Doctoral Dissertation

Recovery of patient-specific anatomy and kinematics of musculoskeletal structures from medical images

Yuta Hiasa

March 15, 2019

Graduate School of Information Science

Nara Institute of Science and Technology

A Doctoral Dissertation
submitted to Graduate School of Information Science,
Nara Institute of Science and Technology
in partial fulfillment of the requirements for the degree of
Doctor of ENGINEERING

Yuta Hiasa

Thesis Committee:

Professor Yoshinobu Sato	(Supervisor)
Professor Yasuhiro Mukaigawa	(Co-supervisor)
Associate Professor Yoshito Otake	(Co-supervisor)
Assistant Professor Mazen Soufi	(Co-supervisor)

Recovery of patient-specific anatomy and kinematics of musculoskeletal structures from medical images[*]

Yuta Hiasa

Abstract

The patient-specific anatomy and kinematics are important for preoperative planning, rehabilitation, biomechanical simulation, and so on. The purpose of this thesis is to automate patient-specific recovering of them from multi-modal clinical images, which we treat as three separate challenges: automatic recovery of 1) musculoskeletal anatomy, 2) musculoskeletal kinematics, and 3) from different modalities. Specifically, we tackled following clinical problems.

Automatic recovery of musculoskeletal anatomy: An automated segmentation method from CT based on a hierarchical multi-atlas method has achieved high accuracy, however computational inefficiency has been one of its drawbacks. We focus on segmentation from CT using convolutional neural networks (CNN), which yielded significant improvement of accuracy and computational time. Then, we introduce an active-learning method using Bayesian approach to effectively increase training dataset size.

Automatic recovery of musculoskeletal kinematics: Previous analysis of rib-cage motion using x-ray video has been shown to be effective for evaluating respiratory function, but it has been limited to 2D. We focus on a method to recover 3D rib motion based on 2D-3D registration of x-ray video and single-time-phase CT.

We introduce the following two components into the conventional intensity-based 2D–3D registration pipeline: i) a rib-motion model based on a uniaxial joint to constrain the search space and ii) local contrast normalization as a pre-process of x-ray video to improve the cost function of the optimization parameters, which is often called the landscape.

Automatic recovery from different modalities: Finally, we focus on MR-to-CT synthesis to realize modality-independent segmentation. We extend the Cycle-GAN approach by adding the gradient consistency loss to improve the accuracy at the boundaries of musculoskeletal structures.

Keywords:

Rib motion, Musculoskeletal segmentation, Image synthesis, Convolutional Neural Networks, 2D-3D registration, Dynamic radiograph, CT, MRI

Contents

List of Figures

vii

xi

List of Tables

1 Introduction

1.1 Background

The development of physics so far has realized imaging devices such as radiography, computed tomography (CT) and magnetic resonance imaging (MRI) which are essential for understanding the inside of human body in-vivo.

To understand the human anatomy, Voxel-Man (Pflesser et al. (2001)) and Visible Human Project (Spitzer and Whitlock (1998)) constructed volumetric detailed human anatomy modeling representing few cadavers using CT, MRI, and ex-vivo cross-sectional photography, as shown in Figure 1.1. Those anatomy were recovered by manual or semi-automated segmentation. To understand the 3D human kinematics, computational simulation of the human kinematics using biomechanics modeling has been performed in a number of studies and investigated musculo-tendon forces and joint contact forces (Rajagopal et al. (2016); Seth et al. (2018); Shu et al. (2018)). These techniques are important for preoperative planning, rehabilitation, as shown in Figure 1.2.

Patient-specific human anatomy also plays an important role in biomechanics modeling. Recent studies have demonstrated that personalization of the model parameters, such as the size of the bones, length of the muscle and tendon, physical properties of muscle-tendon complex, and geometry of the muscles, improve the accuracy in the prediction (Moissenet et al. (2017); Chèze, Laurence et al. (2015); Taddei et al. (2012); Webb et al. (2014)). In vivo 3D imaging techniques to recover the human kinematics could contribute to biomechanics modeling's patient-specific adaptation.
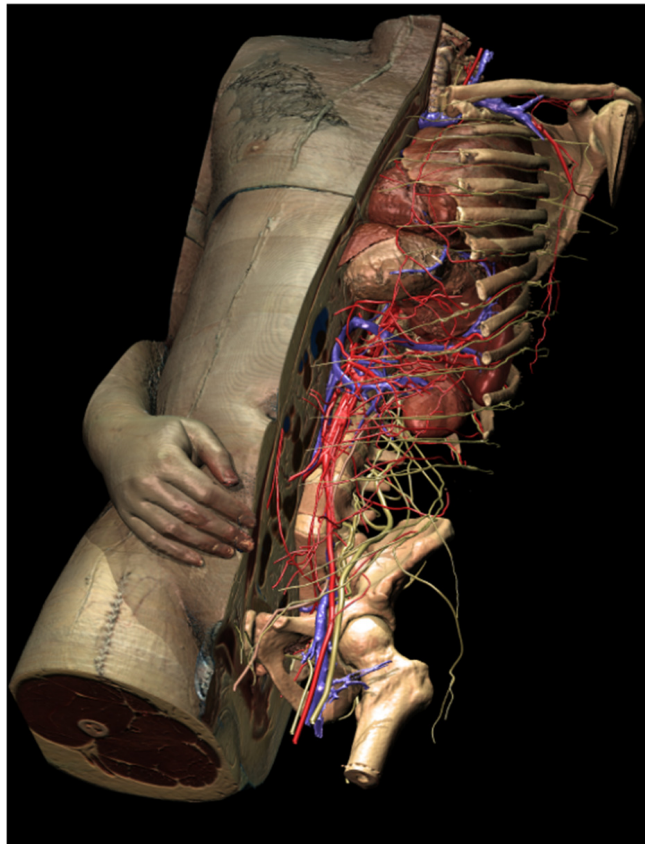
Figure 1.1: Volumetric detailed human anatomy modeling for one representative cadaver. Note that this figure was taken from Voxel-Man (Pflesser et al. (2001)), © 2001 IOS Press.

