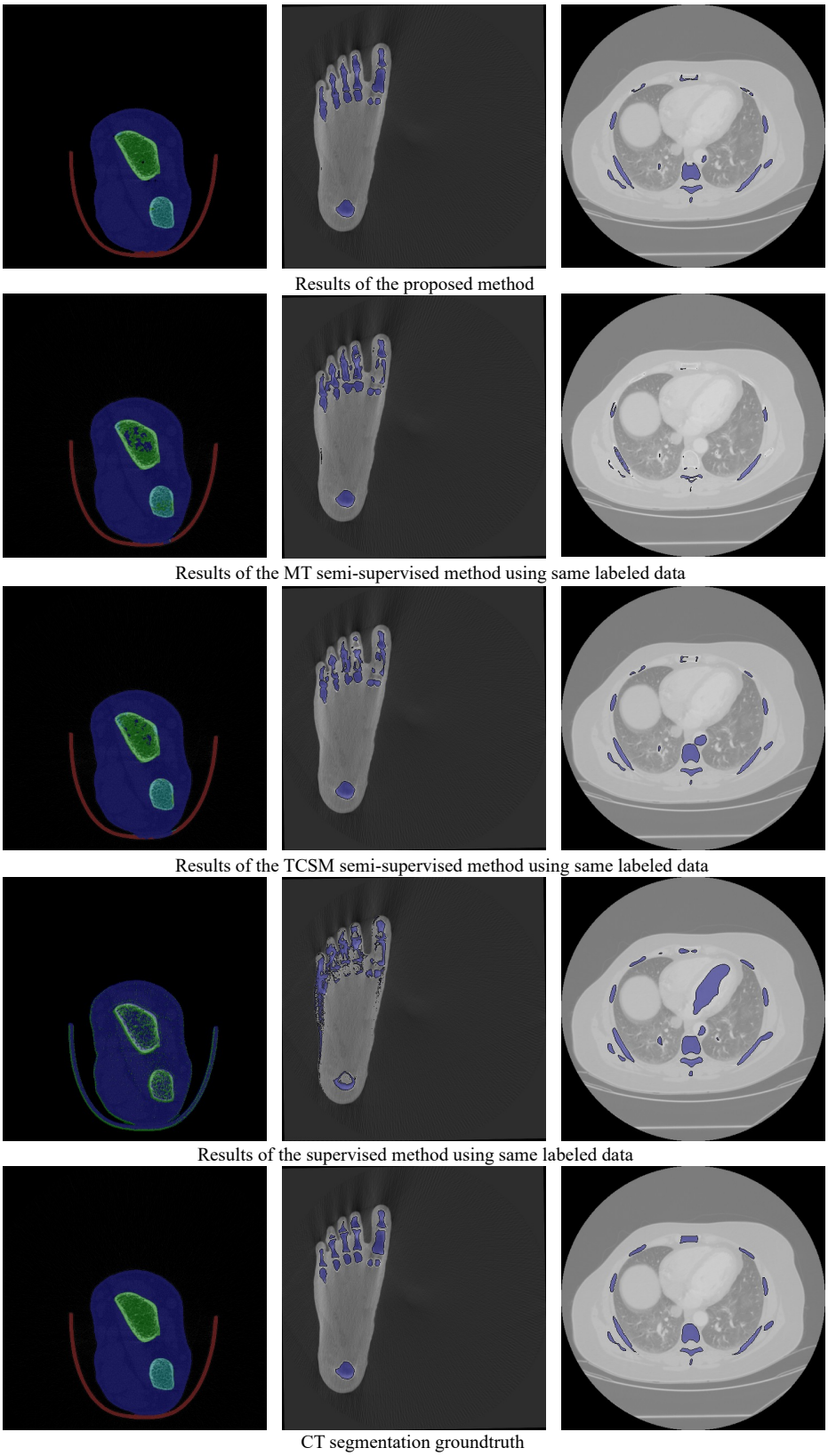

Fig. 5.10 Results comparison of the proposed method (first row), MT method (second row), TCSM method (third row), the supervised method with the same labeled slices (fourth row), and the segmentation groundtruth (last row). First column: models via 2 labeled wrist CT slices; Second column: models via 20 labeled foot CT slices; Third column: models via 20 labeled USEvillaBone CT slices.

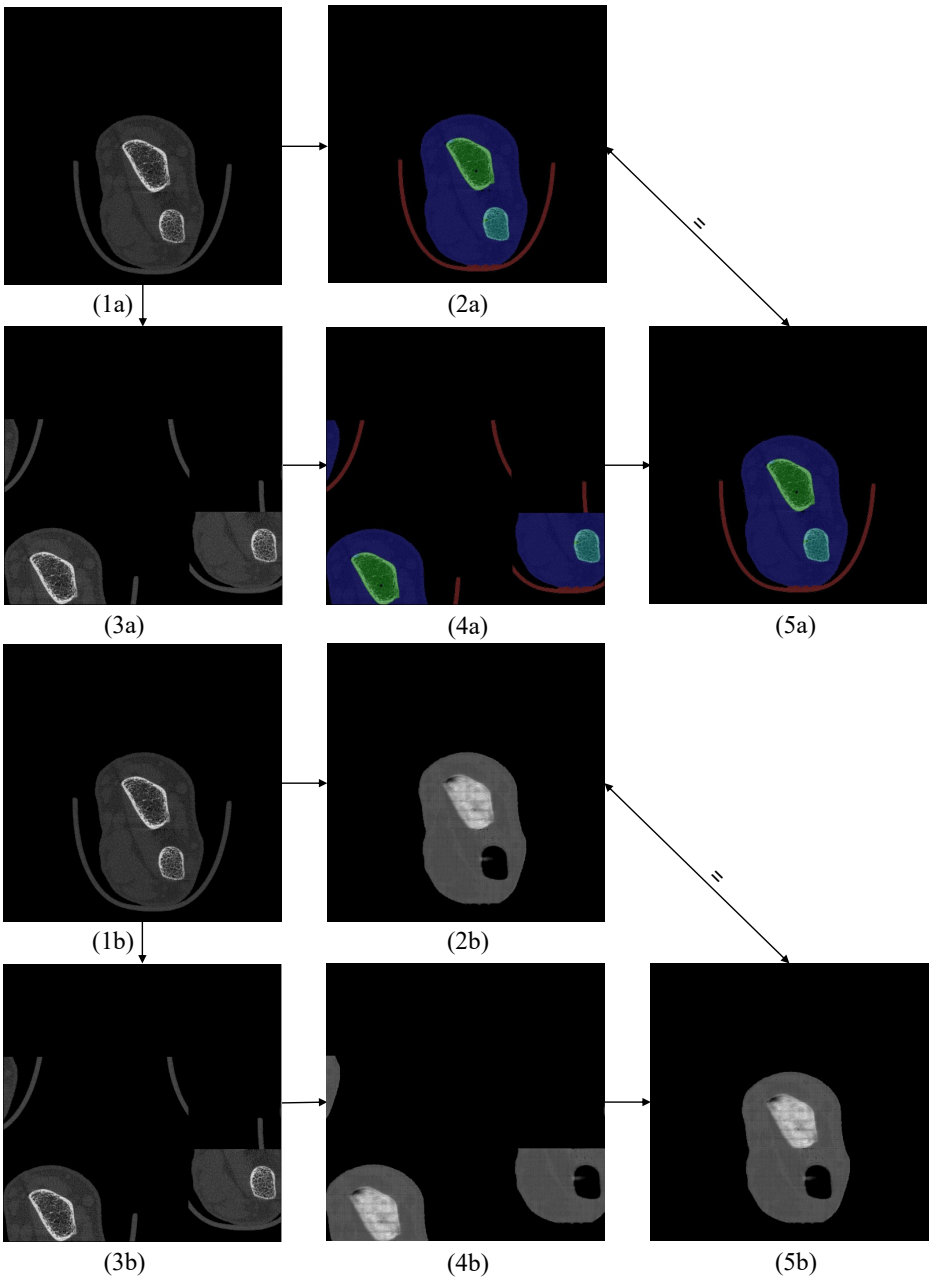

Fig. 5.11 Illustration of segmentation result and feature map of original slice and patch-shuffled slice using the proposed method. Image 1a and 3a, 1b and 3b were the original slice and patch-shuffled slice, respectively. Image 2a and 4a, 2b and 4b were the corresponding segmentation results and feature maps. Image 5a and 5b were the inverse patch-shuffled image and feature map of image 4a and 4b. Image 5a and 5b were consistent with image 2a and 2b.

Chapter 6

Bone Health Analysis via Bone Fracture Prediction using Wrist CT

6.1 Introduction

Fragility fracture, caused by the decrease of bone mass, is a major public bone health problem as lifespan increases. The bone fracture will highly increase the mortality rates in the first year and has an impact on the risk of death for up to ten years [43, 44]. Treatment based on the bone health assessment is an important way to prevent fracture, and two methods are used to determine bone health in clinical currently. The most common way is to measure bone mineral density (BMD). A BMD T-score of -2.5 standard deviation (SD) or lower indicates the presence of osteoporosis and a high risk of bone fracture [45]. The second method is to evaluate the fracture risks of patients by measuring the fracture risk assessment (FRAX) score [46]. The FRAX score integrates BMD at the femoral neck (FN) and several clinical risk factors to calculate the fracture risk in the next ten years.

The BMD assessment is based on the Dual-energy X-ray absorptiometry scanning in current medical practice. However, the operator needs to receive training and experience to analyze the X-ray images, and inappropriate diagnosis would be made due to the poor-quality X-ray images [191]. Some works investigated using bone CT scans to estimate the BMD value. Yasaka et al. [120] designed a CNN structure with

four consecutive convolutional and max-pooling layers and three fully connected layers to measure the BMD value from lumbar vertebrae CT. The input lumbar vertebrae CT images were manually cropped from the original CT images, and data augmentation methods like parallel shifting, rotation, and noise-adding were used to generate 60 different variations for each CT image. 1665 CT images of 183 patients from a medical institution were used to train the proposed method. The estimated BMD values of 45 CT scans of the same institution and 50 scans from another institution demonstrated a significant correlation with the real DXA BMD (correlation value of 85.2% and 84.0% for the two test sets, respectively) and the AUC of osteoporosis diagnosis on the two test set was 96.5% and 97.0%, respectively. Krishnaraj et al. [95] reported a deep learning method to simulate lumbar DEXA BMD scores from chest and abdomen CT with a two key module. The first module aimed to segment the four vertebrae bones (L1 $\sim$ L4) based on a cascade of two U-Net models. The first U-Net, performed in the sagittal view of CT, generated the binary segmentation map of vertebral bones. The second U-Net fused data, from the sagittal and coronal view and the first segmentation map of the first Unet, for the multiclass segmentation of the L1 $\sim$ L4 vertebra bones. The second module was the BMD calculation part based on linear regression and the vertebra segmentation results. The accuracy, sensitivity, and specificity for osteoporosis or osteopenia detection on 1389 patients were 82.0%, 84.4%, and 72.7%, respectively. These studies revealed the promising performance of employing CT data for BMD estimation.

However, only a small portion of the population suffers from osteoporosis (BMD T-score lower than -2.5 SD), while a large number of fractures occur within the stage of osteopenia where BMD T-score is between -2.5 and -1.0 SD or the stage of normal BMD where BMD T-score is above -1.0 SD [192]. A large population would be excluded when providing treatment to prevent bone fracture if the BMD T-score is employed to assess bone health. The FRAX score, considering the BMD and other clinical factors, only provided the fracture risk in the next ten years, which is too long for medical intervention. A more accurate bone health assessment, that considers both the large population with osteopenia or normal BMD and the small group of osteoporosis, is needed to avoid the fragility fracture.

Recent studies have attempted to apply the DL methods to classify the bone fracture, including humerus fracture [3], hip fracture [4], and wrist fracture [193]. These retrospective studies aimed to classify the fractured and non-fracture bone and can not determine bone health before the bone fractures happen. A prospective deep learning study that could predict bone fracture using CT data has not been explored so far.

The wrist CT has been used for bone health assessment in recent years [5, 6, 194]. Compared with the BMD T-score assessment from the spine or hip part, the wrist CT is more convenient during scanning. Johnson et al. [5] demonstrated that the HU value on capitate of wrist CT was highly correlated with the BMD T-score at pelvic, vertebra, and femoral neck. Chapurlat et al. [6] assessed bone health by measuring the deterioration of the bone microstructures on the wrist part and identified more patients with bone fracture risk than BMD T-score and FRAX in the osteopenia/normal population. These works revealed the high potential for using wrist CT in bone health assessment.

This chapter developed a deep learning-based method to determine bone health by predicting the future fragility fracture from wrist CT. We collected data from three population-based cohorts, which were followed for 6.16 years on average. The unstructured data from the three cohorts were processed to construct a structured wrist database for model designing. The proposed method could identify the bones at high risk of fracture before the fracture happens, and extensive evaluations have been performed. We tested whether the proposed method could recognize the participants at intermediate risk (five-years) of fragility fracture or major fragility fracture. We also compared the performance with the BMD T-score and FRAX score, considering different age groups. To the best of our knowledge, this was the first research of using the DL methods to identify the bones with the upcoming fragility fracture using the wrist CT scan.

The contributions of this work were summarized as follows:

- A deep learning-based fracture prediction model was proposed, and the real unstructured clinical raw data, including the patient clinical information and CT

Dicom data, was processed to construct a structured database for model training and evaluation.

- Extensive model evaluation experiments were conducted on different data selection groups considering the fracture type and participant's age. Compared with the clinical index of BMD T-score and FRAX, the proposed model could identify the bones with fragility fracture or major fragility fracture within five years with high AUC values.
- This was the first research using the deep learning models to identify the individuals with upcoming fragility fracture using wrist CT.

6.2 Structured Clinical Database Construction

6.2.1 Clinical and Wrist CT Raw Data Collection

We aimed to conduct a prospective study on bone fracture prediction. This required collecting CT and clinical data in the population and observing the fracture situation in the participants over a period of time. The data collection should consider the following criteria.

- The participants' age, gender, medical treatment condition, and possible duration of the follow-up period should be considered when recruiting candidates.
- Details of the fracture of the participants, such as when the fracture occurred, whether the fracture type was a fragility fracture, and whether it was a major or non-major fragility fracture, should also be recorded.
- The BMD T-score and the FRAX value of the participants should be collected for comparison with the developed model.

Following the criteria, we have collected three population-based cohorts, the OFELY cohort [195], the QUALYOR cohort [196], and the GERICO cohort [197].

OFELY Cohort

The OFELY cohort was collected by a large-population-based project that started in February 1992 in France [195]. 1039 women joined this project, and the wrist CT of 589 women (average age of 68.16, age range from 42 to 94) were scanned during 2006 $\sim$ 2008. The participants in the OFELY cohort were followed for 8.75 years on average (follow-up range from 0 $\sim$ 10.65 years). Participants in the OFELY cohort underwent only one scan at enrollment throughout the entire participation period.

QUALYOR Cohort

The QUALYOR cohort recruited 1539 women for bone health assessment [196]. The participants were from two cities in France, including 1042 women from Lyon and 497 from Orléans. The average age of the QUALYOR participants was 65.9, ranging from 50 to 87. The average follow-up duration was 5.67 years, ranging from 0.83 to 7.84 years. Each participant had one CT scan at enrollment, and some participants went through another CT scan during the follow-up period.

GERICO Cohort

The GERICO cohort was composed of 758 women, and 196 men, average aged at 65.1 (range from 62.9 to 68.1), in the Geneva retirees group of Switzerland [197]. The participants were followed from 2.0 to 7.9 years, with an average follow-up period of 5.0 years. The participants in GERICO went through several wrist CT scanning like the QUALYOR cohort.

The wrist part of the participants at OFELY, QUALYOR, and GERICO were scanned by the Scanco HRpQCT (Xtreme CT, Scanco Medical AG, Brüttisellen, Switzerland). Experts measured the Femoral neck BMD T-score and FRAX index at the enrollment of the participants at the three cohorts. The Femoral neck BMD T-score was also measured during the follow-up period for most participants in the QUALYOR cohort. The bone fracture time, fracture type, major fracture type during the follow-up period were recorded by experts in the three cohorts.

6.2.2 Clinical Raw Data Processing

The clinical data, such as ID, age, medical treatment, follow-up time, fracture information, etc., were recorded as unstructured data in different formats in the three cohorts. After we collected the clinical data, the first step was to process the unstructured raw data as structured data in the same format across the three cohorts.

Clinical Raw Data of OFELY

The collected raw data of the OFELY cohort included the participant's health information, bone measurement, and fracture history. Table 6.1 depicted the raw clinical data in OFELY. The participant's ID in OFELY cohort (ID), cohort name (Cohort), age (Age), medical treatment information (IsTreated), pre-fracture history before enrollment (HasPrevFx), fracture time in the follow-up period (FxWithinTime), Fragility fracture type (IsFragilityFx), major fracture type (IsMoF), follow-up duration (FollowupTime), wrist scan name (ScanName), femoral neck BMD T-score (TscoreFN), spine BMD (TscoreSpine), hip BMD (TscoreHip), FRAX score based on femoral neck BMD T-score (FraxMajorBMD), FRAX score based on hip (FraxHipBMD) and other bone measurements (TBS $\sim$ StraxSFS) were collected.

Table 6.1 Clinical raw data recorded at OFELY cohort.

ID	Cohort	Age	IsTreated
HasPrevFx	FxWithinTime	IsFragilityFx	IsMoF
FollowupTime	ScanName	TscoreFN	TscoreSpine
TscoreHip	FraxMajorBMD	FraxHipBMD	TBS
ScancoD100	ScancoDcomp	ScancoDtrab	ScancoBVTV
ScancoCtPo	StraxSPS	StraxSTS	StraxSFS

Clinical Raw Data of QUALYOR

The QUALYOR cohorts continuously collected the participant's clinical data during the whole follow-up period. In total, there were three clinical raw data table of QUALYOR cohort, which were depicted at Table 6.2, Table 6.3, Table 6.4.

