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

1 Introduction **1**

1.1 Background and Motivation 1

1.2 Objectives and Contributions 6

1.3 Thesis Organization 7

2 Literature Review **11**

2.1 Classification Tasks for Bone CT Analysis 11

2.2 Segmentation Tasks for Bone CT Analysis 13

2.3 Regression Tasks for Bone CT Analysis 24

2.4 Generative Tasks for Bone CT Analysis 30

2.5 Discussion and Conclusion 33

3 Anatomical Segmentation of Human Foot CT **39**

3.1 Introduction 39

3.2 The Multi-Stage FootSeg Method for Foot Anatomical Segmentation . 43

3.2.1 Foot Standardization 44

3.2.2 Foot Bone Region Segmentation 44
3.2.3 Foot Bone Classification 45
3.3 Experimental Results 48
3.3.1 Data Collection, Annotation and Database Construction 48
3.3.2 Implementation Details and Evaluation Matrix 49
3.3.3 Bone Region Segmentation Results 50
3.3.4 Bone Classification Results 50
3.4 Conclusion 52

4 Instance Segmentation of Human Wrist CT 53

4.1 Introduction 53
4.2 Overview of the Proposed Methods 56
4.2.1 Overview of the Semi-Automatic Construction Method of the Wrist Instance Segmentation Database 57
4.2.2 Overview of the Edge-Enhanced Wrist Instance Segmentation Model 59
4.3 Semi-Automatic Construction Method of the Wrist Annotation Database 59
4.3.1 OTSU-Based Radius, Ulna, Muscle-Cast, and Background Segmentation Method 60
4.3.2 U-net-Based Cast Segmentation Method 62
4.3.3 Results of the Semi-Automatic Construction Method of the Wrist Instance Annotation Database 63
4.3.4 Data Annotation Time Analysis 65
4.4 Edge-Enhanced Wrist Instance Segmentation Model 66
4.4.1 Method of the Edge-Enhanced Wrist Instance Segmentation Model 66
4.4.2 Wrist Instance Segmentation Results 67
4.5 Conclusion 69

5 Semi-Supervised Segmentation of Bone CT **71**

5.1 Introduction 71

5.1.1 Motivation for the Semi-Supervised Segmentation of Bone CT . 71

5.1.2 The Particular CT Attribute for Semi-Supervised Segmentation of Bone CT 73

5.2 Method of Patch-Shuffle-Based Semi-Supervised Segmentation of Bone CT 77

5.2.1 Overview of the Proposed Method 77

5.2.2 Patch-Shuffle-Based Semi-Supervised Segmentation 78

5.3 Experiments and Analysis 82

5.3.1 Datasets for Model Evaluation 82

5.3.2 Experiment Settings and Evaluation Metrics 84

5.3.3 Results on the Wrist CT Scan Dataset 88

5.3.4 Results on the Foot Bone CT Dataset 90

5.3.5 Results on the USEvillaBone Dataset 91

5.3.6 Qualitative Analysis on the Segmentation Results and Feature Maps 92

5.4 Conclusion 92

6 Bone Health Analysis via Bone Fracture Prediction using Wrist CT **95**

6.1 Introduction 95

6.2 Structured Clinical Database Construction 98

6.2.1 Clinical and Wrist CT Raw Data Collection 98

6.2.2 Clinical Raw Data Processing 100

6.2.3 Wrist CT Raw Data Processing 106

6.2.4 The Structured Wrist Database 108

6.3 Data Selection and Method of Bone Fracture Prediction in Next Five Years . . . 111
6.3.1 Data Statistics of Age, BMD T-score, FRAX Score, Fracture and Non-Fracture Number in Different Years . . . 111
6.3.2 Data Selection According to Five Year Selection Criteria . . . 114
6.3.3 Method of Multi-Task Based Bone Fracture Prediction in Next Five Years . . . 115
6.4 Results . . . 120
6.4.1 Evaluation of the Proposed Model . . . 120
6.4.2 Our Model Performance on Fragility Fracture Prediction . . . 120
6.4.3 Results of Major Fragility Fracture Prediction on All Ages . . . 123
6.4.4 Results of Fragility Fracture Prediction on Ages > 65 . . . 125
6.4.5 Results of Major Fragility Fracture Prediction on Ages > 65 . . . 125
6.4.6 Results of Fragility Fracture Prediction on Ages > 70 . . . 126
6.4.7 Results of Major Fragility Fracture Prediction on Ages > 70 . . . 129
6.4.8 Results Comparison of Models using Different Wrist Parts as Input on Fragility Fracture Prediction on All Ages . . . 130
6.4.9 Results of Prediction of Age and Longest Health Year before the Bone Fragility Fracture and Results using other Deep Learning Models . . . 130
6.4.10 Visualization of Region of Interest for Bone Fracture Prediction using Heatmap . . . 131
6.5 Conclusion . . . 132