In clinical situation, the imaging modality is selected considering the advantages and disadvantages of each modality. For example, radiography is a relatively accessible and cheap modality but limited to two-dimensional (2D) analysis. CT has static 3D spatial information with high spatial resolution and it can be acquired in a clinically acceptable time without motion artifact. However, CT requires a larger number of time phases to increase temporal resolution and also requires a larger radiation dose limiting its applicability in a routine clinical setting. While MRI has 3D information and advantages in terms of radiation exposure and superior soft-tissue contrast, its long scan time and patient motion
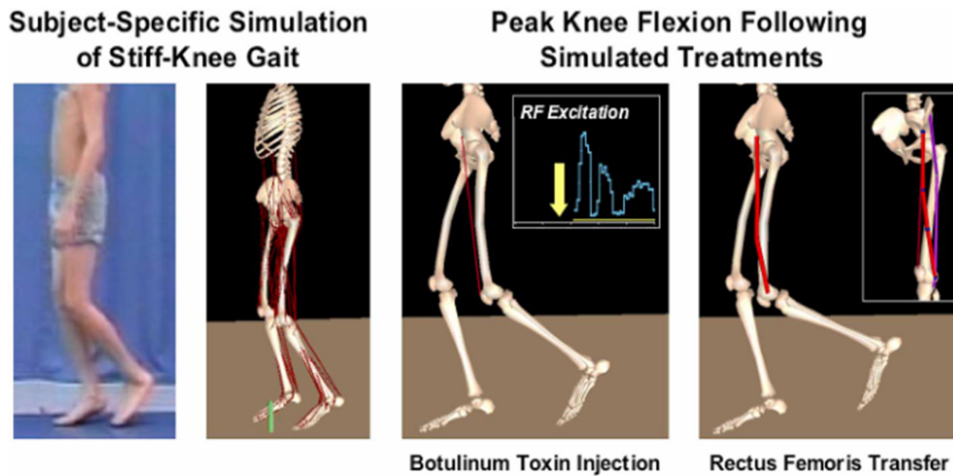
Figure 1.2: Computational simulation of the human kinematics using musculoskeletal modeling. Note that this figure was taken from Delp et al. (2007), the copyright line © 2007 IEEE appear prominently with each reprinted figure.

cause artifacts. And, its small field-of-view (FOV) limits its applicability.

The ultimate goal of our research is to automate patient-specific detailed recovering of full-body anatomy and kinematics from multi-modal clinical images. In this thesis, we especially focus on three separate challenges: automatic recovery of 1) musculoskeletal anatomy, 2) musculoskeletal kinematics, and 3) from different modalities.

1.2 Challenges

Automatic recovery of musculoskeletal anatomy

The recovery of anatomy has been utilized in a number of applications in brain, abdomen, chest, and so on (Litjens et al. (2017)). But, segmentation of individual muscles from CT images is still difficult problem due to the low tissue contrast at the border between neighboring muscles, especially in the area where many muscles are contiguously packed such as the hip and thigh regions.

In order to make the patient-specific muscle modeling available for clinical routine, we propose an application of deep learning. In addition to the segmentation accuracy, we also focus on one issue universal in machine learning based segmentation, which is the manual annotation labor. Especially in segmentation of individual muscle, the manual annotation of detailed human anatomy requires expert knowledge and costly experts' time. The active-learning based on uncertainty estimates in segmentation could help reduce this bottleneck.

Automatic recovery of musculoskeletal kinematics

The recover of human kinematics has been utilized in a number of applications. In chest region, clinical imaging approaches have been proposed for the rib motion analysis. One is multi-phase CT imaging (Beyer et al. (2014)) which directly acquire 3D transformation parameters between two time phases of the rib motion, however, increasing time resolution or follow-up examination is difficult due to the amount of patient dose. In recent years, dynamic radiography based on an x-ray video system with a flat panel detector has been developed (Tanaka (2016)). It is a low-dose and cost-effective functional imaging method in addition to achieving high temporal resolution, whereas its analysis is still limited to 2D.

In this thesis, we combine dynamic chest radiography and a single-time-phase CT to formulate a method for recovering 3D rib motion with high temporal resolution while minimizing increase of radiation dose, which is based on 2D-3D registration. Compared with other typical skeletal x-ray images, the edge detection and labeling of individual bones in the chest radiography are troublesome because other bones (e.g., scapula, clavicle, and sternum) and soft tissues including blood vessels inside the lungs are superimposed, and several ribs are arranged in parallel making anatomical correspondences ambiguous. Therefore, we also need to consider a method to facilitate automation in extraction and recognition of the edges of the ribs used for 2D-3D registration.

Automatic recovery from different modalities

Some previous studies aimed at recover of musculoskeletal anatomy in MRI (Gilles and Magnenat-Thalmann (2010); Ranzini et al. (2017)), but the issues in these studies were the requirement of multiple sequences and devices. The challenge in segmentation of MRI is that there is no standard unit as in CT. Therefore, manually traced label data are necessary for training on each sequence and each imaging device. Thus, MR-to-CT synthesis realizes modality independent segmentation (Hamarneh et al. (2008)).

Image synthesis has been extensively studied using the patch-based learning (Torrado-Carvajal et al. (2016)) as well as deep learning, specifically, convolutional neural networks (CNN) (Zhao et al. (2017)) and generative adversarial networks (GAN) (Kamnitsas et al. (2017)). The conventional approaches required the paired training data, i.e., images of the same patient from multiple modalities that are registered, which limited the application.

1.3 Contributions

Automatic recovery of musculoskeletal anatomy

We demonstrate a significantly improved accuracy and computation time in segmentation of individual 19 muscles from CT of the hip and thigh regions by application of a 2D convolutional neural network (CNN). The slice-by-slice (2D) segmentation approach, as opposed to the volumetric (3D) approach, is advantageous in various realistic clinical contexts such as the case where only few slices are obtained for the reduced radiation exposure in diagnosis of the body composition as well as CTs with various slice intervals. Contribution is three-fold; 1) investigation of performance of the 2D CNN, specifically using the U-Net architecture, 2) extensive quantitative evaluation by fully annotated 20 clinical CTs with the expert trace and partially annotated 18 CTs that are publicly available from The Cancer Imaging Archive (TCIA) database, and comparison with the

5

previous method proposed by Yokota et al. (2018), 3) analysis of uncertainty metric and its potential to application for selection of samples when increasing the training dataset in an active learning manner.