For the first four years of the follow-up period, the QUALYOR cohort collected the baseline clinical data such as the participant's ID, age, gender, weight, height, BMI, medical treatment history, fracture history, fracture situation like the fracture type and major fracture type within four years follow-up period, BMD T-score at the femoral neck, FraxMajorBMD, the scan name at enrollment and other bone measurements. Table 6.2 depicted these recorded clinical information.

Table 6.2 Clinical raw data including the participant's information and fracture situation for the first four years of follow-up period recorded at QUALYOR cohort.

ID	Cohort	Age	Gender
Weight	Height	BMI	IsTreated
HasPrevFx	FxWithinTime	IsFragilityFx	IsMoF
FollowupTime	ScanName	TscoreFN	TscoreSpine
TscoreHip	FraxMajorBMD	FraxHipBMD	TBS
ScancoD100	ScancoDcomp	ScancoDtrab	ScancoBVTV
ScancoCtPo	StraxSPS	StraxSTS	StraxSFS

Then, after six years of the enrollment of each participant, the participant's bone fracture situation between four years to six years of the following up period was collected, as demonstrated in Table 6.3. The description of bone fracture (bone frx V6), bone fracture side of the human body (side of frx V6), fragility fracture information (fragility frx V6), vertebrae fracture information (Vert frx grade), fracture date (date frx V6), major fracture information (trauma degree V6), fracture site (code frx V6) were recorded. For participants with another fracture after the fourth year, a re-frx column was recorded.

Table 6.3 Clinical raw data of the participant's fracture information during the follow-up period from the fourth year that recorded at QUALYOR cohort.

ID V6	V6	Date V6	falls nb since V5 (48m)	Frx Yes/No		
bone frx V6	side of frx V6	fragility frx V6	Vert frx grade V6	date frx V6	trauma degree V6	code frx V6
bone re-frx V6	Side of re-frx V6	fragility re-frx V6	Vert re-frx grade V6	date re-frx V6	trauma degree re-frx V6	code re-frx V6

Besides, during the first four years of the follow-up period, the participant's spine, neck, and hip BMD T-score were recorded at twelve-month intervals, as listed in Table 6.4. The measurement date (date_V2 ∼ date_V5), and spine BMD T-score (SPINE_BMD_V2 ∼ SPINE_BMD_V5), femoral neck BMD T-score (NECK_BMD_V2 ∼ NECK_BMD_V5), hip BMD T-score (HIP_BMD_V2 ∼ HIP_BMD_V5) were recorded.

Table 6.4 Clinical raw data of the participant's BMD T-score at spine, femoral neck, and hip during the follow-up period recorded at QUALYOR cohort.

ID	date_V2	SPINE_BMD_V2	NECK_BMD_V2	HIP_BMD_V2
	date_V3	SPINE_BMD_V3	NECK_BMD_V3	HIP_BMD_V3
	date_V4	SPINE_BMD_V4	NECK_BMD_V4	HIP_BMD_V4
	date_V5	SPINE_BMD_V5	NECK_BMD_V5	HIP_BMD_V5

Clinical Raw Data Processing of GERICO

The clinical information of participants of the GERICO cohort was investigated twice during the whole follow-up period, one at the enrollment and one at the ending of the follow-up. The patient health information and the bone measurements were collected at the enrollment, and the fracture information was collected at the end of the follow-up period. Table 6.5 depicted the collected items of the clinical raw data. The participant's ID, date of birth (DOB), gender (Gender), age at enrollment (Baseline_age), BMI at enrollment (Baseline_BMI), medical treatment (Opdrug_duringfollowup) including the detailed drug type (Opdrug_duringfollowup_class), and the use of tibolone (MHTorTibolon_duringfollowup), follow-up duration (Followupduration), follow-up visit date (Followupvisit), femoral neck BMD T-score of at enrollment (Baseline_FemNeck_Tscore), Frax at enrollment (FRAX_MOF), and other bone measurements at enrollment (Baseline_Spine_BMD ∼ FRAX_BMDTBS_HIP) were recorded. The previous fracture information before enrollment (Priorlowtraumafracture adult), previous major fracture information before enrollment (PriorMOF45years), fracture information during follow-up period including the fragility fracture type (IncidentFracture), fracture site (Fr_site_1 ∼ Fr_site_5), fracture date (Fr_date_1 ∼ Fr_date_5), major fragility fracture type (Fr_trauma_1 ∼ Fr_date_5) were recorded. During follow-up, the maximum number of fractures in GERICO participants was five.

Table 6.5 Clinical raw data of the participant's health and fracture information recorded at GERICO cohort.

ID	DOB	Gender	Baselinevisit
Baseline_age	Baseline_BMI	Opdrug_ duringfollowup	Opdrug_during followup_class
MHTorTibolon_ duringfollowup	Followupvisit	Followupduration	Baseline_ FemNeck_ Tscore
FRAX_MOF	Baseline_ Spine_BMD	Baseline_ TotHip_BMD	Baseline_Rad_ 13d_BMD
Baseline_Rad_ ultradistal_BMD	Baseline_ Spine_Tscore	Baseline_ TotHip_Tscore	FRAX_HIP
FRAX_BMD_ MOF	FRAX_BMD_ HIP	TBS	FRAX_BMDTBS_ MOF
FRAX_BMDTBS_ HIP	Priorlowtrauma fractureadult	PriorMOF45years	IncidentFracture
Fr_site_1	Fr_date_1	Fr_trauma_1	Fr_site_2
Fr_date_2	Fr_trauma_2	Fr_site_3	Fr_date_3
Fr_trauma_3	Fr_site_4	Fr_date_4	Fr_trauma_4
Fr_site_5	Fr_date_5	Fr_trauma_5	

Integration of Clinical Data from OFELY, QUALYOR, and GERICO

The clinical raw data were collected in different formats in OFELY, QUALYOR, and GERICO cohorts. The unstructured data impeded the usage of the three cohorts, and we designed a procedure to integrate the clinical raw data of the three cohorts into structured clinical data. We designed several tables to record the structured clinical data, including a participant information table to record the participant's information of age, gender, enrollment date, follow-up duration, and medical treatment, a fracture information table to record the fracture information, a BMD information table to record the BMD values, a FRAX information table to record the FRAX values, and a miscellaneous information table to record the other information in the three cohorts. The details of these tables were depicted below.

1. Participant information table

The participant information table (as depicted in Table 6.6) recorded the information about participants in the three cohorts, including the cohort name (CohortName), participant ID (CohortID_PatientID), date of birth (DOB), gender (Gender), enroll-

ment date (EnrollmentDate), enrollment age (EnrollmentAge), follow-up duration (FollowupDuration), medical treatment (TreatmentBeforeEnrollment), and drug usage (TreatmentDuringFollowup).

Table 6.6 Clinical data in participant information table.

CohortName	CohortID_PatientID	DOB	Gender
EnrollmentDate	EnrollmentAge	FollowupDuration	
TreatmentBeforeEnrollment		TreatmentDuringFollowup	

2. Fracture information table

The fracture information of the three cohorts was recorded in different formats, and we formatted them as depicted in Table 6.7. The participant and cohort ID (CohortID_PatientID) was used to correlate with the other clinical tables. For each fracture of the participant during the follow-up period, we recorded the site of fracture (FxSite), date of fracture (FxDate), fracture time since enrollment (FxWithinTimeFromBaseline), fragility fracture or not (IsFragility), major fragility fracture or not (IsMoF), previous fracture information before enrollment (IsPrevalent), and previous fracture information before the recorded fracture during the follow-up period (HasPreFx).

Table 6.7 Clinical data in the fracture information table.

CohortID_PatientID	FxSite	FxDate	FxWithinTimeFromBaseline
IsFragility	IsMoF	IsPrevalent	HasPreFx

3. BMD information table

The BMD scores of each participant were organized at the BMD information table (Table 6.8). We used the CohortID_PatientID as the keyword to connect with the other tables. The VisitTag indicated whether the BMD was measured at enrollment (recorded as "baseline") or during the follow-up period (recorded as "follow-up"), and the MeasurementDate indicated the measurement time. The femoral neck BMD T-score(TscoreFN), spine BMD T-score (TscoreSpine), hip BMD T-score (TscoreHip), and other BMD scores (TBS $\sim$ RadUltradistalBMD) were recorded at the BMD information table. The femoral neck BMD T-score(TscoreFN) was the main clinical factor for bone health analysis and was used for comparison with the proposed model in the following section.

Table 6.8 Clinical data in the BMD information table.

CohortID_PatientID	VisitTag	MeasurementDate	TscoreFN
TscoreSpine	TscoreHip	TBS	SpineBMD
FemNeckBMD	TotHipBMD	Rad13dBMD	RadUltradistalBMD

4. FRAX information table

The FRAX scores of each participant were collected at the FRAX information table (Table 6.9). The CohortID_PatientID, VisitTag, and MeasurementDate had a similar meaning with the BMD information table. The FRAX scores, such as FraxMajorBMD, FraxHipBMD, and the other FRAX scores, were listed in the FRAX information table. FraxMajorBMD was compared with the proposed model in the following sections due to its wide usage in the clinical environment.

Table 6.9 Clinical data in the FRAX information table.

CohortID_PatientID	VisitTag	MeasurementDate	FraxMajorBMD
FraxHipBMD	FraxMOF	FraxHip	FraxBMDTBSMof
FraxBMDTBSHip			

5. Miscellaneous information table

Some clinical items were not recorded in all cohorts, such as the BMI, medical treatment details, and other unused bone measurements like ScancoD100. We listed them in the miscellaneous information table, as depicted in Table 6.10.

Table 6.10 Clinical data in the miscellaneous information table.

CohortID_PatientID	VisitTag	BMI	MHTorTibolonTreated
TreatedClass	ScancoD100	ScancoDcomp	ScancoCtPo
ScancoDtrab	ScancoBVTV	StraxSPS	StraxSTS
StraxSFS			

Finally, the clinical raw data were formatted as structured data in the above six tables. The structured data in the participant information table, fracture information table, BMD information table, and Fracture information table were used for the data selection, development and training of the deep learning model, and results evaluation in the following sections.

6.2.3 Wrist CT Raw Data Processing

Except for the clinical data, the wrist CT data was another type of raw data collected in the three cohorts. The wrist CT was scanned at each participant's enrollment in the OFELY cohort and the participant's enrollment and follow-up period in the QUALYOR and GERICO cohort. In the following sections, we used the baseline CT to indicate the CT scan at enrollment and the follow-up CT to indicate the CT scan at the follow-up duration.

Both the baseline CT and follow-up CT were in the Dicom data format, which contained the CT header and CT image data. The CT header contained the participant's information, scanning information, and the scanning machine parameter. The CT image data was stored as a data block, and the slice images could be extracted from it. Since the CT scans of the three cohorts were scanned on different machines, a phantom calibration procedure was performed to place the CT data of the three cohorts in the same distribution, and the phantom calibration file was stored in a calibration table.

We first extracted the CT header from the CT data and extracted the calibration parameter from the calibration file. The extracted information was used to create a scan information table to connect with the structured clinical data in the previous section. For the CT image data, we extracted the slice images of each CT scan and generated the instance segmentation map of each slice image.

Scan Info Table Generation from the CT Header and Calibration File

The CT header stored the raw data of participant's information, the scanning information, and the scanner information as depicted in Table 6.11. The pat_id indicated the participant's ID. The scan_id was the unique scan ID of each wrist CT, and scan_date was the scanning time. The slice number of each wrist CT, voxel size, and slice size were stored at total_slice_count, voxel_size, dimension_x, and dimension_y.

The calibration table included wrist CT file name, calibration slope, and calibration intercept. We used the CT file name as a keyword to link the calibration slope and intercept with the CT header table. The useful information from the CT header and the calibration table made up a scan info table, as depicted in Table 6.12.

Table 6.11 Data stored in the CT header.

scan_name	scan_type	scan_id	scan_site_id
total_slice_count	slice_index	dimension_x	scan_date
voxel_size	manufacturer	pat_no	meas_no
is_discarded	dimension_y	pat_id	workflow_type

Table 6.12 Scan data stored at the scan information table.

CohortID_PatientID	ScanName	ScanSite	ScanDate
VisitTag	ScannerManufacture	ScannerSite	ScannerMachineID
Slope	Intercept		

CohortID_PatientID was used to link with the other clinical tables. The VisitTag indicated whether the scan was a baseline scan or a follow-up scan, calculated according to the ScanDate value in this scan info table (Table 6.12) and the EnrollmentDate in the participant information table (Table 6.6). The ScanSite was the scanning part, and the ScannerSite was the scanner location. The Slope and Intercept were used to calibrate the data scanned by different machines using the Eq. 6.1.

$$Calibrated\_pixel\_value = \frac{Pixel\_value - Intercept}{Slope} \tag{6.1}$$

Anatomical Segmentation of CT Data

Each wrist CT contained 110 slices with the slice resolution of 1536×1536. The slices of each wrist CT were extracted from the Dicom data block first. The examples from the OFELY, QUALYOR, and GERICO cohorts were listed in Fig. 6.1.

We used the U-net model trained in Chapter 4 for the instance segmentation of the wrist CT. The segmentation mask containing the radius, ulna, muscle, cast holder, and background were generated on all wrist CT scans. Fig. 6.2 depicted several segmentation results. To check the segmentation results, we generated the stack image of the segmentation mask and the stack image of CT slices over the axial view by preserving the maximum pixel value. An overlap image was generated by putting the stack segmentation mask over the stack slice image. All the segmentation results

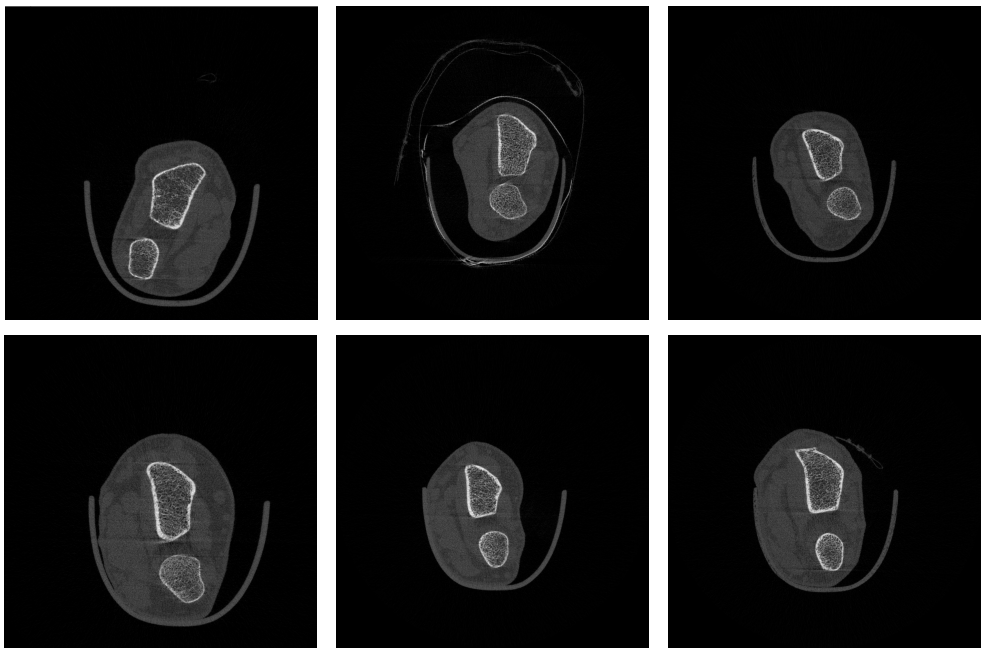

Fig. 6.1 Examples of the OFELY, QUALYOR and GERICO cohorts.

were manually checked by verifying the overlap images. The segmentation results demonstrated high accuracy and could be used for deep learning model development.

6.2.4 The Structured Wrist Database

After processing the clinical raw data and the CT data, we have established a wrist database with structured clinical information to develop the bone fracture prediction model to assess bone health. As depicted in Fig. 6.3, the established wrist database consisted of five clinical information tables, one scan information table, and one wrist CT scan database. The five clinical information tables and the scan information table were linked with the CohortID_PatientID and the VisitTag. The CT scans in the CT database were linked with the scan information table via the ScanName.

Since the women population was more vulnerable to bone fragility fracture, we focused on the women participants in the three cohorts. In total, there were 2666 women participants and 4768 wrist CT scans in the three cohorts. The five clinical

information tables of the women participants and the scan information table of the corresponding wrist CT scans were used to create a new table for the convenience of the later study. The items in the new table were depicted at Table 6.13.