7 Conclusion and Future Work **135**
7.1 Conclusion . . . 135
7.2 Future Work . . . 137

References **139**

List of figures

1.1 Comparison of CT, MRI and X-ray data on the human spine. 2

1.2 Scanning procedure (left part) and scanning examples (right part) of weight-bearing CBCT. 3

1.3 Illustration of the Hounsfield scale for bone, muscle, and air on CT of the wrist. 5

2.1 Illustration of applications, research trends, and tasks proportion of deep learning based bone CT analysis. 35

2.2 Several parts to be considered for deep-learning based bone CT analysis. 37

3.1 Illustration of the anatomical segmentation of human foot bones. 40

3.2 Illustration of different foot scans. 41

3.3 Bone point number of different foot bones. 42

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. 42

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. 45

3.6 Examples of different size of the bone patch images. 47

3.7 The CNN model for the feature extraction of patch image. 48

3.8 The anatomical segmentation results on the test dataset. 51

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). 54

4.2 The framework of semi-automatic construction of the wrist instance segmentation database. 57

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). 58

4.4 Edge-enhanced wrist segmentation framework. 59

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. 61

4.6 Example of successful (first row) and unsuccessful (second row) segmentation results from the OTSU-based method. 64

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). 69

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). 70

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. . 74

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. 75

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. 76

5.4 Examples of original slices (left) and patch-shuffled slices (right) of wrist CT. 76

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. 79

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. 81

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). 83

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). 85

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). 86

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. 93

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. 94

6.1 Examples of the OFELY, QUALYOR and GERICO cohorts. 108

6.2 Examples of the segmentation results. 109

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. 110

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. 117

6.5 Examples of the heatmaps (part one). 133

6.6 Examples of the heatmaps (part two). 134

List of tables

2.1 Classification task details on head-related CT. 14
2.2 Classification task details on whole-body, spine, chest, lower body, foot and bone CT. 15
2.3 Segmentation task details on head-related CT. 20
2.6 Segmentation task details on whole body, shoulder, chest, upper body, hand, femur and bone CT. 22
2.4 Segmentation task details on spinal CT. 25
2.5 Segmentation task details on pelvic CT. 26
2.7 Regression task details on head CT. 28
2.8 Regression task details on spine CT. 29
2.9 Regression task details on whole body, upper body, and chest CT. . . . 30
2.10 Generative task details on head CT. 33
2.11 Generative task details on chest, spinal, abdominal, pelvic, hip, lower body, femur and bone CT. 34

3.1 The performance comparison of different models. 51

4.1 Segmentation results and acceptance rate of OTSU-based segmentation model and U-net-based cast segmentation model. 65
4.2 Result comparison of IOU on wrist segmentation of U-net and the proposed model. 67

5.1 HU values of different body tissues and materials. 74
5.2 Comparison of mIOU of different semi-supervised methods under different number of training data. (Unit: %) 88
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: %) 89
5.4 Results of mIOU of the proposed method under different number of labeled training data. (Unit: %) 90
5.5 Results of mIOU of different semi-supervised methods on the foot bone CT dataset. (Unit: %) 91
5.6 Results of mIOU of different semi-supervised methods on the USEvillaBone dataset. (Unit: %) 91

6.1 Clinical raw data recorded at OFELY cohort. 100
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. 101
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. 101
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. 102
6.5 Clinical raw data of the participant's health and fracture information recorded at GERICO cohort. 103
6.6 Clinical data in participant information table. 104
6.7 Clinical data in the fracture information table. 104
6.8 Clinical data in the BMD information table. 105
6.9 Clinical data in the FRAX information table. 105
6.10 Clinical data in the miscellaneous information table. 105

6.11 Data stored in the CT header. 107
6.12 Scan data stored at the scan information table. 107
6.13 The clinical information and scan information table of the women participants. 111
6.14 Statistics of Age, BMD T-score, FRAX, and duration of follow-up of the cohorts. 112
6.15 Statistics of fracture number in different follow-up years. 113
6.16 Statistics of fracture number within different follow-up years. 113
6.17 Statistics of no-fracture number within different follow-up years. 114
6.18 Statistics of case number in different cohorts in the selected data with five year as data select criteria. 115
6.19 Statistics on age, BMD T-score and FRAX of the selected data according to the five-year criteria. 116
6.20 Data number in the four datasets. 121
6.21 Fragility fracture prediction results on data of all ages. 122
6.22 Fragility fracture prediction results on data of all ages with BMD T-score recorded. 122
6.23 Fragility fracture prediction results on data of all ages with FRAX index recorded. 123
6.24 Major fragility fracture prediction results on data of all ages. 123
6.25 Major Fragility fracture prediction results on data of all ages with BMD T-score recorded. 124
6.26 Major fragility fracture prediction results on data of all ages with FRAX index recorded. 124
6.27 Fragility fracture prediction results on data of ages > 65. 125
6.28 Fragility fracture prediction results on data of ages > 65 with BMD T-score recorded. 125

6.29 Fragility fracture prediction results on data of ages > 65 with FRAX index recorded. 126

6.30 Major fragility fracture prediction results on data of ages > 65. 126

6.31 Major fragility fracture prediction results on data of ages > 65 with BMD T-score recorded. 127

6.32 Major fragility fracture prediction results on data of ages > 65 with FRAX index recorded. 127

6.33 Fragility fracture prediction results on data of ages > 70. 127

6.34 Fragility fracture prediction results on data of ages > 70 with BMD T-score recorded. 128

6.35 Fragility fracture prediction results on data of ages > 70 with FRAX index recorded. 128

6.36 Major fragility fracture prediction results on data of ages > 70. 129

6.37 Major fragility fracture prediction results on data of ages > 70 with BMD T-score recorded. 129

6.38 Major fragility fracture prediction results on data of ages > 70 with FRAX index recorded. 130

6.39 Comparison of models using different CT parts as input on fragility fracture prediction on all ages. 131

6.40 Comparison of different deep learning models for fragility fracture prediction task. 132