Automatic recovery of musculoskeletal kinematics

The proposed 2D-3D registration method includes the following two components: 1) rib-motion model of a uniaxial joint whose axis location is softly constrained by anatomical knowledge to improve robustness while maintaining accuracy, and 2) local contrast normalization (LCN) for pre-processing of x-ray video images to facilitate automation. Unlike the previous medical applications, which aimed at improving performance in feature extraction, we aimed at improving 2D-3D registration using LCN. In this study, for the first time to the best of our knowledge, we demonstrated that LCN actually changed the cost-function space, which results in avoiding failure due to convergence to the local optimum solutions.

Automatic recovery from different modalities

We extend the CycleGAN approach by adding the gradient consistency (GC) loss to encourage edge alignment between images in the two domains and using an order-of-magnitude larger training data set (302 MR and 613 CT volumes) in order to overcome the larger variation and improve the accuracy at the boundaries. We investigated dependency of image synthesis accuracy on 1) the number of training data and 2) incorporation of the GC loss. To demonstrate the applicability of proposed method, we also investigated a segmentation accuracy on synthesized images.

1.4 Paper organization

Figure 1.3 shows the diagram of this thesis which is organized as follows. In Chapter 2, we discuss automatic segmentation of hip and thigh muscles from CT

using CNN. In Chapter 3, we discuss recovery of three-dimensional rib motion from dynamic chest radiography (x-ray video) based on registration method. In Chapter 4, we discuss an image synthesis method between CT and MRI to realize modality-independent segmentation, and Chapter 5 concludes this thesis.
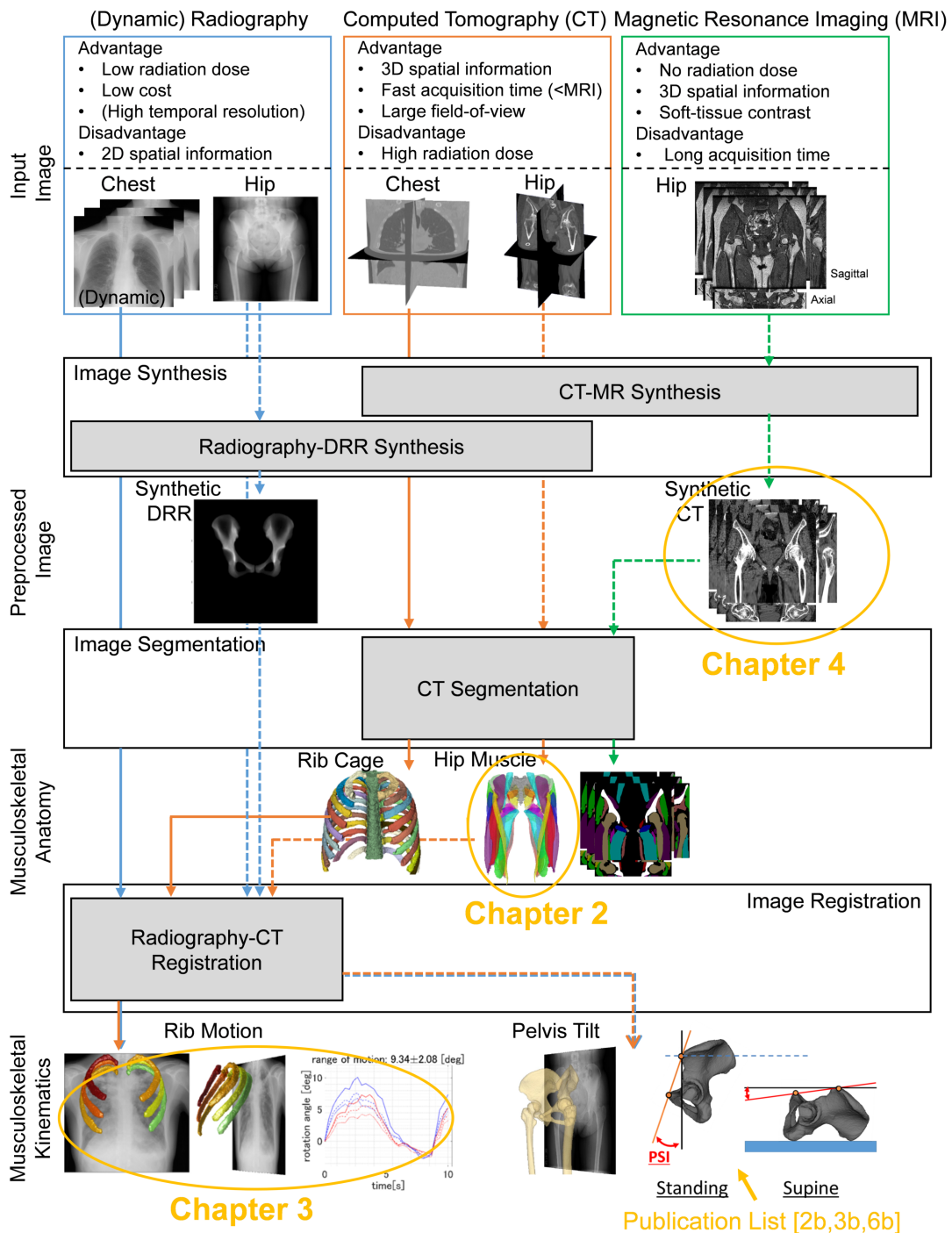
Figure 1.3: Recovery of patient-specific anatomy and kinematics of musculoskeletal structures from medical images. The solid and dashed arrows indicate the chest and hip regions, respectively. The color indicates different input modality.