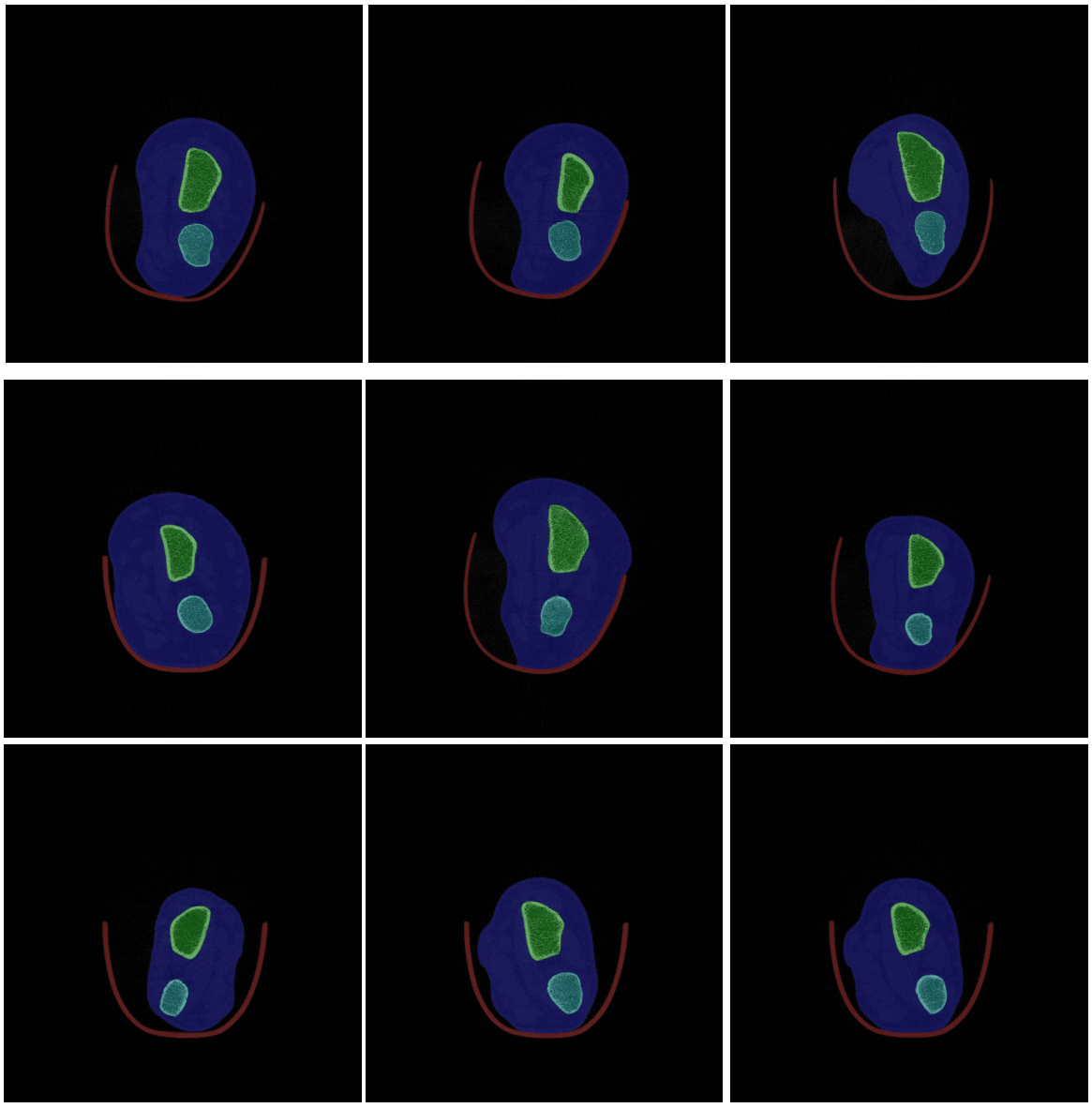

Fig. 6.2 Examples of the segmentation results.

Table 6.13 contained all the required clinical information and scan information for the development of the bone fracture prediction model. The other required data can be calculated from items in Table 6.13. For example, the participant's age at scanning

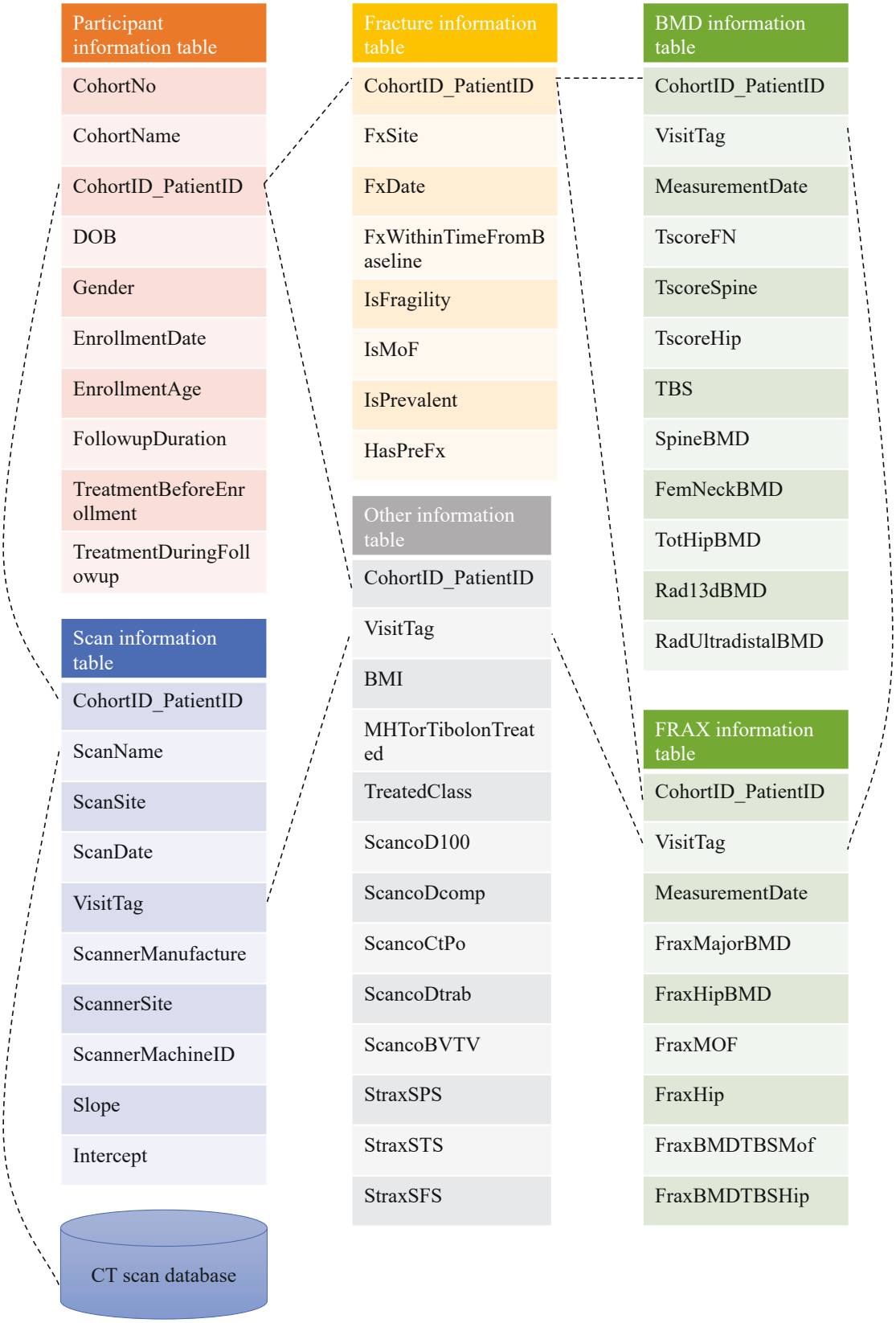

Fig. 6.3 The wrist database structure illustration. The five clinical information tables and the scan information table can be linked via the CohortID_PatientID and the VisitTag item. The scan information table use the ScanName to link with the CT scan database.

Table 6.13 The clinical information and scan information table of the women participants.

CohortID_PatientID	ScanName	ScanSite	ScanDate
VisitTag	ScanTimeSinceEnroll	DOB	Gender
EnrollmentDate	EnrollmentAge	FollowupDuration	IsTreated
HasPreFx	FxWithinTime	IsFragilityFx	IsMoF
ScanPath	NextFxWithinTime	BMD T-score	Frax
Slope	Intercept		

time could be calculated from EnrollmentAge and ScanTimeSinceEnroll. The BMD T-score indicated the BMD T-score value at the femoral neck, and the Frax denoted the FRAXMajorBMD value in the clinical usage.

6.3 Data Selection and Method of Bone Fracture Prediction in Next Five Years

6.3.1 Data Statistics of Age, BMD T-score, FRAX Score, Fracture and Non-Fracture Number in Different Years

We performed statistics on the data of the three cohorts to better understand the data distribution for data selection and model development. The total number (indicated as n under the cohort name), age (average and standard deviation (SD)), femoral neck BMD T-score (average and SD), FRAX score (average and SD), follow-up duration of participants (average and SD) in all cohorts and each cohort were listed at Table 6.14. The age of participants was around 66.11 years old. The femoral neck BMD T-score and FRAX score demonstrated that the three cohorts covered a wide range of the population in different bone health conditions. The average duration of follow-up was around 6.16 years, which was enough for the bone health assessment.

To develop the deep learning model for the bone fracture prediction, we also calculate the number of participants with bone fracture in different follow-up years. Table 6.15, Table 6.16, and Table 6.17 listed the number of participants with fracture

Table 6.14 Statistics of Age, BMD T-score, FRAX, and duration of follow-up of the cohorts.

	All cohorts (n= 2666)	OFELY, France (n = 568)	QUALYOR, France (n= 1427)	GERICO, Switzerland (n = 671)
Age (years) (average/SD)	66.1090 (6.4423)	67.9824 (8.5322)	65.8698 (6.7530)	65.0319 (1.4297)
BMD T-score (average/SD)	-1.5228 (0.7481)	-1.3598 (0.8198)	-1.6999 (0.5293)	-1.2704 (0.9639)
FRAX score (%) (average/SD)	8.0660 (5.6772)	7.9881 (6.8071)	6.3476 (3.8200)	11.9851 (6.1094)
Duration of follow-up (years) (average/SD)	6.1628 (2.1146)	8.8210 (1.8288)	5.6735 (1.2853)	4.9532 (1.8622)

from one $\sim$ ten years, and the number of participants without fracture from one $\sim$ ten years.

Table 6.15 listed the total number of participants with bone fracture in different years. There were 87 fracture cases in the first year of the follow-up period, 93 in the second year, 73 in the third year, 52 in the fourth year, and 42 in the fifth year. 31, 41, 17, 11, and 12 fracture cases occurred at the sixth, seventh, eighth, ninth, and tenth year, respectively. Only one fracture case occurred after the tenth year.

Table 6.16 depicted the fracture case number within different follow-up years. There were 347 fracture cases within the first five years of the follow-up period, and only 113 fractures happened after the first five years. Most of the fractures (75.43% of the fractures) happened in the first five years.

Table 6.17 listed the number of cases without fracture within different follow-up years after the scanning. 3694, 3552, 2444, 2058 cases did not fracture until the first, second, third, and fourth year ended after scanning. There were 1753 cases that were fracture free until the fifth year after scanning. With the follow-up period growing, the cases without fracture decreased quickly due to the ending of follow-up or bone fracture. 1664, 452, 391, 301, and 70 cases had no fracture until the sixth, seventh, eighth, ninth, and tenth year ended after scanning.

Table 6.15 Statistics of fracture number in different follow-up years.

Fragility fracture within	OFELY	QUALYOR	GERICO	Total
0∼1 year	16	53	18	87
1∼2 years	15	60	18	93
2∼3 years	11	47	15	73
3∼4 years	15	24	13	52
4∼5 years	7	22	13	42
5∼6 years	9	19	3	31
6∼7 years	20	13	8	41
7∼8 years	14	1	2	17
8∼9 years	11	0	0	11
9∼10 years	12	0	0	12
10 years∼ follow-up ends	1	0	0	1

Table 6.16 Statistics of fracture number within different follow-up years.

Fragility fracture within	OFELY	QUALYOR	GERICO	Total
0∼1 year	16	53	18	87
0∼2 years	31	113	36	180
0∼3 years	42	160	51	253
0∼4 years	57	184	64	305
0∼5 years	64	206	77	347
0∼6 years	73	225	80	378
0∼7 years	93	238	88	419
0∼8 years	107	239	90	436
0∼9 years	118	239	90	447
0∼10 years	130	239	90	459
0∼ follow-up ends	131	239	90	460

Table 6.17 Statistics of no-fracture number within different follow-up years.

No-fracture within	OFELY	QUALYOR	GERICO	Total
1 year	548	2286	860	3694
2 years	531	2182	839	3552
3 years	515	1278	651	2444
4 years	491	1189	378	2058
5 years	462	981	310	1753
6 years	450	956	258	1664
7 years	420	4	28	452
8 years	391	0	0	391
9 years	301	0	0	301
10 years	70	0	0	70

6.3.2 Data Selection According to Five Year Selection Criteria

The data selection was crucial to the model training. Enough positive cases (fracture cases) and negative cases (cases without fracture until the target follow-up period ends) should be selected for the model training according to the data selection criteria. Table 6.16, and Table 6.17 indicated that with the follow-up period extended, the fracture cases increased and no-fracture cases decreased. In order to get enough positive and negative data, we selected the scans which had fractures within five years after scanning as the positive cases and the scans which had no fracture for at least five years after scanning as the negative cases. In this way, we have selected 347 positive cases and 1753 negative cases. The statistics of the case number in different cohorts in the selected data were depicted in Table 6.18.

In total, 2100 cases were selected. For the positive cases, 64 scans were from the OFELY cohort, 206 scans were from the QUALYOR cohort, and 77 wrist scans were from the GERICO cohort. Among the 347 positive cases, there were 174 cases with major fragility fractures. For the negative scans, 462 scans were from the OFELY cohort, 981 scans were from the QUALYOR cohort, and 360 wrist scans were from the GERICO cohort.

Table 6.19 investigated the data distribution of the selected data in the three cohorts, including the age, BMD T-score, and FRAX. The average value and the standard deviation (SD) of age, BMD T-score, and FRAX were calculated. Since the

Table 6.18 Statistics of case number in different cohorts in the selected data with five year as data select criteria.

	All cohorts	OFELY, France	QUALYOR, France	GERICO, Switzerland
Total case number	2100	526	1187	387
Negative case number (no fracture for at least five years after scanning)	1753	462	981	310
Positive case number (fragility fractures within five years after scanning)	347	64	206	77
Positive case number (major fragility fractures within five years after scanning)	174	42	121	11

BMD T-score or FRAX value of some cases was missing during data collection, the total number of cases with recorded BMD T-score or FRAX were also calculated. Table 6.19 revealed that the age, BMD T-score, and FRAX values were different among the different cohorts, which indicated that the data distributions of the three cohorts were not identical. Therefore, we could not use one or two cohorts as training and test the model on the others. Instead, we mixed the three cohorts for the model development and evaluation in the following sections.

6.3.3 Method of Multi-Task Based Bone Fracture Prediction in Next Five Years

The multi-task learning strategy could extract more discriminative features and both the age and the maximum non-fracture year were highly correlated with the occurrence of bone fracture. With the age growing, the bone would lose gradually and if the non-fracture year was short, the bone loss would be more severe. Inspired by the multi-task learning, we modified the Densenet model [61] for the fragility fracture prediction and proposed a multi-task based bone fracture prediction framework as depicted in Fig. 6.4.

As shown in Fig. 6.4, the structure of the fracture prediction model was based on the DenseNet121 [61] model. The input was 110 CT scan slices after preprocessing to

Table 6.19 Statistics on age, BMD T-score and FRAX of the selected data according to the five-year criteria.

		Age (average/ SD/ case number)	BMD T-score (average/ SD/ case number)	FRAX (average/ SD/ case number)
All participants	Fracture	68.65170 (7.3828) 347	-1.7556 (0.6881) 320	9.8327 (7.3489) 260
	Non-fracture	65.6169 (6.3426) 1753	-1.5119 (0.7231) 1749	7.4476 (5.2817) 1745
OFELY	Fracture	72.1719 (8.6015) 64	-1.7373 (0.7899) 64	11.6625 (7.8863) 64
	Non-fracture	66.7316 (7.9811) 462	-1.2821 (0.8076) 458	7.0387 (5.7530) 457
QUALYOR	Fracture	68.4135 (7.8024) 206	-1.8260 (0.5109) 206	7.1685 (4.5261) 146
	Non-fracture	65.2743 (6.3613) 981	-1.6997 (0.5167) 981	6.1216 (3.6034) 978
GERICO	Fracture	66.3632 (2.3618) 77	-1.4887 (1.0314) 50	15.2700 (9.3182) 50
	Non-fracture	65.0394 (1.3937) 310	-1.2572 (0.9415) 310	12.2339 (6.2108) 310

remove the cast holder, and the DenseNet121 model was used as the backbone for scan feature extraction. A multi-task learning strategy, which employed the scan feature to predict the fracture situation of the patient within five years after CT scanning, predict the age of the patient at scanning date, and the longest bone health year (non-fracture year, i.e., how many years would the patient stay healthy without fracture during the following up period) was applied to train the fracture prediction model. The preprocessing of input CT slice data, the data augmentation for model training, and transfer learning for model training were described in the following sections.

CT Slice Preprocessing

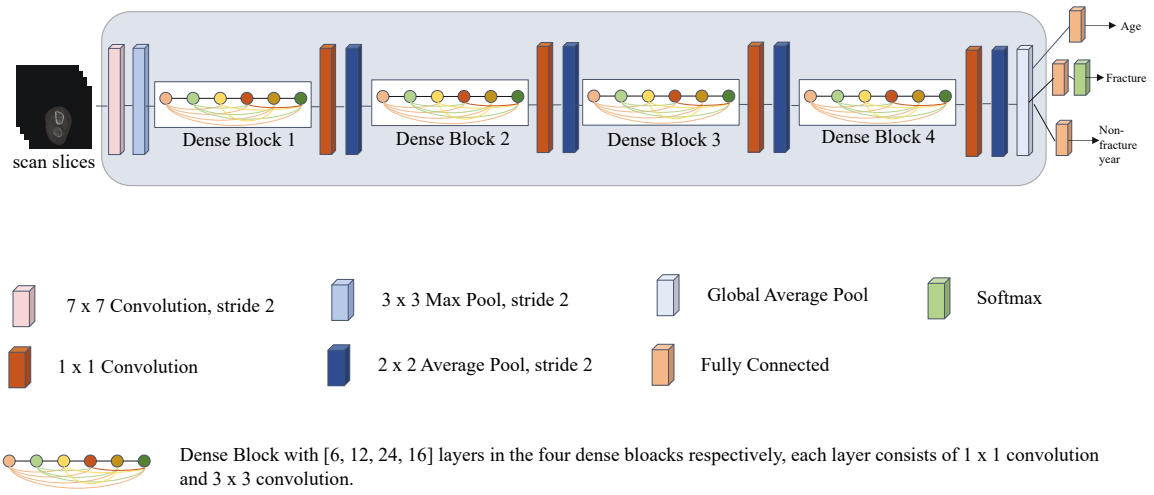

Fig. 6.4 Model structure of the fracture prediction model. The 110 scan slices were the input and only the wrist parts including the muscle, radius, and ulna have been used as input while the cast holder was removed. The DenseNet121 was used as the backbone and the output feature after the global average pool was a 256-dimension feature. A multi-task learning model used the 256-dimension feature for age prediction, fracture prediction and longest bone health year (non-fracture year) prediction.

The wrist CT consisted of muscle, radius bone, ulna bone, cast, and background. The cast was made of carbon fiber which was used to hold the wrist during scanning.

The pixel value of the cast part was within the same range as the muscle part in the CT slice. We have used the U-net model to generate the segmentation mask of the wrist slice during the database construction. The cast parts were removed in the 110 input slices using the segmentation map from the U-Net model. The size of the original CT slice was 1536 × 1536, while the wrist part only occupied part of the slice. We cropped a patch of size 840 around the wrist part to increase the wrist part ratio in the input slices. Since the parameter of CT value was not the same in different CT machines, the intercept and slope value of the CT scan was used to calibrate the pixel value as Eq. 6.1.

Data Augmentation

Data augmentation was widely used to avoid overfitting and train a more robust neural network model. The CNN model could be more invariant to rotation, scaling, and translation after the data augmentation. For the proposed fracture prediction model, the bone shape should be preserved to keep the bone microstructure after the data augmentation. Only the random rotation, random horizontal flipping, and random brightness augmentation have been performed on the input slices. The rotation (rotation degree change between -30 and 30 degrees) and brightness augmentation (pixel value change between -50 and 50) were within a limited range to keep the wrist CT structure. The 110 slices in the same wrist CT scan were augmented via the same parameter. The wrist part was put in the central area of the CT slices after data augmentation. During the testing, the scan without data augmentation was used.

Transfer Learning and Model Training

The pre-trained models from large image datasets have learned a good representation of the images and could extract discriminative image features than the randomly initialized models. Transfer learning has been proven as an efficient way to train a deep neural network model by finetuning the pre-trained models. The model could converge quickly in a new image dataset by transfer learning, which was suitable for the medical image tasks.

We employed the DenseNet121 as the backbone for the fracture prediction task. The DenseNet used the dense connection between layers and achieved better feature use efficiency with fewer parameters. The first layer has a 110-dimension input instead of three in the original DenseNet model. We add a transition layer using 1×1 convolution after the last dense block of DenseNet to reduce the feature dimension from 1024 to 256 to extract a compact feature. A multi-task learning strategy was used to achieve more robust features. The extracted feature was utilized to predict the fracture situation of the scan within five years during the follow-up period, the age of the participant at the time of the CT scanning, and the longest healthy time of the patient without any bone fracture from the scanning time. The cross-entropy loss was used as the fracture situation prediction loss; the mean squared error (MSE) loss was used as the loss for both age prediction and non-fracture year prediction. L1 regularization was performed on the weights of the classification layer to reduce overfitting. The model was optimized by ADAM optimizer using the four losses with weight one on the first three losses and 0.01 on the L1 regularization loss. The learning rate was set as 5e-6 for the DenseNet backbone and 5e-5 for the other layers, including the transition layer and the three multi-learning branches. Pytorch was utilized to implement the model.

During training, the model weights were initialized based on the model trained on the ImageNet dataset [198]. The selected wrist data as shown in Fig. 6.18 was divided into training, validation, and testing with a split ratio of 6:2:2. To reduce the data noise and achieve more stable results, we trained four models based on the different selections of the training and validation set. During the testing, the output of the bone health score was calculated based on the average output of the four models. The model was trained on the NVIDIA GeForce GTX 1080Ti GPU using the ADAM optimization method [199].

6.4 Results

6.4.1 Evaluation of the Proposed Model

The fragility fracture prediction DL model was evaluated on the testing dataset with 422 scans. We compared the result with the BMD T-score and FRAX. Since the BMD T-score and FRAX on some scans were missing, we only calculated the results of our model with the data with BMD T-score or FRAX during the comparison with BMD T-score or FRAX. Major fragility fracture was a more severe fragility fracture and would increase the death risk significantly. The major fragility fracture prediction result was also assessed. We also calculated the results of our model on population above 70 years old and above 65 years old. The population over 70 or 65 years old had more fracture risk than other groups. Both the fragility fracture and major fragility fracture were calculated.

The area under the receiver operator characteristic (ROC) curve (AUC) was used to measure the performance of the proposed model. The AUC number indicated the ability of fragility fracture prediction, and the higher the AUC, the better the model was at disguising the bone with a high chance of the occurrence of fragility fracture. The 95% confidence interval (CI) of the AUC was used to better describe the model performance. The Youden index was derived from the AUC analysis, and the sensitivity and specificity were also calculated at the Youden index. The Youden index, sensitivity, and specificity were calculated using a Python script, and the CI values were derived from the formula in [200].

6.4.2 Our Model Performance on Fragility Fracture Prediction

We have established four datasets with different compositions of selected training and validation data. The training data, validation number, and the testing number in the four datasets were as Table 6.20.

Table 6.21 showed the results on the fragility fracture prediction on all ages population. The AUC, confusion matrix, specificity and sensitivity on threshold 0.5, Youden index, and specificity and sensitivity on Youden index were calculated. The

Table 6.20 Data number in the four datasets.

Data selection set	Train set number, Total, neg : pos	Validation set number	Test set number, Total, neg : pos
1	1259, 1051 : 208	419, 350 : 69	422, 352 : 70
2	1256, 1050 : 206	422, 351 : 71	
3	1259, 1051 : 208	419, 350 : 69	
4	1260, 1051 : 209	418, 350 : 68	

proposed model achieved high AUC value of 77.48% (95% CI: 70.74% $\sim$ 84.22%). The specificity and sensitivity on the threshold of 0.5 were 76.14% and 62.86%. The Youden index was calculated from the ROC curve and was usually used for medical analysis. The Youden index on the proposed ensemble model, which used the average outputs of the model from dataset one $\sim$ four, was 0.4867, and specificity and sensitivity on the threshold of Youden index were 71.31% and 75.71%. However, the Youden indices were different among the models from dataset one $\sim$ four. Therefore, it was not an optimal choice to select the Youden index for the clinical analysis. Instead, we chose the fixed threshold of 0.5 as the clinical index for the bone health analysis in the following sections.

Table 6.22 showed the result comparison with BMD T-score. There were 352 negative cases, 70 positive cases in the test set. However, the BMD T-score of some participants was not recorded during data collection. Therefore, we only compared the model performance with participants with BMD T-score recorded. In total, there were 350 negative cases and 64 positive cases in comparison with BMD T-score. The AUC of the proposed model was 76.77%, while the AUC of BMD T-score was 53.70%. The proposed model gained an improvement of 23.07% compared with BMD T-score.

Similar to the comparison with the BMD T-score, the FRAX scores of some participants were not recorded. In total, there were 350 negative cases, 54 positive cases with FRAX score recorded in the test set. Table 6.23 showed the result comparison with FRAX. Only the data with FRAX were considered in the experiment, and the proposed model (AUC: 73.32%) outperformed the FRAX index (AUC: 60.00%) by 13.32%. Since the FRAX index used the clinical information, such as height, weight, fracture history, and medical treatment during analysis, while the proposed model only

Table 6.21 Fragility fracture prediction results on data of all ages.

Results	AUC (%) (95% CI)	Confusion matrix on clinical index 0.5	Specificity (TNR)/ Sensitivity(TPR) on clinical index of 0.5 (%)	Youden index	Specificity (TNR)/ Sensitivity(TPR) on Youden index (%)
dataset 1	72.17 (65.02 ∼79.31)	[191 161] [16 54]	54.26 / 77.14	0.519	64.49 / 72.86
dataset 2	75.24 (68.31 ∼82.17)	[296 56] [36 34]	84.09 / 48.57	0.4521	62.50 / 80.00
dataset 3	72.21 (65.07 ∼79.36)	[237 115] [23 47]	67.33 / 67.14	0.5254	71.31 / 65.71
dataset 4	71.75 (64.58 ∼78.93)	[279 73] [37 33]	79.26 / 47.14	0.4545	63.35 / 70.00
Ensemble model	77.48 (70.74 ∼84.22)	[268 84] [26 44]	76.14 / 62.86	0.4867	71.31 / 75.71

Table 6.22 Fragility fracture prediction results on data of all ages with BMD T-score recorded.

Testing number with BMD T-score recorded, total, neg : pos	Proposed model AUC (%) (95% CI)	Confusion matrix on clinical index 0.5	Specificity (TNR)/ Sensitivity(TPR) on clinical index of 0.5 (%)	BMD T-score AUC (%) (95% CI)	Confusion matrix on BMD T-score of -2.5	Specificity (TNR)/ Sensitivity(TPR) on BMD T-score of -2.5 (%)
414, 350 : 64	76.77 (69.68 ∼83.86)	[266 84] [24 40]	76.00 / 62.50	53.70 (45.89 ∼61.50)	[331 19] [58 6]	94.57 / 9.38

used the wrist CT scan only, the results revealed that the proposed model was better than the FRAX index even without the consideration of the clinical information.

Table 6.23 Fragility fracture prediction results on data of all ages with FRAX index recorded.

Testing number with FRAX recorded, total, neg : pos	Proposed model AUC (%) (95% CI)	Confusion matrix on clinical index 0.5	Specificity (TNR)/ Sensitivity(TPR) on clinical index of 0.5 (%)	FRAX AUC (%) (95% CI)	Confusion matrix on FRAX of 20	Specificity (TNR)/ Sensitivity(TPR) on FRAX of 20 (%)
404, 350 : 54	73.32 (65.35 ∼81.29)	[266 84] [23 31]	76.00 / 57.41	60.00 (51.53 ∼68.47)	[336 14] [49 5]	96.00 / 9.26

6.4.3 Results of Major Fragility Fracture Prediction on All Ages

Table 6.24 showed the result on the major fracture prediction on all ages population. The AUC of the proposed model was 74.62% on major fracture prediction of all ages. We also compared the result with the BMD (Table 6.25) or FRAX (Table 6.26) with the scans have BMD T-score or FRAX. When only the data with BMD T-score were considered, the proposed model (AUC: 74.66%) gained an improvement of 15.19% compared with BMD T-score (AUC: 59.47%). When only the data with FRAX score were considered, the proposed model (AUC: 69.08%) gained an improvement of 11.88% compared with FRAX score (AUC: 57.20%).

Table 6.24 Major fragility fracture prediction results on data of all ages.

Testing number, total, neg : pos	Proposed model AUC (%) (95% CI)	Confusion matrix on clinical index 0.5	Specificity (TNR)/ Sensitivity(TPR) on clinical index of 0.5 (%)
383, 352 : 31	74.62 (64.39 ∼84.86)	[268 84] [13 18]	76.14 / 58.06

Table 6.25 Major Fragility fracture prediction results on data of all ages with BMD T-score recorded.

Testing number with BMD recorded, total, neg : pos	Proposed model AUC (%) (95% CI)	Confusion matrix on clinical index 0.5	Specificity (TNR)/ Sensitivity(TPR) on clinical index of 0.5 (%)	BMD T-score AUC (%) (95% CI)	Confusion matrix on BMD T-score of -2.5	Specificity (TNR)/ Sensitivity(TPR) on BMD T-score of -2.5 (%)
381, 350 : 31	74.66 (64.43 ∼84.89)	[266 84] [13 18]	76.00 / 58.06	59.47 (48.55 ∼70.39)	[331 19] [27 4]	94.57 / 12.90

Table 6.26 Major fragility fracture prediction results on data of all ages with FRAX index recorded.

Testing number with FRAX recorded, total, neg : pos	Proposed model AUC (%) (95% CI)	Confusion matrix on clinical index 0.5	Specificity (TNR)/ Sensitivity(TPR) on clinical index of 0.5 (%)	FRAX AUC (%) (95% CI)	Confusion matrix on FRAX of 20	Specificity (TNR)/ Sensitivity(TPR) on FRAX of 20 (%)
374 350 : 24	69.08 (57.02 ∼81.15)	[266 84] [12 12]	76.00 / 50.00	57.20 (44.91 ∼69.49)	[336 14] [22 2]	96.00 / 8.33

6.4.4 Results of Fragility Fracture Prediction on Ages > 65

Table 6.27 showed the result on the fragility fracture prediction on ages > 65 population. The proposed model achieved an AUC of 76.65% on fragility fracture prediction on the population > 65 years old. We also compared the results with the BMD T-score (Table 6.28) or FRAX (Table 6.29) with the scans have BMD T-score / FRAX. The proposed model outperformed the BMD T-score by 22.99% and outperformed the FRAX index by 11.99%.

Table 6.27 Fragility fracture prediction results on data of ages > 65.

Testing number, total, neg : pos	Proposed model AUC (%) (95% CI)	Confusion matrix on clinical index 0.5	Specificity (TNR)/ Sensitivity(TPR) on clinical index of 0.5 (%)
236, 186 : 50	76.65 (68.46 ∼ 84.83)	[130 56] [15 35]	69.89 / 70.00

Table 6.28 Fragility fracture prediction results on data of ages > 65 with BMD T-score recorded.

Testing number with BMD recorded, total, neg : pos	Proposed model AUC (%) (95% CI)	Confusion matrix on clinical index 0.5	Specificity (TNR)/ Sensitivity(TPR) on clinical index of 0.5 (%)	BMD T-score AUC (%) (95% CI)	Confusion matrix on BMD T-score of -2.5	Specificity (TNR)/ Sensitivity(TPR) on BMD T-score of -2.5 (%)
228, 184 : 44	76.11 (67.38 ∼84.84)	[128 56] [13 31]	69.57 / 70.45	53.12 (43.51 ∼62.73)	[174 10] [39 5]	94.57 / 11.36

6.4.5 Results of Major Fragility Fracture Prediction on Ages > 65

Table 6.30 demonstrated the result on the major fracture prediction on ages > 65 population. The AUC of the proposed value was 75.05%. We also compared the result with the BMD T-score (Table 6.31) or FRAX (Table 6.32) with the scans had

Table 6.29 Fragility fracture prediction results on data of ages > 65 with FRAX index recorded.

Testing number with FRAX recorded, total, neg : pos	Proposed model AUC (%) (95% CI)	Confusion matrix on clinical index 0.5	Specificity (TNR)/ Sensitivity(TPR) on clinical index of 0.5 (%)	FRAX AUC (%) (95% CI)	Confusion matrix on FRAX of 20	Specificity (TNR)/ Sensitivity(TPR) on FRAX of 20 (%)
218, 184:34	71.10 (60.78 ∼81.42)	[128 56] [12 22]	69.57 / 64.71	59.11 (48.31 ∼69.91)	[172 12] [29 5]	93.48 / 14.71

BMD T-score or FRAX. Compared with BMD T-score, the proposed model gained an improvement of 17.46%, and compared with the FRAX index, the proposed model achieved an improvement of 8.93%.

Table 6.30 Major fragility fracture prediction results on data of ages > 65.

Testing number, total, neg : pos	Proposed model AUC (%) (95% CI)	Confusion matrix on clinical index 0.5	Specificity (TNR)/ Sensitivity(TPR) on clinical index of 0.5 (%)
208, 186 : 22	75.05 (62.86 ∼87.24)	[130 56] [7 15]	69.89 / 68.18

6.4.6 Results of Fragility Fracture Prediction on Ages > 70

Table 6.33 showed the result on the fracture prediction on ages > 70 population. The proposed model's AUC performance was 77.46% on fragility fracture prediction on population > 70 years old. We also compared the result with the BMD (Table 6.34)or FRAX (Table 6.35) with the scans contained BMD T-score / FRAX. The proposed model outperformed the BMD T-score by 20.69% and outperformed the FRAX index by 5.92%.

Table 6.31 Major fragility fracture prediction results on data of ages > 65 with BMD T-score recorded.

Testing number with BMD recorded, total, neg : pos	Proposed model AUC (%) (95% CI)	Confusion matrix on clinical index 0.5	Specificity (TNR)/ Sensitivity(TPR) on clinical index of 0.5 (%)	BMD T-score AUC (%) (95% CI)	confusion matrix on BMD T-score of -2.5	Specificity (TNR)/ Sensitivity(TPR) on BMD T-score of -2.5 (%)
206, 184 : 22	75.02 (62.83 ∼87.22)	[128 56] [7 15]	69.57 / 68.18	57.56 (44.45 ∼70.67)	[174 10] [18 4]	94.57 / 18.18

Table 6.32 Major fragility fracture prediction results on data of ages > 65 with FRAX index recorded.

Testing number with FRAX recorded, total, neg : pos	Proposed model AUC (%) (95% CI)	Confusion matrix on clinical index 0.5	Specificity (TNR)/ Sensitivity(TPR) on clinical index of 0.5 (%)	FRAX AUC (%) (95% CI)	Confusion matrix on FRAX of 20	Specificity (TNR)/ Sensitivity(TPR) on FRAX of 20 (%)
199, 184:15	67.64 (52.20 ∼83.09)	[128 56] [6 9]	69.57 / 60.00	58.71 (43.04 ∼74.38)	[172 12] [13 2]	93.48 / 13.33

Table 6.33 Fragility fracture prediction results on data of ages > 70.

Testing number, total, neg : pos	Proposed model AUC (%) (95% CI)	Confusion matrix on clinical index 0.5	Specificity (TNR)/ Sensitivity(TPR) on clinical index of 0.5 (%)
96, 71 : 25	77.46 (65.81 ∼89.12)	[46 25] [7 18]	64.79 / 72.00

Table 6.34 Fragility fracture prediction results on data of ages > 70 with BMD T-score recorded.

Testing number with BMD recorded, total, neg : pos	Proposed model AUC (%) (95% CI)	Confusion matrix on clinical index 0.5	Specificity (TNR)/ Sensitivity(TPR) on clinical index of 0.5 (%)	BMD T-score AUC (%) (95% CI)	Confusion matrix on BMD T-score of -2.5	Specificity (TNR)/ Sensitivity(TPR) on BMD T-score of -2.5 (%)
93, 70 : 23	77.83 (65.79 ∼89.86)	[45 25] [6 17]	64.29 / 73.91	57.14 (43.32 ∼70.96)	[66 4] [19 4]	94.29 / 17.39

Table 6.35 Fragility fracture prediction results on data of ages > 70 with FRAX index recorded.