2 Automated muscle segmentation from clinical CT using 2D Bayesian U-Net for patient-specific musculoskeletal modeling

Although volumetric models representing patient-specific muscle anatomy provide higher accuracy in the simulation (Webb et al. (2014)), the segmentation of the volumetric anatomy of individual muscles from patient-specific medical images remains a time consuming and expert-knowledge-demanding task that precludes application in clinical practice. Thus, our focus in this chapter is automated segmentation of individual muscles for personalization of musculoskeletal model.

2.1 Related work

Segmentation of muscle tissue and fat tissue has been studied extensively for the analysis of muscle/fat composition. (Note that we refer to muscle tissue here as an object including all muscles, not an individual muscle.) Ulbrich (2018) and Karlsson et al. (2015) implemented an algorithm for automated segmentation of the muscle and fat tissues from MRI by a multi-atlas method where multiple atlas images (i.e., a pair of grayscale image and its label image) are non-rigidly
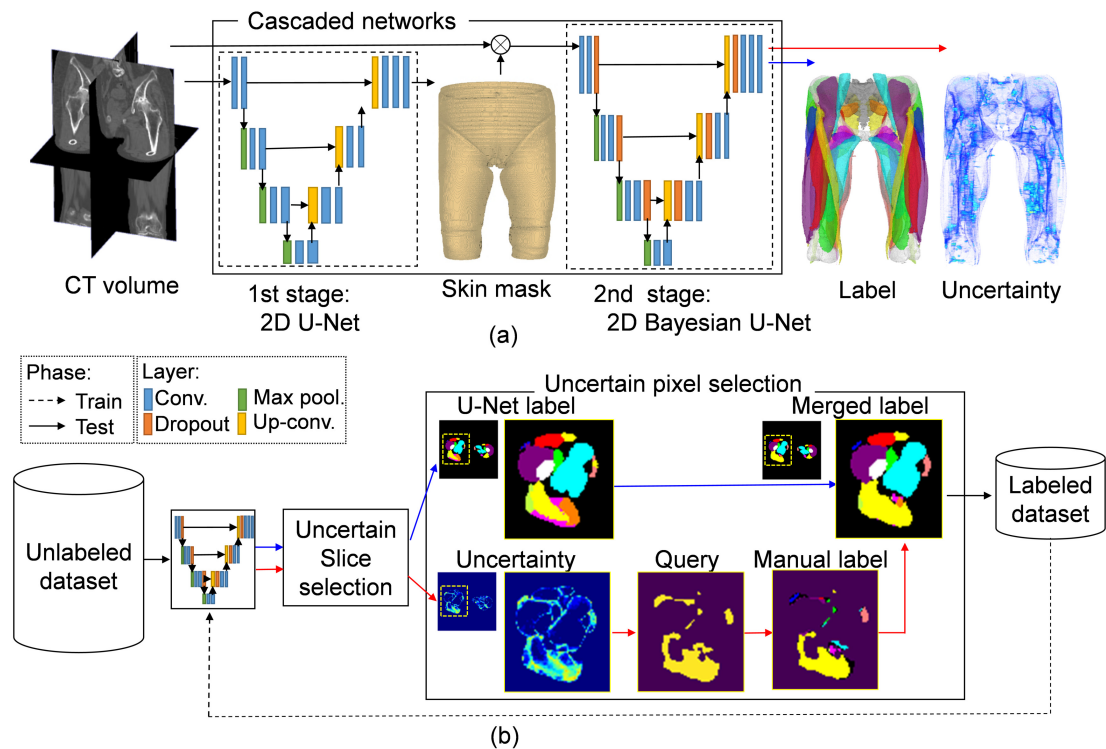
Figure 2.1: Workflow of proposed methods. (a) Segmentation and uncertainty estimation. The skin surface is first segmented by the deterministic U-Net, then individual muscles are segmented and the model uncertainty is predicted by Bayesian U-Net. (b) Active-learning method. First, the networks are applied to unlabeled images and the segmentation and uncertainty are automatically predicted. The uncertain pixels are queried to experts and relabeled manually while the pixels with high confidence are used directly as the training dataset in the next iteration. See texts for detail.

registered to the target image and a voting scheme produces the final segmentation. Lee et al. (2017) used deep learning for segmentation of the muscle and fat tissues in one 2D abdominal CT slice.

Segmentation of individual muscles is much harder problem due to the low tissue contrast at the border between neighboring muscles, especially in the area

where many muscles are contiguously packed such as the hip and thigh regions. Handsfield et al. (2014) manually performed segmentation of 35 individual muscles from MRI of the lower leg in order to investigate the relationship between muscle volume and height or weight. Andrews and Hamarneh (2015) proposed an automated segmentation method for 11 thigh muscles from MRI using a probabilistic shape representation and adjacency information. They evaluated the method with images of the middle part of the left thigh (20 cm in length from above the knee) and reported an accuracy of 0.808 average Dice coefficient. The targeted region along the femur bone has similar appearance in axial slices resulting in less complexity in segmentation problem compared to the hip region.

In CT images, due to the lower soft tissue contrast compared to MRI, segmentation of individual muscles is even difficult. Yokota et al. (2018) addressed automated segmentation of individual muscles from CT of the hip and thigh regions. The target region was broader than Andrews and Hamarneh (2015) covering from the origin to insertion of 19 muscles. They introduced a hierarchization of the multi-atlas segmentation method so that the target region becomes gradually complex in a hierarchical manner, namely, skin surface, all muscle tissues as one object, and individual muscles at each hierarchy. They reported 0.838 average Dice coefficient. Although their algorithm produced a reasonable accuracy in this highly challenging problem, due to the large number of non-rigid registration required in the multi-atlas method, computational load was prohibitive when considering application to clinical routine (41 minutes for segmentation of one CT volume using a high performance server with 60 cores).

In order to make the patient-specific muscle modeling available for clinical routine, we propose an application of deep learning. We investigate the segmentation accuracy as well as the metric indicating uncertainty of the segmentation using the framework of Bayesian deep learning. Gal and Ghahramani (2016) found that the dropout (Srivastava et al. (2014)) is equivalent to approximating the Bayesian inference, which allows estimation of the model uncertainty. It measures the degree of difference of the test samples from the training dataset, i.e., it captures

the lack of training data, which is called *epistemic* uncertainty (Kendall and Gal (2017)). The method has been applied to brain lesion segmentation (Nair et al. (2018); Eaton-Rosen et al. (2018)) and surgical tool segmentation (Hiasa et al. (2018a)). Two example use cases of the uncertainty metric explored in this study are; prediction of segmentation accuracy without using the ground truth similar to the goal of Valindria et al. (2017) and the active-learning framework (Maier-Hein et al. (2016); Yang et al. (2017)) for the reduction of manual annotation cost.

2.2 Methods

2.2.1 Overview

Figure 2.1 shows the workflow of the proposed methods. We first segment the skin surface using a 2D U-Net to isolate the body from surrounding objects such as the scan table. Then the individual muscles are segmented and the model uncertainty is predicted using the Bayesian U-Net, which is described in Section 2.2.3. The average model uncertainty of each muscle region computed in a cross validation within the training dataset and dice coefficient of the muscle computed using the ground truth segmentation was linearly regressed (Fig. 2.8(a)) for prediction of dice coefficient of an unknown test dataset solely from the model uncertainty without using the ground truth. We tested the proposed active-learning framework, shown in Fig. 2.1(b), on a simulated environment using a fully annotated dataset by assuming initially a partial annotation and gradually increasing the amount of manual annotation according to the proposed procedure as described in Section 2.2.4.

2.2.2 Datasets

Two datasets were used to evaluate the proposed method: 1) fully annotated non-public clinical CT dataset and 2) partially annotated publicly available CT

dataset.

Osaka University Hospital THA dataset (THA dataset)

The dataset consists of 20 CT volumes which were scanned at Osaka University Hospital, Suita, Japan, for diagnosis of patients subjected to THA surgery. The field of view was 360×360 mm and the matrix size was 512×512. The original slice interval was 2.0 mm for the region including pelvis and proximal femur, 6.0 mm for the femoral shaft region, and 1.0 mm for the distal femur region. In this study, the CT volumes were resampled so that the slice thickness becomes 1.0 mm throughout the entire volume. 19 muscles around the hip and thigh regions and 3 bones (pelvis, femur, sacrum) were manually annotated by an expert surgeon (Figure 2.2) which took about 40 hours per volume. This dataset was used for training and cross validation for the accuracy evaluation and prediction.

TCIA soft-tissue sarcoma dataset (TCIA dataset)

The dataset obtained from TCIA collections [*] contains CT and MR volumes from 51 patients with soft-tissue sarcomas (STSs) (Vallières et al. (2015)). In this study, we selected 18 CT volumes that include the entire hip and thigh region. The CT volumes were resampled so that the field of view becomes 360×360 mm, which is the same as THA dataset (the cropping range was determined so that the volume center stays the same) and the slice interval becomes 1.0 mm thoughout the volume. The gluteus medius muscle was manually traced by a computer scientist and verified by an expert surgeon. This dataset was not used in the training nor in the parameter tuning and only used for evaluation of generalizability of the model trained with THA dataset (see Section 2.3.2 for detail).

---

[*]`http://www.cancerimagingarchive.net`

| | Hip muscles | | | Thigh muscles | | |

Color	Hip muscle	#	Color	Thigh muscle	#
yellow	Gluteus maximus	#01	cyan	Adductor	#01
green	Gluteus medius	#02	light green	Biceps	#02
pink	Gluteus minimus	#03	pink	Gracilis	#03
purple	Iliacus	#04	red	Rectus femoris	#04
blue	Obturator extermus	#05	yellow	Sartorius	#05
red	Obturator intermus	#06	orange	Semimembranosus	#06
magenta	Pectineus	#07	magenta	Semitendinosus	#07
orange	Piriformis	#08	blue	Tensor fasciae latae	#08
cyan	Psoas major	#09	purple	Vastus lateralis and vastus intermedius	#09
			green	Vastus medialis	#10

Figure 2.2: Training dataset used in this study. CT dataset consists of 20 labeled volumes. Each color shows individual musculoskeletal structures. The target muscles are separately visualized according to the functional group (in columns) and their region (hip and thigh, in rows).

2.2.3 Estimation of uncertainty metric

The underlying algorithm of the proposed accuracy estimates follows that of Gal and Ghahramani (2016) which used the dropout at the test phase allowing to approximate the posterior distribution base on the probabilistic softmax output obtained from the stochastic dropout sampling. We use the mean and variance of the output from the multiple samplings as the segmentation result and the uncertainty estimate. Below, we briefly summarize the general theoretical background described in Gal and Ghahramani (2016), formulate the uncertainty that we employed in this paper, and propose structure-wise metrics that can be used in a multi-class segmentation for active learning.

Let a "*deterministic*" neural network represents $p(y|\mathbf{x}) = \text{Softmax}(\mathbf{f}(\mathbf{x}; \mathbf{W}))$. A "*probabilistic*" Bayesian neural network is given by marginalization over the weight $\mathbf{W}$ as

$$p(y = c|\mathbf{x}, \mathbf{X}, \mathbf{Y}) = \int p(y = c|\mathbf{x}, \mathbf{W})p(\mathbf{W}|\mathbf{X}, \mathbf{Y})\mathrm{d}\mathbf{W} \qquad (2.1)$$

where $\mathbf{x}$ denotes the input image, y is the output label of a pixel, c is the label class, $\mathbf{X}$ and $\mathbf{Y}$ are sets of images and labels in the training dataset, respectively, and $p(\mathbf{W}|\mathbf{X}, \mathbf{Y})$ is the posterior distribution. Gal and Ghahramani (2016) proved that approximation of the posterior distribution is equivalent to the dropout masked distribution $q(\mathbf{W})$, where $\mathbf{W} = \mathbf{M}\cdot\text{diag}(\mathbf{z})$ and $\mathbf{z} \sim \text{Bernoulli}(\theta)$, where $\mathbf{M}$ denotes the variational parameters, θ is the dropout ratio. Then, Eq.(2.1) can be approximated by minimizing the Kullback-Leibler (KL) divergence $\text{KL}(q(\mathbf{W})||p(\mathbf{W}|\mathbf{X}, \mathbf{Y}))$ as follows.