Testing number with FRAX recorded, total, neg : pos	Proposed model AUC (%) (95% CI)	Confusion matrix on clinical index 0.5	Specificity (TNR)/ Sensitivity(TPR) on clinical index of 0.5 (%)	FRAX AUC (%) (95% CI)	Confusion matrix on FRAX of 20	Specificity (TNR)/ Sensitivity(TPR) on FRAX of 20 (%)
87, 70:17	73.36 (58.81 ∼87.91)	[45 25] [5 12]	64.29 / 70.59	67.44 (52.21 ∼82.67)	[65 5] [12 5]	92.86 / 29.41

6.4.7 Results of Major Fragility Fracture Prediction on Ages > 70

Table 6.36 showed the result on the major fracture prediction on ages > 70 population. The AUC of the proposed method was 77.57%. We also compared the result with the BMD T-score (Table 6.37) or FRAX (Table 6.38) with the scans had BMD T-score or FRAX index. Compared with BMD T-score, the proposed model gained an improvement of 19.56%, and compared with the FRAX index, the proposed model achieved an improvement of 10.15%.

Table 6.36 Major fragility fracture prediction results on data of ages > 70.

Testing number, total, neg : pos	Proposed model AUC (%) (95% CI)	Confusion matrix on clinical index 0.5	Specificity (TNR)/ Sensitivity(TPR) on clinical index of 0.5 (%)
84, 71 : 13	77.57 (61.93 ∼93.21)	[46 25] [4 9]	64.79 / 69.23

Table 6.37 Major fragility fracture prediction results on data of ages > 70 with BMD T-score recorded.

Testing number with BMD recorded, total, neg : pos	Proposed model AUC (%) (95% CI)	Confusion matrix on clinical index 0.5	Specificity (TNR)/ Sensitivity(TPR) on clinical index of 0.5 (%)	BMD T-score AUC (%) (95% CI)	Confusion matrix on BMD T-score of -2.5	Specificity (TNR)/ Sensitivity(TPR) on BMD T-score of -2.5 (%)
83, 70:13	77.47 (61.80 ∼93.14)	[45 25] [4 9]	64.29 / 69.23	57.91 (40.38 ∼75.44)	[66 4] [10 3]	94.29 / 23.08

Table 6.38 Major fragility fracture prediction results on data of ages > 70 with FRAX index recorded.

Testing number with FRAX recorded, total, neg : pos	Proposed model AUC (%) (95% CI)	Confusion matrix on clinical index 0.5	Specificity (TNR)/ Sensitivity(TPR) on clinical index of 0.5 (%)	FRAX AUC (%) (95% CI)	Confusion matrix on FRAX of 20	Specificity (TNR)/ Sensitivity(TPR) on FRAX of 20 (%)
79, 70 : 9	73.17 (53.65 ∼92.70)	[45 25] [3 6]	64.29 / 66.67	63.02 (42.43 ∼83.61)	[65 5] [7 2]	92.86 / 22.22

6.4.8 Results Comparison of Models using Different Wrist Parts as Input on Fragility Fracture Prediction on All Ages

We also trained the fragility fracture prediction model using the radius part, ulna part, and muscle part as input, respectively. The outputs of the model via radius, model via ulna, and model via muscle were average as another ensemble model. We compared the results on fragility fracture prediction of the model via radius, model via ulna, model via muscle, model by averaging the output of the above three models, and the proposed model. Table 6.39 listed the results, and the results revealed that the proposed model achieved the best AUC compared with the other four models. This indicated that the radius, ulna, and muscle were both important for predicting bone fracture.

6.4.9 Results of Prediction of Age and Longest Health Year before the Bone Fragility Fracture and Results using other Deep Learning Models

The proposed model was also used to predict age and the participant's longest health year before the bone fragility fracture. The mean absolute error of age prediction was 4.71, and the mean absolute error of the predicted longest health year before the bone fragility fracture was 4.22. The doctors could use these information during the

Table 6.39 Comparison of models using different CT parts as input on fragility fracture prediction on all ages.

	Model AUC (%) (95% CI)	Confusion matrix on clinical index 0.5	Specificity (TNR)/ Sensitivity(TPR) on clinical index of 0.5 (%)
Radius	72.60 (65.48 ∼79.72)	[215 137] [19 51]	61.08 / 72.86
Ulna	68.39 (61.04 ∼75.74)	[240 112] [32 38]	68.18 / 54.29
Muscle	69.28 (61.97 ∼76.58)	[247 105] [30 40]	70.17 / 57.14
Average of output of Radius, Ulna, and Muscle model	77.05 (70.28 ∼83.83)	[260 92] [23 47]	73.86 / 67.14
Proposed model	77.48 (70.74 ∼84.22)	[268 84] [26 44]	76.14 / 62.86

diagnosis. We also compared the performance between different deep learning models, such as DenseNet, ResNet, VGGNet, and GoogLeNet, on fragility fracture prediction task during the model backbone selection. Since each scan contained 110 slices and the data size was more than 500 MB, reading data from disk consumed a lot of time during training. We used the average axil-view value of the 110 slices as input to save training time during model selection. Table 6.40 listed the results of the above models. The results indicated that DenseNet could use the model parameters more efficiently for bone fracture prediction. Besides, compared with the results of DenseNet, the proposed model using multi-task training strategy achieved a better result.

6.4.10 Visualization of Region of Interest for Bone Fracture Prediction using Heatmap

Understanding where was considered as the most important part for predicting the bone fracture by the neural network was crucial for the model interpretation. The Grad-CAM [159] was utilized to generate the heatmap of the model and illustrate the learning behavior of the model. The CT scan data was fed into the proposed deep

Table 6.40 Comparison of different deep learning models for fragility fracture prediction task.

	Model AUC (%) (95% CI)		Model AUC (%) (95% CI)
VGGNet	72.40 (65.27 ∼79.54)	DenseNet	75.53 (68.62 ∼82.44)
GoogLeNet	73.99 (66.96 ∼81.01)	Proposed model using average slice	76.25 (69.41 ∼83.10)
ResNet	75.19 (68.26 ∼82.13)	Proposed model using 110 slices	77.48 (70.74 ∼84.22)

learning model, and a prediction score was generated after the forward propagation. The gradients of the specified convolutional layer of the predicted class were calculated through the backpropagation and were then pooled channel-wise. The feature maps of the targeted convolutional layer were weighted with the corresponding gradients to yield the heatmap. The background of the CT data was removed, and the prediction score was weighted to the heatmap to illustrate the decision more clearly. The heatmap was then overlapped on the CT scan data to interpret the decision of the model. The brighter the color in the generation map, the greater the possibility of bone fragility fracture in the next five years. The results were shown in Fig. 6.5 and Fig. 6.6, and the model put more weight on the radius part when predicting the bone fracture situation.

6.5 Conclusion

This study provided data supporting the hypothesis that the deep learning method had the ability to predict the fragility fracture. We selected data from three population-based cohorts and constructed a structured wrist database from the unstructured clinical data. The wrist scans were used to train and test the proposed model performance. The results demonstrated the high performance of the deep learning model in the fragility fracture prediction and major fragility fracture prediction within five years in different age groups. Moreover, the results showed superior performance in fragility fracture and major fragility fracture prediction of the deep learning model, compared with the BMD and FRAX scores that have been widely used in assessing bone health.

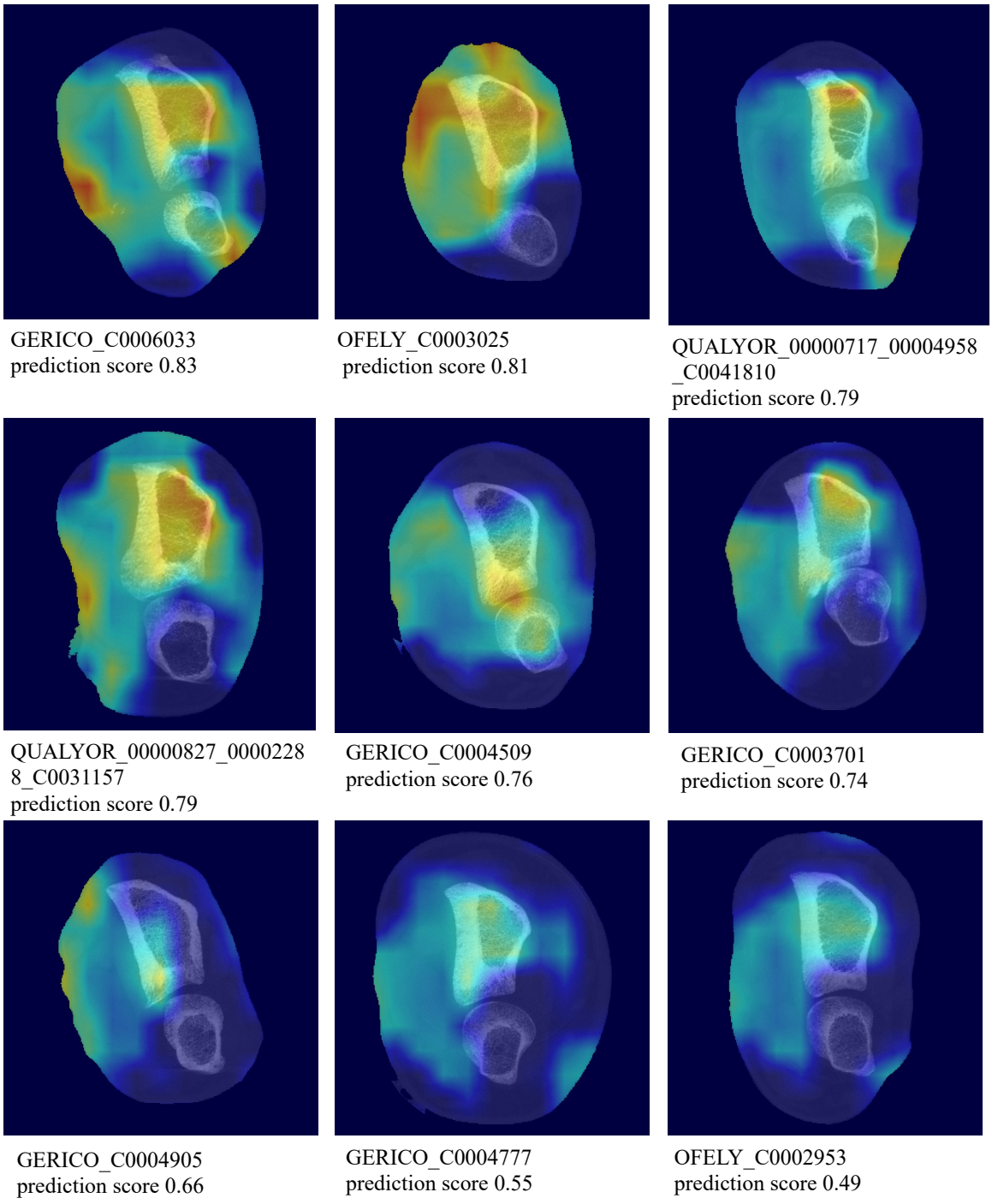

Fig. 6.5 Examples of the heatmaps (part one).

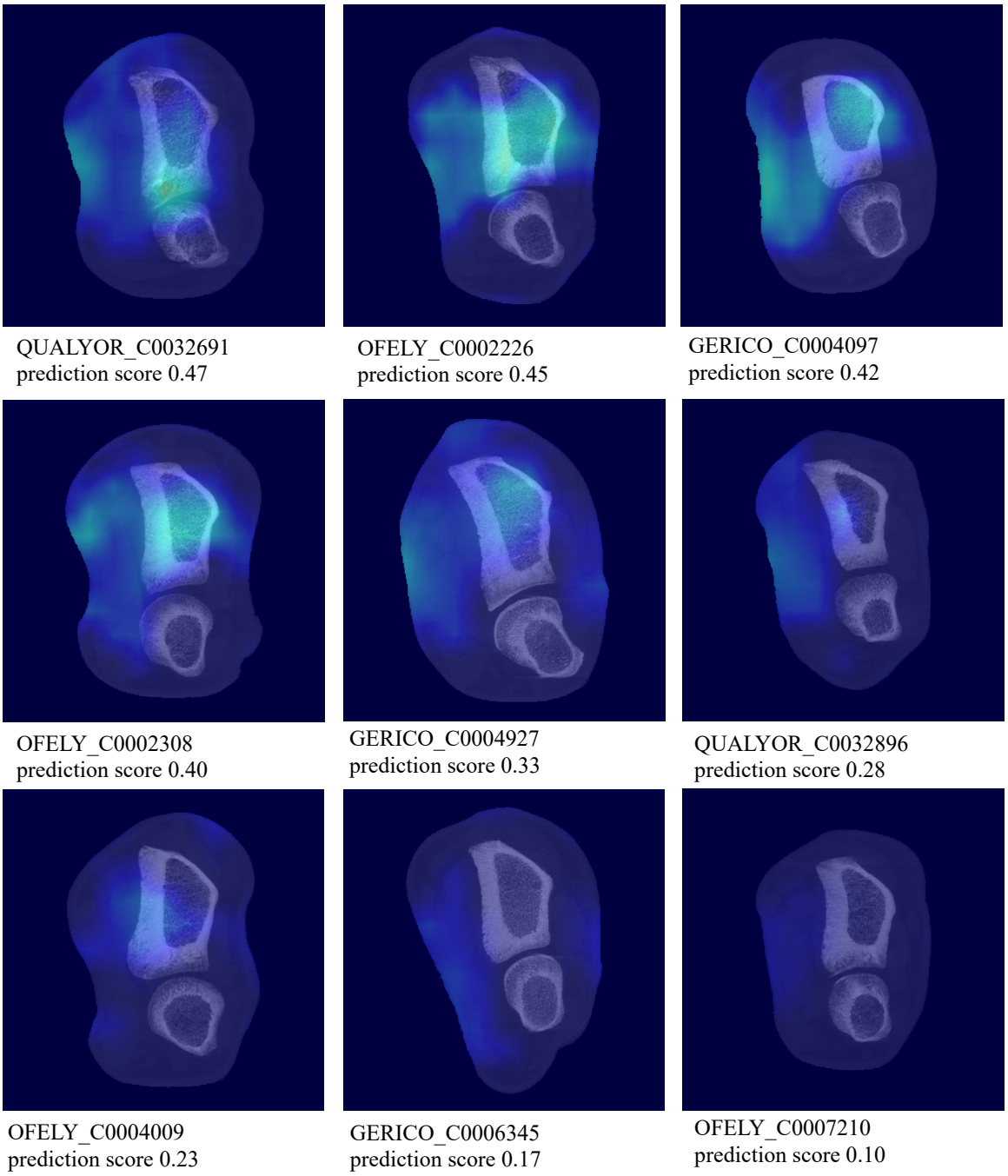

Fig. 6.6 Examples of the heatmaps (part two).

Chapter 7

Conclusion and Future Work

7.1 Conclusion

As a great medical invention during the past decades, CT has greatly facilitated the clinical assessment in orthopedics. This thesis focused on the automatic analysis of bone CT via deep learning methods to assist the medical diagnosis. Bone segmentation is the fundamental stage to various medical applications, and bone health analysis is an important part of clinical practice. We conducted several works in the two areas, including the anatomical segmentation of human foot weight-bearing CBCT scan, instance segmentation of wrist CT, semi-supervised bone CT segmentation, and bone health analysis via predicting the fragility fracture at intermediate risk (5-years) using wrist CT.

Firstly, we focused on the anatomical segmentation of human foot CT scan in chapter 3. Thirty-one foot bones have been identified using a three-stage framework, preprocessing, bone region segmentation, and bone pixel classification. The proposed method could automatically segment the bones from the Dicom CT data, and the proposed method generated accurate results on foot anatomical segmentation.

Afterward, we worked on another segmentation task, i.e., instance segmentation of wrist CT in chapter 4. In order to reduce the manual annotation workload, a semi-automatic method that annotated 5k wrist CT slices by employing the Otsu-based method and the U-net-based method was proposed. Our method only required

fewer manual annotations, saved much time, and alleviated the annotation workload significantly. We also proposed an edge-enhanced segmentation model for the instance segmentation of wrist CT slice. The proposed model achieved better performance compared with the U-net model. The training procedure was more stable and was not vulnerable to over-fitting.

During the studies on foot and wrist CT segmentation, we found manual annotation was time-consuming and laborious. Therefore, we considered employing the semi-supervised method to develop a bone segmentation model in chapter 5. We leveraged the fact that bone had a specific Hounsfield unit compared with the other human parts and proposed a novel patch-shuffle-based semi-supervised segmentation method for bone CT segmentation. The proposed method has been evaluated on three different datasets. The experiment results demonstrated the high effectiveness of the proposed method on bone CT segmentation.

Finally, we developed a deep learning method to determine bone health by predicting the fragility fracture at intermediate risk (five-years). We collected data from three population-based cohorts, and the participants in the three cohorts have been followed for 6.16 years on average. We constructed a structured wrist database from the unstructured clinical raw data. The proposed model achieved high AUC values on predicting the fragility fracture and major fragility fracture in different age groups. What's more, the proposed method demonstrated superior performance than the widely used BMD T-score and FRAX index.

In conclusion, we have developed various deep learning-based methods for bone CT analysis, including different segmentation and classification tasks. The multi-stage based methods were widely used in the thesis and achieved remarkable results. For example, we developed a three-stage based framework for the anatomical segmentation of foot CT in chapter 3, and a two-branch based method for the annotation of wrist CT in chapter 4. The multi-stage based framework have greatly reduced the difficulty of the deep learning model design. The semi-supervised learning has greatly facilitated the deep learning model development. Both the self-learning based semi-supervised segmentation method in chapter 4 and the consistency-based semi-supervised segmentation method in chapter 5 have achieved ideal segmentation results and demonstrated the potential to reduce

the annotation workload for segmentation tasks. Besides, the bone segmentation is the fundamental stage to various medical applications. We have used the segmentation results in chapter 4 as the input for the classification model in chapter 6, which could help to extract more efficient features by removing irrelevant background areas.