$$p(y = c|\mathbf{x}, \mathbf{X}, \mathbf{Y}) \approx \int p(y = c|\mathbf{x}, \mathbf{W})q(\mathbf{W})\mathrm{d}\mathbf{W} \qquad (2.2)$$

$$\approx \frac{1}{T}\sum_{t=1}^{T}\text{Softmax}(\mathbf{f}(\mathbf{x}, \mathbf{W})). \qquad (2.3)$$

where T is the number of dropout samplings. This Monte Carlo estimation is called "*MC dropout*" (Gal and Ghahramani (2016)). We employed the predictive

variance as the metric indicating uncertainty which is defined as

$$
\begin{aligned}
&Var(y = c|\mathbf{x}, \mathbf{X}, \mathbf{Y}) \\
&\approx \frac{1}{T} \sum_{t=1}^{T} \text{Softmax}(\mathbf{f}(\mathbf{x}, \mathbf{W}))^T \text{Softmax}(\mathbf{f}(\mathbf{x}, \mathbf{W})) \\
&- p(y|\mathbf{x}, \mathbf{X}, \mathbf{Y})^T p(y|\mathbf{x}, \mathbf{X}, \mathbf{Y}).
\end{aligned} \tag{2.4}
$$

In this paper, we propose two structure-wise uncertainty metrics; 1) predictive structure-wise variance (PSV) and 2) predictive Dice coefficient (PDC). PSV represents the predictive variance per unit area of the pixels that are classified as the target structure. The metric is defined as

$$
PSV(s = c|\mathbf{x}) = \frac{1}{|Z|} \sum_{i \in Z} \sum_{k} Var_i(y = k|\mathbf{x}) \tag{2.5}
$$

where $Z = \{i| \arg\max_{k} p(y_i = k) = c\}$, indicating all pixels that are classified as class c ($argmax$ represents selection of the class with the highest probability for the pixel i). PDC is computed by linear regression of PSV and actual Dice coefficient of the target structure.

$$
PDC(s = c|\mathbf{x}) = \alpha \cdot PSV(s = c|\mathbf{x}) + \beta \tag{2.6}
$$

where α is the linear coefficient and β is the bias.

As for the network architecture, we extend the U-Net model by inserting the dropout layer before each max pooling layer and after each up-convolution layer as shown in Fig. 2.1(a), which is the same approach as Bayesian SegNet proposed by Kendall et al. (2015). We call the U-Net extended by *MC dropout "Bayesian U-Net"*.

2.2.4 Bayesian active-learning

A common practical situation in segmentation problems entails a scenario where there is a small-scale labeled dataset and a large-scale unlabeled dataset. The

active-learning method is a way known to be effective in that scenario by interactively expanding the training dataset using the experts' input.

In order to determine the pixels to query to the experts, the proposed method first selects uncertain slices in the unlabeled dataset, which we call slice selection step, and then select uncertain pixels in the selected slices, which we call pixel selection step. The slice selection step follows Yang et al. (2017) that utilized uncertainty and similarity metric to determine the query slices, which is summarized as follows: Letting $\mathcal{D}_u$ be unlabeled dataset, a subset of uncertain slices $\mathcal{D}_c \subseteq \mathcal{D}_u$ is selected. Then, a subset of representative slices $\mathcal{D}_r \subseteq \mathcal{D}_c$ is selected, which is considered to be representative of the distribution of the image features of all unlabeled dataset $\mathcal{D}_u$. Details of the algorithm is formulated in Algorithm 1.

In this paper, we propose a method for reducing the number of pixels to query, that is, to be manually labeled in the selected slices, using the uncertainty. The label for the j-th queried pixel in i-th slice, $\hat{Y}_{ij}$, is given as

$$\hat{Y}_{ij} = \begin{cases} \arg\max\limits_{k} \ p(y = k | \mathbf{x}) & (\sum\limits_{k} Var_{ij}(y = k) < T) \\ Y_{ij}^{manual} & (otherwise) \end{cases} \tag{2.7}$$

where Y_{ij}^{manual} denotes the manually provided label. T is a criterion used to distinguish the confident and the uncertain pixels. In our experiments, T was empirically determined as 2.5×10^{-3}.

2.2.5 Implementation details

As pre-processing, intensity of the CT volumes is normalized such that $[-150, 350]$ HU is mapped to $[0, 255]$ (smaller than -150 HU and larger than 350 HU are clamped to 0 and 255, respectively). At the training phase, data augmentation is performed to artificially generate variations of the posture of the patient by translation of $[-25, +25]$ % of the matrix size, rotation of $[-10, +10]$ deg, scale of $[-35, +35]$ %, shear transform with the shear angle of $[-\pi/8, +\pi/8]$ rad, and

Algorithm 1 Slice selection

Require: unlabeled dataset $\mathcal{D}_u$; uncertain slices $\mathcal{D}_c \subseteq \mathcal{D}_u$; representative slices
$\mathcal{D}_r = \emptyset$

1: **while** $|\mathcal{D}_r| < N$ **do**
2:      $l, i_{best} \leftarrow 0, 0$
3:      **for** $i \leftarrow 1$ to $|D_c|$ **do**
4:          $\hat{\mathcal{D}}_r, m \leftarrow \mathcal{D}_r \cup \{I_{c,i}\}, 0$
5:          **for** $j \leftarrow 1$ to $|\mathcal{D}_u|$ **do**
6:              $n \leftarrow 0$
7:              **for** $k \leftarrow 1$ to $|\hat{\mathcal{D}}_r|$ **do**
8:                  $s \leftarrow similarity\ (I_{u,j},\ \hat{I}_{r,k})$
9:                  **if** $s > n$ **then**
10:                      $n \leftarrow s$
11:                  **end if**
12:              **end for**
13:              $m \leftarrow m + n$
14:          **end for**
15:          **if** $l > m$ **then**
16:              $l, i_{best} \leftarrow m, i$
17:          **end if**
18:      **end for**
19:      $\mathcal{D}_r \leftarrow \mathcal{D}_r \cup \{I_{c,i_{best}}\}$
20: **end while**

horizontal flipping. These augmentations allow the model to be invariant to FOV of the scan, patient's size, rotation, and translation. As post-processing, the largest connected component is extracted to obtain the final output for each muscle.

2.2.6 Comparison with conventional method

We have implemented the current state-of-the-art method for automated segmentation of individual muscles from CT based on the hierarchical multi-atlas label fusion Yokota et al. (2018) using the same parameters as the original paper and compared the results with our CNN-based method. In addition to the U-Net, as a reference, we tested another network architecture, FCN-8s (Long et al. (2015)), which is also a common fully convolutional network based on VGG16.

We used Dice coefficient (DC) (Dice (1945)) and absolute symmetric surface distance (ASD) (Styner et al. (2008)) as error metrics. The statistical significance was tested by paired t-test with Bonferroni correction.

2.3 Results

2.3.1 Network architecture selection and comparison with conventional method

To demonstrate the applicability of deep neural networks, we quantitatively evaluated the segmentation accuracy using 20 labeled clinical CTs. Leave-one-out cross-validation (LOOCV) was performed where each network was trained with 19 CTs, and tested with the remaining one. Twenty-three class classifications (three bones, 19 muscles and background) were performed. We initialized the weights in the same way as in He et al. (2015), and then trained the networks using Adaptive moment estimation (Adam) (Kingma and Ba (2014)) for 1×10^5 iterations at the learning rate of 0.0001. The batch-size was 3.
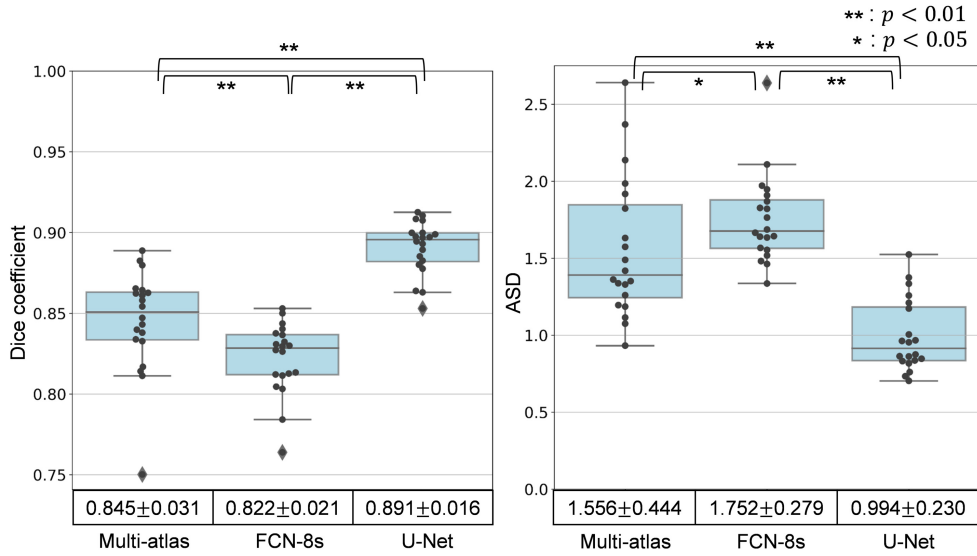
Figure 2.3: Accuracy of muscle segmentation for 20 patients with conventional multi-atlas method, FCN-8s and U-Net. Box and whisker plots for two error metrics. (a) Dice coefficient (DC) and (b) average asymmetric surface distance (ASD). Boxes denote the 1st/3rd quartiles, the median is marked with the horizontal line within each box, and outliers are marked with diamond. The accuracy of 19 muscles were averaged in advance (i.e., 20 data points for each box plot)

Fig. 2.3 summarizes the segmentation accuracy of the muscles. The DC and ASD over 19 muscles were averaged and plotted as box plots (i.e., 20 data points in each plot) for the multi-atlas method, FCN-8s and U-Net. The average and standard deviation over 20 patients were 0.845 ± 0.031, 1.556 ± 0.444 mm (mean$\pm$std), 0.822 ± 0.021, 1.752 ± 0.279 mm and 0.891 ± 0.016, 0.994 ± 0.230 mm, respectively. Compared with the conventional multi-atlas method (Yokota et al. (2018)) and FCN-8s, U-Net yielded statistically significant improvement ($p < 0.01$) in both DC and ASD.

Fig. 2.4 shows the heatmap visualization of ASD for individual muscles of each patient with multi-atlas method and U-Net. The blue color shows the lower ASD. The accuracy improvement is clearly observed for almost all the muscles
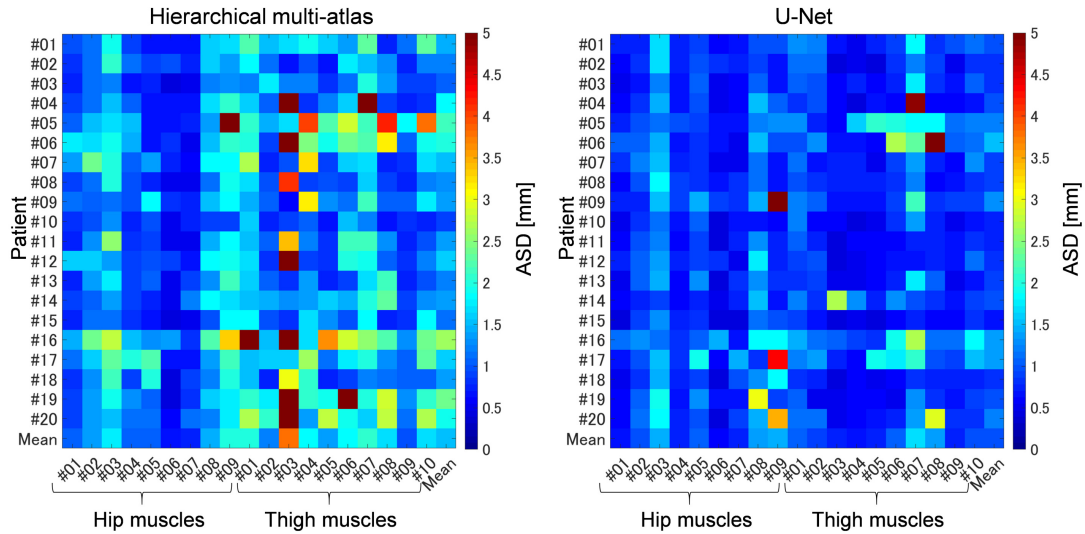
Figure 2.4: Heatmap visualization of ASD for individual muscle structures with respect to each patient with multi-atlas method and U-Net. The blue color shows higher segmentation accuracy.