7.2 Future Work

This thesis focused on developing deep learning approaches to analyze bone CT. The conducted works have demonstrated that deep learning methods could greatly assist bone CT analysis in clinical treatment. To further design more advanced methods in this field, the interesting directions for future work are presented as follows.

- The current foot anatomical segmentation model is not an end-to-end framework, and the time to process one scan would use fifteen minutes. The future work will aim to propose an end-to-end framework to simplify the framework and accelerate the processing speed. We will also consider to introduce the long-tail distribution based classification or segmentation methods to replace the data sampling methods in the classification model in future work.
- We only consider two bones in the wrist CT data. In the future, it is interesting to consider more bones in the wrist data, use the semantic edge information to improve the model performance further, and accelerate the annotation time through semi-supervised learning methods.
- We used the consistency learning method for the semi-supervised task. In the future, we could extend our approach to more segmentation tasks and investigate introducing the pseudo labels to our work.
- The fracture prediction model could identify people at intermediate risk of bone fracture. In the future, it is valuable to extend the work on imminent risks, such as one-year or two-year, by exploring a larger population, collecting more data, and designing 3D models for the fracture prediction on imminent risk.

References

[1] R. Ramdhian-Wihlm, J.-M. Le Minor, M. Schmittbuhl, J. Jeantroux, P. Mac Mahon, F. Veillon, J.-C. Dosch, J.-L. Dietemann, and G. Bierry, "Cone-beam computed tomography arthrography: an innovative modality for the evaluation of wrist ligament and cartilage injuries," *Skeletal radiology*, vol. 41, no. 8, pp. 963–969, 2012.

[2] C. de Cesar Netto, A. Bernasconi, L. Roberts, P. A. Pontin, F. Lintz, G. H. Saito, A. Roney, A. Elliott, and M. O'Malley, "Foot alignment in symptomatic national basketball association players using weightbearing cone beam computed tomography," *Orthopaedic journal of sports medicine*, vol. 7, no. 2, p. 2325967119826081, 2019.

[3] S. W. Chung, S. S. Han, J. W. Lee, K.-S. Oh, N. R. Kim, J. P. Yoon, J. Y. Kim, S. H. Moon, J. Kwon, H.-J. Lee, *et al.*, "Automated detection and classification of the proximal humerus fracture by using deep learning algorithm," *Acta orthopaedica*, vol. 89, no. 4, pp. 468–473, 2018.

[4] M. A. Badgeley, J. R. Zech, L. Oakden-Rayner, B. S. Glicksberg, M. Liu, W. Gale, M. V. McConnell, B. Percha, T. M. Snyder, and J. T. Dudley, "Deep learning predicts hip fracture using confounding patient and healthcare variables," *NPJ digital medicine*, vol. 2, no. 1, pp. 1–10, 2019.

[5] C. C. Johnson, E. B. Gausden, A. J. Weiland, J. M. Lane, and J. J. Schreiber, "Using hounsfield units to assess osteoporotic status on wrist computed tomography scans: comparison with dual energy x-ray absorptiometry," *The Journal of hand surgery*, vol. 41, no. 7, pp. 767–774, 2016.

[6] R. Chapurlat, M. Bui, E. Sornay-Rendu, R. Zebaze, P. D. Delmas, D. Liew, E. Lespessailles, and E. Seeman, "Deterioration of cortical and trabecular microstructure identifies women with osteopenia or normal bone mineral density at imminent and long-term risk for fragility fracture: A prospective study," *Journal of Bone and Mineral Research*, vol. 35, no. 5, pp. 833–844, 2020.

[7] T. Klinder, J. Ostermann, M. Ehm, A. Franz, R. Kneser, and C. Lorenz, "Automated model-based vertebra detection, identification, and segmentation in ct images," *Medical image analysis*, vol. 13, no. 3, pp. 471–482, 2009.

[8] H. Scherf and R. Tilgner, "A new high-resolution computed tomography (ct) segmentation method for trabecular bone architectural analysis," *American*

Journal of Physical Anthropology: The Official Publication of the American Association of Physical Anthropologists, vol. 140, no. 1, pp. 39–51, 2009.

[9] J. Zhang, C.-H. Yan, C.-K. Chui, and S.-H. Ong, "Fast segmentation of bone in ct images using 3d adaptive thresholding," *Computers in biology and medicine*, vol. 40, no. 2, pp. 231–236, 2010.

[10] H. M. Kang, M. G. Kim, S. H. Boo, K. H. Kim, E. K. Yeo, S. K. Lee, and S. G. Yeo, "Comparison of the clinical relevance of traditional and new classification systems of temporal bone fractures," *European Archives of Oto-Rhino-Laryngology*, vol. 269, no. 8, pp. 1893–1899, 2012.

[11] J. Dunklebarger, B. Branstetter IV, A. Lincoln, M. Sippey, M. Cohen, B. Gaines, and D. Chi, "Pediatric temporal bone fractures: current trends and comparison of classification schemes," *The Laryngoscope*, vol. 124, no. 3, pp. 781–784, 2014.

[12] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio, *Deep learning*, vol. 1. MIT press Cambridge, 2016.

[13] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, 2017.

[14] P. Druzhkov and V. Kustikova, "A survey of deep learning methods and software tools for image classification and object detection," *Pattern Recognition and Image Analysis*, vol. 26, no. 1, pp. 9–15, 2016.

[15] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. Van Der Laak, B. Van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Medical image analysis*, vol. 42, pp. 60–88, 2017.

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

[17] A. A. Novikov, D. Major, M. Wimmer, D. Lenis, and K. Bühler, "Deep sequential segmentation of organs in volumetric medical scans," *IEEE transactions on medical imaging*, vol. 38, no. 5, pp. 1207–1215, 2018.

[18] G. Fan, H. Liu, Z. Wu, Y. Li, C. Feng, D. Wang, J. Luo, W. Wells, and S. He, "Deep learning–based automatic segmentation of lumbosacral nerves on ct for spinal intervention: a translational study," *American Journal of Neuroradiology*, vol. 40, no. 6, pp. 1074–1081, 2019.

[19] T. J. Netherton, D. J. Rhee, C. E. Cardenas, C. Chung, A. H. Klopp, C. B. Peterson, R. M. Howell, P. A. Balter, and L. E. Court, "Evaluation of a multiview architecture for automatic vertebral labeling of palliative radiotherapy simulation ct images," *Medical physics*, vol. 47, no. 11, p. 5592, 2020.

[20] G. Fan, H. Liu, D. Wang, C. Feng, Y. Li, B. Yin, Z. Zhou, X. Gu, H. Zhang, Y. Lu, *et al.*, "Deep learning-based lumbosacral reconstruction for difficulty prediction of percutaneous endoscopic transforaminal discectomy at l5/s1 level: A retrospective cohort study," *International Journal of Surgery*, vol. 82, pp. 162–169, 2020.

[21] V. Malinda and D. Lee, "Lumbar vertebrae synthetic segmentation in computed tomography images using hybrid deep generative adversarial networks," in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pp. 1327–1330, IEEE, 2020.

[22] A. Zakirov, M. Ezhov, M. Gusarev, V. Alexandrovsky, and E. Shumilov, "Dental pathology detection in 3d cone-beam ct," *arXiv preprint arXiv:1810.10309*, 2018.

[23] L. You, G. Zhang, W. Zhao, M. Greives, L. David, and X. Zhou, "Automated sagittal craniosynostosis classification from ct images using transfer learning," *Clinics in surgery*, vol. 5, 2020.

[24] I. Kim, D. Misra, L. Rodriguez, M. Gill, D. K. Liberton, K. Almpani, J. S. Lee, and S. Antani, "Malocclusion classification on 3d cone-beam ct craniofacial images using multi-channel deep learning models," in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pp. 1294–1298, IEEE, 2020.

[25] N. Papandrianos, E. Papageorgiou, A. Anagnostis, and K. Papageorgiou, "Bone metastasis classification using whole body images from prostate cancer patients based on convolutional neural networks application," *PloS one*, vol. 15, no. 8, p. e0237213, 2020.

[26] Y. Wu, X. Lu, J. Hong, W. Lin, S. Chen, S. Mou, G. Feng, R. Yan, and Z. Cheng, "Detection of extremity chronic traumatic osteomyelitis by machine learning based on computed-tomography images: A retrospective study," *Medicine*, vol. 99, no. 9, 2020.

[27] Q. Lin, T. Li, C. Cao, Y. Cao, Z. Man, and H. Wang, "Deep learning based automated diagnosis of bone metastases with spect thoracic bone images," *Scientific Reports*, vol. 11, no. 1, pp. 1–15, 2021.

[28] A. L. Godoy-Santos, A. Bernasconi, M. Bordalo-Rodrigues, F. Lintz, C. F. T. Lôbo, and C. d. C. Netto, "Weight-bearing cone-beam computed tomography in the foot and ankle specialty: where we are and where we are going-an update," *Radiologia Brasileira*, vol. 54, pp. 177–184, 2021.

[29] F. Lintz, P. Beaudet, G. Richardi, and J. Brilhault, "Weight-bearing ct in foot and ankle pathology," *Orthopaedics & Traumatology: Surgery & Research*, vol. 107, no. 1, p. 102772, 2021.

[30] M. S. Conti and S. J. Ellis, "Weight-bearing ct scans in foot and ankle surgery," *JAAOS-Journal of the American Academy of Orthopaedic Surgeons*, vol. 28, no. 14, pp. e595–e603, 2020.

[31] J. R. Jastifer and P. A. Gustafson, "Three-dimensional printing and surgical simulation for preoperative planning of deformity correction in foot and ankle surgery," *The Journal of Foot and Ankle Surgery*, vol. 56, no. 1, pp. 191–195, 2017.

[32] L. Jans, I. De Kock, N. Herregods, K. Verstraete, F. Van den Bosch, P. Carron, E. H. Oei, D. Elewaut, and P. Jacques, "Dual-energy ct: a new imaging modality for bone marrow oedema in rheumatoid arthritis," *Annals of the rheumatic diseases*, vol. 77, no. 6, pp. 958–960, 2018.

[33] J. S. You, S. P. Chung, H. S. Chung, I. C. Park, H. S. Lee, and S. H. Kim, "The usefulness of ct for patients with carpal bone fractures in the emergency department," *Emergency Medicine Journal*, vol. 24, no. 4, pp. 248–250, 2007.

[34] J. G. Snel, H. W. Venema, and C. A. Grimbergen, "Deformable triangular surfaces using fast 1-d radial lagrangian dynamics-segmentation of 3-d mr and ct images of the wrist," *IEEE transactions on medical imaging*, vol. 21, no. 8, pp. 888–903, 2002.

[35] J. Duryea, M. Magalnick, S. Alli, L. Yao, M. Wilson, and R. Goldbach-Mansky, "Semiautomated three-dimensional segmentation software to quantify carpal bone volume changes on wrist ct scans for arthritis assessment," *Medical physics*, vol. 35, no. 6Part1, pp. 2321–2330, 2008.

[36] F. Salaffi, M. Carotti, A. Ciapetti, A. Ariani, S. Gasparini, and W. Grassi, "Validity of a computer-assisted manual segmentation software to quantify wrist erosion volume using computed tomography scans in rheumatoid arthritis," *BMC musculoskeletal disorders*, vol. 14, no. 1, pp. 1–8, 2013.

[37] E. M. A. Anas, A. Rasoulian, P. S. John, D. Pichora, R. Rohling, and P. Abolmaesumi, "A statistical shape+ pose model for segmentation of wrist ct images," in *Medical Imaging 2014: Image Processing*, vol. 9034, p. 90340T, International Society for Optics and Photonics, 2014.

[38] Y. Gordienko, P. Gang, J. Hui, W. Zeng, Y. Kochura, O. Alienin, O. Rokovyi, and S. Stirenko, "Deep learning with lung segmentation and bone shadow exclusion techniques for chest x-ray analysis of lung cancer," in *International Conference on Computer Science, Engineering and Education Applications*, pp. 638–647, Springer, 2018.

[39] Y. Huo, Z. Xu, K. Aboud, P. Parvathaneni, S. Bao, C. Bermudez, S. M. Resnick, L. E. Cutting, and B. A. Landman, "Spatially localized atlas network tiles enables 3d whole brain segmentation from limited data," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 698–705, Springer, 2018.

[40] P. Liu, H. Han, Y. Du, H. Zhu, Y. Li, F. Gu, H. Xiao, J. Li, C. Zhao, L. Xiao, *et al.*, "Deep learning to segment pelvic bones: large-scale ct datasets and baseline models," *International Journal of Computer Assisted Radiology and Surgery*, vol. 16, no. 5, pp. 749–756, 2021.

[41] X. Li, Y. Peng, and M. Xu, "Edge-enhanced instance segmentation of wrist ct via a semi-automatic annotation database construction method," in *2021 Digital Image Computing: Techniques and Applications (DICTA)*, pp. 01–08, IEEE.

[42] Y. Deng, C. Wang, Y. Hui, Q. Li, J. Li, S. Luo, M. Sun, Q. Quan, S. Yang, Y. Hao, *et al.*, "Ctspine1k: A large-scale dataset for spinal vertebrae segmentation in computed tomography," *arXiv preprint arXiv:2105.14711*, 2021.

[43] J. T. Schousboe, "Mortality after osteoporotic fractures: what proportion is caused by fracture and is preventable?," *Journal of Bone and Mineral Research*, vol. 32, no. 9, pp. 1783–1788, 2017.

[44] T. Tran, D. Bliuc, L. Hansen, B. Abrahamsen, J. van den Bergh, J. A. Eisman, T. van Geel, P. Geusens, P. Vestergaard, T. V. Nguyen, *et al.*, "Persistence of excess mortality following individual nonhip fractures: a relative survival analysis," *The Journal of Clinical Endocrinology & Metabolism*, vol. 103, no. 9, pp. 3205–3214, 2018.

[45] J. A. Kanis, "Assessment of fracture risk and its application to screening for postmenopausal osteoporosis: synopsis of a who report," *Osteoporosis international*, vol. 4, no. 6, pp. 368–381, 1994.

[46] J. Kanis, O. Johnell, A. Odén, H. Johansson, and E. McCloskey, "Frax™ and the assessment of fracture probability in men and women from the uk," *Osteoporosis international*, vol. 19, no. 4, pp. 385–397, 2008.

[47] Y. Miki, C. Muramatsu, T. Hayashi, X. Zhou, T. Hara, A. Katsumata, and H. Fujita, "Classification of teeth in cone-beam ct using deep convolutional neural network," *Computers in biology and medicine*, vol. 80, pp. 24–29, 2017.

[48] X. H. Meng, D. J. Wu, Z. Wang, X. L. Ma, X. M. Dong, A. E. Liu, and L. Chen, "A fully automated rib fracture detection system on chest ct images and its impact on radiologist performance," *Skeletal Radiology*, vol. 50, no. 9, pp. 1821–1828, 2021.

[49] Y. D. Pranata, K.-C. Wang, J.-C. Wang, I. Idram, J.-Y. Lai, J.-W. Liu, and I.-H. Hsieh, "Deep learning and surf for automated classification and detection of calcaneus fractures in ct images," *Computer methods and programs in biomedicine*, vol. 171, pp. 27–37, 2019.

[50] D. Chettrit, T. Meir, H. Lebel, M. Orlovsky, R. Gordon, A. Akselrod-Ballin, and A. Bar, "3d convolutional sequence to sequence model for vertebral compression fractures identification in ct," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 743–752, Springer, 2020.

[51] M. Husseini, A. Sekuboyina, M. Loeffler, F. Navarro, B. H. Menze, and J. S. Kirschke, "Grading loss: a fracture grade-based metric loss for vertebral fracture detection," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 733–742, Springer, 2020.

[52] K. M. Lee, S. Y. Lee, C. S. Han, and S. M. Choi, "Long bone fracture type classification for limited number of ct data with deep learning," in *Proceedings of the 35th Annual ACM Symposium on Applied Computing*, pp. 1090–1095, 2020.

[53] Y. Li, Y. Zhang, E. Zhang, Y. Chen, Q. Wang, K. Liu, J. Y. Hon, H. Yuan, N. Lang, and M.-Y. Su, "Differential diagnosis of benign and malignant vertebral fracture on ct using deep learning," *European Radiology*, pp. 1–8, 2021.

[54] N. A. Farda, J.-Y. Lai, J.-C. Wang, P.-Y. Lee, J.-W. Liu, and I.-H. Hsieh, "Sanders classification of calcaneal fractures in ct images with deep learning and differential data augmentation techniques," *Injury*, vol. 52, no. 3, pp. 616–624, 2021.

[55] J. Bewes, A. Low, A. Morphett, F. D. Pate, and M. Henneberg, "Artificial intelligence for sex determination of skeletal remains: application of a deep learning artificial neural network to human skulls," *Journal of Forensic and Legal Medicine*, vol. 62, pp. 40–43, 2019.

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

[57] S. Belharbi, C. Chatelain, R. Hérault, S. Adam, S. Thureau, M. Chastan, and R. Modzelewski, "Spotting l3 slice in ct scans using deep convolutional network and transfer learning," *Computers in biology and medicine*, vol. 87, pp. 95–103, 2017.

[58] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[59] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1–9, 2015.

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

[61] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4700–4708, 2017.

[62] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2818–2826, 2016.

[63] I. Kim, S. Rajaraman, and S. Antani, "Visual interpretation of convolutional neural network predictions in classifying medical image modalities," *Diagnostics*, vol. 9, no. 2, p. 38, 2019.

[64] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "Mobilenets: Efficient convolutional neural networks for mobile vision applications," *arXiv preprint arXiv:1704.04861*, 2017.

[65] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1251–1258, 2017.

[66] A. Klein, J. Warszawski, J. Hillengaß, and K. H. Maier-Hein, "Automatic bone segmentation in whole-body ct images," *International journal of computer assisted radiology and surgery*, vol. 14, no. 1, pp. 21–29, 2019.

[67] S. Noguchi, M. Nishio, M. Yakami, K. Nakagomi, and K. Togashi, "Bone segmentation on whole-body ct using convolutional neural network with novel data augmentation techniques," *Computers in biology and medicine*, vol. 121, p. 103767, 2020.

[68] J. Egger, B. Pfarrkirchner, C. Gsaxner, L. Lindner, D. Schmalstieg, and J. Wallner, "Fully convolutional mandible segmentation on a valid ground-truth dataset," in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 656–660, IEEE, 2018.

[69] A. Huang, W.-H. Cheng, C.-W. Lee, C.-Y. Yang, and H.-M. Liu, "Multiphase computed tomographic angiography with bone subtraction using 3d multichannel convolution neural networks," in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pp. 1274–1277, IEEE, 2020.

[70] F. Matzkin, V. Newcombe, S. Stevenson, A. Khetani, T. Newman, R. Digby, A. Stevens, B. Glocker, and E. Ferrante, "Self-supervised skull reconstruction in brain ct images with decompressive craniectomy," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 390–399, Springer, 2020.

[71] J. Zhang, M. Liu, L. Wang, S. Chen, P. Yuan, J. Li, S. G.-F. Shen, Z. Tang, K.-C. Chen, J. J. Xia, *et al.*, "Joint craniomaxillofacial bone segmentation and landmark digitization by context-guided fully convolutional networks," in *International conference on medical image computing and computer-assisted intervention*, pp. 720–728, Springer, 2017.

[72] C. Lian, F. Wang, H. H. Deng, L. Wang, D. Xiao, T. Kuang, H.-Y. Lin, J. Gateno, S. G. Shen, P.-T. Yap, *et al.*, "Multi-task dynamic transformer network for concurrent bone segmentation and large-scale landmark localization with dental cbct," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 807–816, Springer, 2020.

[73] M. J. Lee, H. Hong, K. W. Shim, and S. Park, "Mgb-net: Orbital bone segmentation from head and neck ct images using multi-graylevel-bone convolutional networks," in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, pp. 692–695, IEEE, 2019.

[74] N. Torosdagli, D. K. Liberton, P. Verma, M. Sincan, J. S. Lee, and U. Bagci, "Deep geodesic learning for segmentation and anatomical landmarking," *IEEE transactions on medical imaging*, vol. 38, no. 4, pp. 919–931, 2018.

[75] J. Fauser, I. Stenin, M. Bauer, W.-H. Hsu, J. Kristin, T. Klenzner, J. Schipper, and A. Mukhopadhyay, "Toward an automatic preoperative pipeline for image-guided temporal bone surgery," *International journal of computer assisted radiology and surgery*, vol. 14, no. 6, pp. 967–976, 2019.

[76] X. Li, Z. Gong, H. Yin, H. Zhang, Z. Wang, and L. Zhuo, "A 3d deep supervised densely network for small organs of human temporal bone segmentation in ct images," *Neural Networks*, vol. 124, pp. 75–85, 2020.

[77] C. Neves, E. Tran, I. Kessler, and N. Blevins, "Fully automated preoperative segmentation of temporal bone structures from clinical ct scans," *Scientific Reports*, vol. 11, no. 1, pp. 1–11, 2021.

[78] S. Nikan, K. Van Osch, M. Bartling, D. G. Allen, S. A. Rohani, B. Connors, S. K. Agrawal, and H. M. Ladak, "Pwd-3dnet: A deep learning-based fully-automated segmentation of multiple structures on temporal bone ct scans," *IEEE Transactions on Image Processing*, vol. 30, pp. 739–753, 2020.

[79] J. Wang, Y. Lv, J. Wang, F. Ma, Y. Du, X. Fan, M. Wang, and J. Ke, "Fully automated segmentation in temporal bone ct with neural network: a preliminary assessment study," *BMC Medical Imaging*, vol. 21, no. 1, pp. 1–11, 2021.

[80] S. Nikan, S. K. Agrawal, and H. M. Ladak, "Fully automated segmentation of the temporal bone from micro-ct using deep learning," in *Medical Imaging 2020: Biomedical Applications in Molecular, Structural, and Functional Imaging*, vol. 11317, p. 113171U, International Society for Optics and Photonics, 2020.

[81] Q. Li, K. Chen, L. Han, Y. Zhuang, J. Li, and J. Lin, "Automatic tooth roots segmentation of cone beam computed tomography image sequences using u-net and rnn," *Journal of X-ray Science and Technology*, vol. 28, no. 5, pp. 905–922, 2020.

[82] J. Jaskari, J. Sahlsten, J. Järnstedt, H. Mehtonen, K. Karhu, O. Sundqvist, A. Hietanen, V. Varjonen, V. Mattila, and K. Kaski, "Deep learning method for mandibular canal segmentation in dental cone beam computed tomography volumes," *Scientific reports*, vol. 10, no. 1, pp. 1–8, 2020.

[83] F. C. Setzer, K. J. Shi, Z. Zhang, H. Yan, H. Yoon, M. Mupparapu, and J. Li, "Artificial intelligence for the computer-aided detection of periapical lesions in cone-beam computed tomographic images," *Journal of endodontics*, vol. 46, no. 7, pp. 987–993, 2020.

[84] Z. Cui, C. Li, and W. Wang, "Toothnet: automatic tooth instance segmentation and identification from cone beam ct images," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 6368–6377, 2019.

[85] M. Ezhov, A. Zakirov, and M. Gusarev, "Coarse-to-fine volumetric segmentation of teeth in cone-beam ct," in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, pp. 52–56, IEEE, 2019.

[86] E. Taghizadeh, O. Truffer, F. Becce, S. Eminian, S. Gidoin, A. Terrier, A. Farron, and P. Büchler, "Deep learning for the rapid automatic quantification and characterization of rotator cuff muscle degeneration from shoulder ct datasets," *European radiology*, vol. 31, no. 1, pp. 181–190, 2021.

[87] P. Sanghani, F. Wong, and H. Ren, "Clavicle bone segmentation from ct images using u-net-based deep learning algorithm," in *Data Analytics in Biomedical Engineering and Healthcare*, pp. 205–214, Elsevier, 2021.

[88] M. T. Löffler, A. Jacob, A. Scharr, N. Sollmann, E. Burian, M. El Husseini, A. Sekuboyina, G. Tetteh, C. Zimmer, J. Gempt, *et al.*, "Automatic opportunistic osteoporosis screening in routine ct: improved prediction of patients with prevalent vertebral fractures compared to dxa," *European Radiology*, pp. 1–9, 2021.

[89] A. Suri, B. C. Jones, G. Ng, N. Anabaraonye, P. Beyrer, A. Domi, G. Choi, S. Tang, A. Terry, T. Leichner, *et al.*, "A deep learning system for automated, multi-modality 2d segmentation of vertebral bodies and intervertebral discs," *Bone*, vol. 149, p. 115972, 2021.

[90] H.-J. Bae, H. Hyun, Y. Byeon, K. Shin, Y. Cho, Y. J. Song, S. Yi, S.-U. Kuh, J. S. Yeom, and N. Kim, "Fully automated 3d segmentation and separation of multiple cervical vertebrae in ct images using a 2d convolutional neural network," *Computer methods and programs in biomedicine*, vol. 184, p. 105119, 2020.

[91] Y. Pan, D. Shi, H. Wang, T. Chen, D. Cui, X. Cheng, and Y. Lu, "Automatic opportunistic osteoporosis screening using low-dose chest computed tomography scans obtained for lung cancer screening," *European radiology*, vol. 30, no. 7, p. 4107, 2020.

[92] S. L. Belal, M. Sadik, R. Kaboteh, O. Enqvist, J. Ulén, M. H. Poulsen, J. Simonsen, P. F. Høilund-Carlsen, L. Edenbrandt, and E. Trägårdh, "Deep learning for segmentation of 49 selected bones in ct scans: first step in automated pet/ct-based 3d quantification of skeletal metastases," *European journal of radiology*, vol. 113, pp. 89–95, 2019.

[93] F. Rehman, S. I. A. Shah, M. N. Riaz, S. O. Gilani, and R. Faiza, "A region-based deep level set formulation for vertebral bone segmentation of osteoporotic fractures," *Journal of digital imaging*, vol. 33, no. 1, pp. 191–203, 2020.

[94] Y. Fang, W. Li, X. Chen, K. Chen, H. Kang, P. Yu, R. Zhang, J. Liao, G. Hong, and S. Li, "Opportunistic osteoporosis screening in multi-detector ct images using deep convolutional neural networks," *European Radiology*, vol. 31, no. 4, pp. 1831–1842, 2021.

[95] A. Krishnaraj, S. Barrett, O. Bregman-Amitai, M. Cohen-Sfady, A. Bar, D. Chettrit, M. Orlovsky, and E. Elnekave, "Simulating dual-energy x-ray absorptiometry in ct using deep-learning segmentation cascade," *Journal of the American College of Radiology*, vol. 16, no. 10, pp. 1473–1479, 2019.

[96] L. Folle, T. Meinderink, D. Simon, A.-M. Liphardt, G. Krönke, G. Schett, A. Kleyer, and A. Maier, "Deep learning methods allow fully automated segmentation of metacarpal bones to quantify volumetric bone mineral density," *Scientific reports*, vol. 11, no. 1, pp. 1–9, 2021.

[97] J. C. G. Sánchez, M. Magnusson, M. Sandborg, Å. C. Tedgren, and A. Malusek, "Segmentation of bones in medical dual-energy computed tomography volumes using the 3d u-net," *Physica Medica*, vol. 69, pp. 241–247, 2020.

[98] Z. Liu, X. Liu, B. Xiao, S. Wang, Z. Miao, Y. Sun, and F. Zhang, "Segmentation of organs-at-risk in cervical cancer ct images with a convolutional neural network," *Physica Medica*, vol. 69, pp. 184–191, 2020.

[99] R. Hemke, C. G. Buckless, A. Tsao, B. Wang, and M. Torriani, "Deep learning for automated segmentation of pelvic muscles, fat, and bone from ct studies for body composition assessment," *Skeletal radiology*, vol. 49, no. 3, pp. 387–395, 2020.

[100] D. Dreizin, Y. Zhou, Y. Zhang, N. Tirada, and A. L. Yuille, "Performance of a deep learning algorithm for automated segmentation and quantification of traumatic pelvic hematomas on ct," *Journal of digital imaging*, vol. 33, no. 1, pp. 243–251, 2020.

[101] F. Chen, J. Liu, Z. Zhao, M. Zhu, and H. Liao, "Three-dimensional feature-enhanced network for automatic femur segmentation," *IEEE journal of biomedical and health informatics*, vol. 23, no. 1, pp. 243–252, 2017.

[102] J. Chen, Y. Li, L. P. Luna, H. W. Chung, S. P. Rowe, Y. Du, L. B. Solnes, and E. C. Frey, "Learning fuzzy clustering for spect/ct segmentation via convolutional neural networks," *Medical physics*, 2021.

[103] Y. Song, H. Lu, H. Kim, S. Murakami, M. Ueno, T. Terasawa, and T. Aoki, "Segmentation of bone metastasis in ct images based on modified hed," in *2019 19th International Conference on Control, Automation and Systems (ICCAS)*, pp. 812–815, IEEE, 2019.

[104] Q. Lin, M. Luo, R. Gao, T. Li, Z. Man, Y. Cao, and H. Wang, "Deep learning based automatic segmentation of metastasis hotspots in thorax bone spect images," *PloS one*, vol. 15, no. 12, p. e0243253, 2020.

[105] N. Moreau, C. Rousseau, C. Fourcade, G. Santini, L. Ferrer, M. Lacombe, C. Guillerminet, M. Campone, M. Colombié, M. Rubeaux, *et al.*, "Deep learning approaches for bone and bone lesion segmentation on 18fdg pet/ct imaging in the context of metastatic breast cancer," in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pp. 1532–1535, IEEE, 2020.

[106] J. Léger, L. Leyssens, C. De Vleeschouwer, and G. Kerckhofs, "Deep learning-based segmentation of mineralized cartilage and bone in high-resolution micro-ct images," in *International Symposium on Computer Methods in Biomechanics and Biomedical Engineering. Springer*, pp. 158–170, 2019.

[107] K. Uemura, Y. Otake, M. Takao, M. Soufi, A. Kawasaki, N. Sugano, and Y. Sato, "Automated segmentation of an intensity calibration phantom in clinical ct images using a convolutional neural network," *International Journal of Computer Assisted Radiology and Surgery*, pp. 1–10, 2021.

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

[109] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3431–3440, 2015.

[110] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn," in *Proceedings of the IEEE international conference on computer vision*, pp. 2961–2969, 2017.

[111] D. Dreizin, Y. Zhou, T. Chen, G. Li, A. L. Yuille, A. McLenithan, and J. J. Morrison, "Deep learning-based quantitative visualization and measurement of extraperitoneal hematoma volumes in patients with pelvic fractures: potential role in personalized forecasting and decision support," *The journal of trauma and acute care surgery*, vol. 88, no. 3, p. 425, 2020.

[112] S. Perera, N. Barnes, X. He, S. Izadi, P. Kohli, and B. Glocker, "Motion segmentation of truncated signed distance function based volumetric surfaces," in *2015 IEEE Winter Conference on Applications of Computer Vision*, pp. 1046–1053, IEEE, 2015.

[113] Y. Hiasa, Y. Otake, M. Takao, T. Ogawa, N. Sugano, and Y. Sato, "Automated muscle segmentation from clinical ct using bayesian u-net for personalized musculoskeletal modeling," *IEEE transactions on medical imaging*, vol. 39, no. 4, pp. 1030–1040, 2019.

[114] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proceedings of the IEEE international conference on computer vision*, pp. 2223–2232, 2017.

[115] S. Liu, D. Xu, S. K. Zhou, O. Pauly, S. Grbic, T. Mertelmeier, J. Wicklein, A. Jerebko, W. Cai, and D. Comaniciu, "3d anisotropic hybrid network: Transferring convolutional features from 2d images to 3d anisotropic volumes," in *International conference on medical image computing and computer-assisted intervention*, pp. 851–858, Springer, 2018.

[116] A. Sekuboyina, M. Rempfler, J. Kukačka, G. Tetteh, A. Valentinitsch, J. S. Kirschke, and B. H. Menze, "Btrfly net: Vertebrae labelling with energy-based adversarial learning of local spine prior," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 649–657, Springer, 2018.

[117] B. Zhang, C. Jia, R. Wu, B. Lv, B. Li, F. Li, G. Du, Z. Sun, and X. Li, "Improving rib fracture detection accuracy and reading efficiency with deep learning-based detection software: a clinical evaluation," *The British Journal of Radiology*, vol. 94, no. 1118, p. 20200870, 2021.

[118] H.-D. Nguyen and S.-H. Kim, "Automatic whole-body bone age assessment using deep hierarchical features," *arXiv preprint arXiv:1901.10237*, 2019.

[119] G. González, G. R. Washko, and R. S. J. Estépar, "Deep learning for biomarker regression: application to osteoporosis and emphysema on chest ct scans," in *Medical Imaging 2018: Image Processing*, vol. 10574, p. 105741H, International Society for Optics and Photonics, 2018.

[120] K. Yasaka, H. Akai, A. Kunimatsu, S. Kiryu, and O. Abe, "Prediction of bone mineral density from computed tomography: application of deep learning with a convolutional neural network," *European radiology*, pp. 1–9, 2020.

[121] M. Pisov, V. Kondratenko, A. Zakharov, A. Petraikin, V. Gombolevskiy, S. Morozov, and M. Belyaev, "Keypoints localization for joint vertebra detection and fracture severity quantification," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 723–732, Springer, 2020.

[122] H. S. Yun, T. J. Jang, S. M. Lee, S.-H. Lee, and J. K. Seo, "Learning-based local-to-global landmark annotation for automatic 3d cephalometry," *Physics in Medicine & Biology*, vol. 65, no. 8, p. 085018, 2020.

[123] Y. Chen, Y. Gao, K. Li, L. Zhao, and J. Zhao, "vertebrae identification and localization utilizing fully convolutional networks and a hidden markov model," *IEEE transactions on medical imaging*, vol. 39, no. 2, pp. 387–399, 2019.

[124] H. Liao, A. Mesfin, and J. Luo, "Joint vertebrae identification and localization in spinal ct images by combining short-and long-range contextual information," *IEEE transactions on medical imaging*, vol. 37, no. 5, pp. 1266–1275, 2018.

[125] R. Jakubicek, J. Chmelik, P. Ourednicek, and J. Jan, "Deep-learning-based fully automatic spine centerline detection in ct data," in *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 2407–2410, IEEE, 2019.

[126] Y. Cai, M. Landis, D. T. Laidley, A. Kornecki, A. Lum, and S. Li, "Multi-modal vertebrae recognition using transformed deep convolution network," *Computerized medical imaging and graphics*, vol. 51, pp. 11–19, 2016.

[127] A. Suzani, A. Seitel, Y. Liu, S. Fels, R. N. Rohling, and P. Abolmaesumi, "Fast automatic vertebrae detection and localization in pathological ct scans-a deep learning approach," in *International conference on medical image computing and computer-assisted intervention*, pp. 678–686, Springer, 2015.

[128] D. Xiao, C. Lian, H. Deng, T. Kuang, Q. Liu, L. Ma, D. Kim, Y. Lang, X. Chen, J. Gateno, *et al.*, "Estimating reference bony shape models for orthognathic surgical planning using 3d point-cloud deep learning," *IEEE Journal of Biomedical and Health Informatics*, 2021.

[129] M. Thies, J.-N. Zäch, C. Gao, R. Taylor, N. Navab, A. Maier, and M. Unberath, "A learning-based method for online adjustment of c-arm cone-beam ct source trajectories for artifact avoidance," *International journal of computer assisted radiology and surgery*, vol. 15, no. 11, pp. 1787–1796, 2020.

[130] S. Jégou, M. Drozdzal, D. Vazquez, A. Romero, and Y. Bengio, "The one hundred layers tiramisu: Fully convolutional densenets for semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pp. 11–19, 2017.

[131] H. Chen, Y. Zhang, W. Zhang, P. Liao, K. Li, J. Zhou, and G. Wang, "Low-dose ct via convolutional neural network," *Biomedical optics express*, vol. 8, no. 2, pp. 679–694, 2017.

[132] Y. Hiasa, Y. Otake, M. Takao, T. Matsuoka, K. Takashima, A. Carass, J. L. Prince, N. Sugano, and Y. Sato, "Cross-modality image synthesis from unpaired data using cyclegan," in *International workshop on simulation and synthesis in medical imaging*, pp. 31–41, Springer, 2018.

[133] F. Liu, H. Jang, R. Kijowski, G. Zhao, T. Bradshaw, and A. B. McMillan, "A deep learning approach for 18 f-fdg pet attenuation correction," *EJNMMI physics*, vol. 5, no. 1, pp. 1–15, 2018.

[134] A. P. Leynes, J. Yang, F. Wiesinger, S. S. Kaushik, D. D. Shanbhag, Y. Seo, T. A. Hope, and P. E. Larson, "Zero-echo-time and dixon deep pseudo-ct (zedd ct): direct generation of pseudo-ct images for pelvic pet/mri attenuation correction using deep convolutional neural networks with multiparametric mri," *Journal of Nuclear Medicine*, vol. 59, no. 5, pp. 852–858, 2018.

[135] D. Kawahara, A. Saito, S. Ozawa, and Y. Nagata, "Image synthesis with deep convolutional generative adversarial networks for material decomposition in dual-energy ct from a kilovoltage ct," *Computers in Biology and Medicine*, vol. 128, p. 104111, 2021.

[136] T.-H. Yong, S. Yang, S.-J. Lee, C. Park, J.-E. Kim, K.-H. Huh, S.-S. Lee, M.-S. Heo, and W.-J. Yi, "Qcbct-net for direct measurement of bone mineral density from quantitative cone-beam ct: A human skull phantom study," *Scientific Reports*, vol. 11, no. 1, pp. 1–13, 2021.

[137] Y. Koike, S. Ohira, Y. Akino, T. Sagawa, M. Yagi, Y. Ueda, M. Miyazaki, I. Sumida, T. Teshima, and K. Ogawa, "Deep learning-based virtual noncontrast ct for volumetric modulated arc therapy planning: Comparison with a dual-energy ct-based approach," *Medical physics*, vol. 47, no. 2, pp. 371–379, 2020.

[138] J. Park, D. Hwang, K. Y. Kim, S. K. Kang, Y. K. Kim, and J. S. Lee, "Computed tomography super-resolution using deep convolutional neural network," *Physics in Medicine & Biology*, vol. 63, no. 14, p. 145011, 2018.

[139] C. You, G. Li, Y. Zhang, X. Zhang, H. Shan, M. Li, S. Ju, Z. Zhao, Z. Zhang, W. Cong, *et al.*, "Ct super-resolution gan constrained by the identical, residual, and cycle learning ensemble (gan-circle)," *IEEE transactions on medical imaging*, vol. 39, no. 1, pp. 188–203, 2019.

[140] I. Guha, S. A. Nadeem, C. You, X. Zhang, S. M. Levy, G. Wang, J. C. Torner, and P. K. Saha, "Deep learning based high-resolution reconstruction of trabecular bone microstructures from low-resolution ct scans using gan-circle," in *Medical*

Imaging 2020: Biomedical Applications in Molecular, Structural, and Functional Imaging, vol. 11317, p. 113170U, International Society for Optics and Photonics, 2020.

[141] Y. Nomura, Q. Xu, H. Shirato, S. Shimizu, and L. Xing, "Projection-domain scatter correction for cone beam computed tomography using a residual convolutional neural network," *Medical physics*, vol. 46, no. 7, pp. 3142–3155, 2019.

[142] W.-A. Lin, H. Liao, C. Peng, X. Sun, J. Zhang, J. Luo, R. Chellappa, and S. K. Zhou, "Dudonet: Dual domain network for ct metal artifact reduction," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 10512–10521, 2019.

[143] Y. Zhang and H. Yu, "Convolutional neural network based metal artifact reduction in x-ray computed tomography," *IEEE transactions on medical imaging*, vol. 37, no. 6, pp. 1370–1381, 2018.

[144] H. Liao, W.-A. Lin, S. K. Zhou, and J. Luo, "Adn: artifact disentanglement network for unsupervised metal artifact reduction," *IEEE transactions on medical imaging*, vol. 39, no. 3, pp. 634–643, 2019.

[145] Q. Mai and J. W. Wan, "Metal artifacts reduction in ct scans using convolutional neural network with ground truth elimination," in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pp. 1319–1322, IEEE, 2020.

[146] Y. Chen, X. Yin, L. Shi, H. Shu, L. Luo, J.-L. Coatrieux, and C. Toumoulin, "Improving abdomen tumor low-dose ct images using a fast dictionary learning based processing," *Physics in Medicine & Biology*, vol. 58, no. 16, p. 5803, 2013.

[147] P. F. Feruglio, C. Vinegoni, J. Gros, A. Sbarbati, and R. Weissleder, "Block matching 3d random noise filtering for absorption optical projection tomography," *Physics in Medicine & Biology*, vol. 55, no. 18, p. 5401, 2010.

[148] J. M. Verburg and J. Seco, "Ct metal artifact reduction method correcting for beam hardening and missing projections," *Physics in Medicine & Biology*, vol. 57, no. 9, p. 2803, 2012.

[149] CarlosAB. http://spineweb.digitalimaginggroup.ca/.

[150] K. Clark, B. Vendt, K. Smith, J. Freymann, J. Kirby, P. Koppel, S. Moore, S. Phillips, D. Maffitt, M. Pringle, *et al.*, "The cancer imaging archive (tcia): maintaining and operating a public information repository," *Journal of digital imaging*, vol. 26, no. 6, pp. 1045–1057, 2013.

[151] H. Pampel, P. Vierkant, F. Scholze, R. Bertelmann, M. Kindling, J. Klump, H.-J. Goebelbecker, J. Gundlach, P. Schirmbacher, and U. Dierolf, "Making research data repositories visible: the re3data. org registry," *PloS one*, vol. 8, no. 11, p. e78080, 2013.

[152] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, "Swin transformer: Hierarchical vision transformer using shifted windows," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 10012–10022, 2021.

[153] H. Cao, Y. Wang, J. Chen, D. Jiang, X. Zhang, Q. Tian, and M. Wang, "Swin-unet: Unet-like pure transformer for medical image segmentation," *arXiv preprint arXiv:2105.05537*, 2021.

[154] K. Cao, C. Wei, A. Gaidon, N. Arechiga, and T. Ma, "Learning imbalanced datasets with label-distribution-aware margin loss," *Advances in neural information processing systems*, vol. 32, 2019.

[155] A. Rajkomar, E. Oren, K. Chen, A. M. Dai, N. Hajaj, M. Hardt, P. J. Liu, X. Liu, J. Marcus, M. Sun, *et al.*, "Scalable and accurate deep learning with electronic health records," *NPJ Digital Medicine*, vol. 1, no. 1, pp. 1–10, 2018.

[156] C. Xiao, E. Choi, and J. Sun, "Opportunities and challenges in developing deep learning models using electronic health records data: a systematic review," *Journal of the American Medical Informatics Association*, vol. 25, no. 10, pp. 1419–1428, 2018.

[157] A. Kendall and Y. Gal, "What uncertainties do we need in bayesian deep learning for computer vision?," *Advances in neural information processing systems*, vol. 30, 2017.

[158] Y. Gal and Z. Ghahramani, "Dropout as a bayesian approximation: Representing model uncertainty in deep learning," in *international conference on machine learning*, pp. 1050–1059, PMLR, 2016.

[159] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-cam: Visual explanations from deep networks via gradient-based localization," in *Proceedings of the IEEE international conference on computer vision*, pp. 618–626, 2017.

[160] M. Brehler, A. Islam, L. Vogelsang, D. Yang, W. Sehnert, D. Shakoor, S. Demehri, J. H. Siewerdsen, and W. Zbijewski, "Coupled active shape models for automated segmentation and landmark localization in high-resolution ct of the foot and ankle," in *Medical Imaging 2019: Biomedical Applications in Molecular, Structural, and Functional Imaging*, vol. 10953, p. 109530P, International Society for Optics and Photonics, 2019.

[161] A. Leardini, M. Benedetti, F. Catani, L. Simoncini, and S. Giannini, "An anatomically based protocol for the description of foot segment kinematics during gait," *Clinical Biomechanics*, vol. 14, no. 8, pp. 528–536, 1999.

[162] O. K. Nwawka, D. Hayashi, L. E. Diaz, A. R. Goud, W. F. Arndt, F. W. Roemer, N. Malguria, and A. Guermazi, "Sesamoids and accessory ossicles of the foot: anatomical variability and related pathology," *Insights into imaging*, vol. 4, no. 5, pp. 581–593, 2013.

[163] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE transactions on systems, man, and cybernetics*, vol. 9, no. 1, pp. 62–66, 1979.

[164] T. B. Sebastian, H. Tek, J. J. Crisco, and B. B. Kimia, "Segmentation of carpal bones from ct images using skeletally coupled deformable models," *Medical Image Analysis*, vol. 7, no. 1, pp. 21–45, 2003.

[165] X. Chen, J. Graham, and C. Hutchinson, "Integrated frameworkfor simultaneous segmentation and registration of carpal bones," in *2011 18th IEEE International Conference on Image Processing*, pp. 433–436, IEEE, 2011.

[166] E. M. A. Anas, A. Rasoulian, A. Seitel, K. Darras, D. Wilson, P. S. John, D. Pichora, P. Mousavi, R. Rohling, and P. Abolmaesumi, "Automatic segmentation of wrist bones in ct using a statistical wrist shape + pose model," *IEEE transactions on medical imaging*, vol. 35, no. 8, pp. 1789–1801, 2016.

[167] W. Xue, "Unet-based fully-automatic segmentation of the capitate from ct images," 2021.

[168] W. Bai, O. Oktay, M. Sinclair, H. Suzuki, M. Rajchl, G. Tarroni, B. Glocker, A. King, P. M. Matthews, and D. Rueckert, "Semi-supervised learning for network-based cardiac mr image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 253–260, Springer, 2017.

[169] S. Sedai, B. Antony, R. Rai, K. Jones, H. Ishikawa, J. Schuman, W. Gadi, and R. Garnavi, "Uncertainty guided semi-supervised segmentation of retinal layers in oct images," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 282–290, Springer, 2019.

[170] Z. Zhao, X. Zhang, C. Chen, W. Li, S. Peng, J. Wang, X. Yang, L. Zhang, and Z. Zeng, "Semi-supervised self-taught deep learning for finger bones segmentation," in *2019 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI)*, pp. 1–4, IEEE, 2019.

[171] D. Cheng, G. Meng, S. Xiang, and C. Pan, "Fusionnet: Edge aware deep convolutional networks for semantic segmentation of remote sensing harbor images," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 10, no. 12, pp. 5769–5783, 2017.

[172] T. Takikawa, D. Acuna, V. Jampani, and S. Fidler, "Gated-scnn: Gated shape cnns for semantic segmentation," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 5229–5238, 2019.

[173] Y. Yuan, J. Xie, X. Chen, and J. Wang, "Segfix: Model-agnostic boundary refinement for segmentation," in *European Conference on Computer Vision*, pp. 489–506, Springer, 2020.

[174] P.-S. Liao, T.-S. Chen, P.-C. Chung, *et al.*, "A fast algorithm for multilevel thresholding," *J. Inf. Sci. Eng.*, vol. 17, no. 5, pp. 713–727, 2001.

[175] C. H. Sudre, W. Li, T. Vercauteren, S. Ourselin, and M. J. Cardoso, "Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations," in *Deep learning in medical image analysis and multimodal learning for clinical decision support*, pp. 240–248, Springer, 2017.

[176] D. Wu, M. Sofka, N. Birkbeck, and S. K. Zhou, "Segmentation of multiple knee bones from ct for orthopedic knee surgery planning," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 372–380, Springer, 2014.

[177] D. D. Ruikar, K. Santosh, and R. S. Hegadi, "Segmentation and analysis of ct images for bone fracture detection and labeling," in *Medical Imaging*, pp. 130–154, CRC Press, 2019.

[178] N. Lessmann, B. Van Ginneken, P. A. De Jong, and I. Išgum, "Iterative fully convolutional neural networks for automatic vertebra segmentation and identification," *Medical image analysis*, vol. 53, pp. 142–155, 2019.

[179] N. Kamiya, "Deep learning technique for musculoskeletal analysis," *Deep learning in medical image analysis*, pp. 165–176, 2020.

[180] S. M. Anwar, I. Irmakci, D. A. Torigian, S. Jambawalikar, G. Z. Papadakis, C. Akgun, J. Ellermann, M. Akcakaya, and U. Bagci, "Semi-supervised deep learning for multi-tissue segmentation from multi-contrast mri," *Journal of Signal Processing Systems*, pp. 1–14, 2020.

[181] W. Burton II, C. Myers, and P. Rullkoetter, "Semi-supervised learning for automatic segmentation of the knee from mri with convolutional neural networks," *Computer methods and programs in biomedicine*, vol. 189, p. 105328, 2020.

[182] H. Liu, H. Xiao, L. Luo, C. Feng, B. Yin, D. Wang, Y. Li, S. He, and G. Fan, "Semi-supervised semantic segmentation of multiple lumbosacral structures on ct," in *International Workshop and Challenge on Computational Methods and Clinical Applications for Spine Imaging*, pp. 47–59, Springer, 2019.

[183] C. A. Peña-Solórzano, D. W. Albrecht, R. Bassed, J. Gillam, P. Harris, and M. Dimmock, "Semi-supervised labelling of the femur in a whole-body post-mortem ct database using deep learning," *Computers in Biology and Medicine*, vol. 122, p. 103797, 2020.

[184] L. Yu, S. Wang, X. Li, C.-W. Fu, and P.-A. Heng, "Uncertainty-aware self-ensembling model for semi-supervised 3d left atrium segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 605–613, Springer, 2019.

[185] H. Zheng, L. Lin, H. Hu, Q. Zhang, Q. Chen, Y. Iwamoto, X. Han, Y.-W. Chen, R. Tong, and J. Wu, "Semi-supervised segmentation of liver using adversarial learning with deep atlas prior," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 148–156, Springer, 2019.

[186] X. Luo, J. Chen, T. Song, and G. Wang, "Semi-supervised medical image segmentation through dual-task consistency," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, pp. 8801–8809, 2021.

[187] X. Li, L. Yu, H. Chen, C.-W. Fu, and P.-A. Heng, "Semi-supervised skin lesion segmentation via transformation consistent self-ensembling model," *arXiv preprint arXiv:1808.03887*, 2018.

[188] X. Li, L. Yu, H. Chen, C.-W. Fu, L. Xing, and P.-A. Heng, "Transformation-consistent self-ensembling model for semisupervised medical image segmentation," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 2, pp. 523–534, 2020.

[189] J.-A. Pérez-Carrasco, B. Acha, C. Suárez-Mejías, J.-L. López-Guerra, and C. Serrano, "Joint segmentation of bones and muscles using an intensity and histogram-based energy minimization approach," *Computer methods and programs in biomedicine*, vol. 156, pp. 85–95, 2018.

[190] A. Tarvainen and H. Valpola, "Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results," *arXiv preprint arXiv:1703.01780*, 2017.

[191] E. M. Lewiecki and N. E. Lane, "Common mistakes in the clinical use of bone mineral density testing," *Nature clinical practice Rheumatology*, vol. 4, no. 12, pp. 667–674, 2008.

[192] E. S. Siris, Y.-T. Chen, T. A. Abbott, E. Barrett-Connor, P. D. Miller, L. E. Wehren, and M. L. Berger, "Bone mineral density thresholds for pharmacological intervention to prevent fractures," *Archives of internal medicine*, vol. 164, no. 10, pp. 1108–1112, 2004.

[193] D. Kim and T. MacKinnon, "Artificial intelligence in fracture detection: transfer learning from deep convolutional neural networks," *Clinical radiology*, vol. 73, no. 5, pp. 439–445, 2018.

[194] J. J. Schreiber, E. B. Gausden, P. A. Anderson, M. G. Carlson, and A. J. Weiland, "Opportunistic osteoporosis screening—gleaning additional information from diagnostic wrist ct scans," *JBJS*, vol. 97, no. 13, pp. 1095–1100, 2015.

[195] M. E. Arlot, E. Sornay-Rendu, P. Garnero, B. Vey-Marty, and P. D. Delmas, "Apparent pre-and postmenopausal bone loss evaluated by dxa at different skeletal sites in women: The ofely cohort," *Journal of Bone and Mineral Research*, vol. 12, no. 4, pp. 683–690, 1997.

[196] R. Chapurlat, J.-B. Pialat, B. Merle, E. Confavreux, F. Duvert, E. Fontanges, F. Khacef, S. L. Peres, A.-M. Schott, and E. Lespessailles, "The qualyor (qualite osseuse lyon orleans) study: a new cohort for non invasive evaluation of bone quality in postmenopausal osteoporosis. rationale and study design," *Archives of osteoporosis*, vol. 13, no. 1, p. 2, 2018.

[197] M. Hars, E. Biver, T. Chevalley, F. Herrmann, R. Rizzoli, S. Ferrari, and A. Trombetti, "Low lean mass predicts incident fractures independently from frax: a prospective cohort study of recent retirees," *Journal of bone and mineral research*, vol. 31, no. 11, pp. 2048–2056, 2016.

[198] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE conference on computer vision and pattern recognition*, pp. 248–255, Ieee, 2009.

[199] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[200] C. Cortes and M. Mohri, "Confidence intervals for the area under the roc curve," *Advances in neural information processing systems*, vol. 17, pp. 305–312, 2005.