except for the psoas major in patient #09 and #17, gracilis in patient #14, semimembranosus in patient #04, and semimembranosus in patient #06. Fig. 2.5 shows example 3D and 2D visualizations of the predicted label for a representative patient. The result with U-Net demonstrates more accurate segmentation near the boundary of muscle structures in comparison with Multi-atlas methods. In FCN-8s, since the output layer is obtained by upsampling and fusing the latent vectors have the resolution under one eighth of the input size, the accuracy seemed to be consistently lower than U-Net. On the other hand, in U-Net, since the output layer is directly fused with the latent vectors which have the resolution same as the input size, the input image and output label could have pixel-wise correspondence.

DC and ASD of bone structures for FCN-8s and U-Net were 0.928 ± 0.026, 0.759 ± 0.368 mm, and 0.976 ± 0.015, 0.224 ± 0.127 mm, respectively. (Note that Yokota et al. (2018) did not use multi-atlas method for bones, thus we did not compare the accuracy of bones with its.)

The average training time was approximately 11 hours with FCN-8s and U-Net, on Intel Xeon processor (2.8 GHz, 4 cores) and NVIDIA GeForce GTX 1080Ti. The average computation time for the two stage segmentation of one CT volume (about 500 2D slices) was approximately 2 minutes excluding file loading, and post-processing was 3 minutes.

We conducted the following experiments about the predictive accuracy and active learning only with U-Net architecture, since its accuracy is significantly higher than the other two methods as shown above.
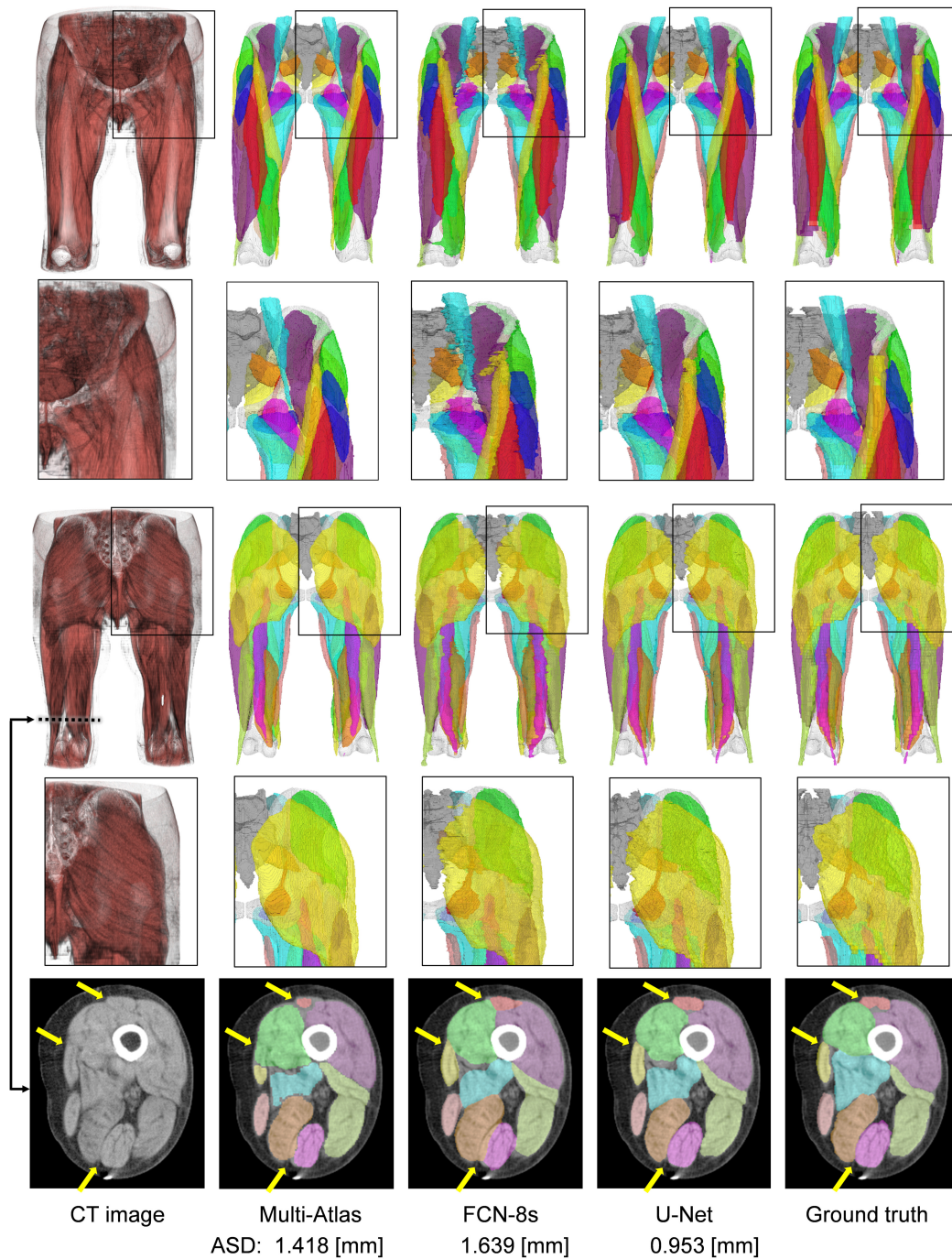
Figure 2.5: Visualization of estimated label for a representative patient (#01). The result with U-Net has distinctly smoother segmentation near the boundary of muscle structures. The region of interest is illustrated as a black line in the lower left 3D visualization.

23

2.3.2 Estimation of uncertainty metric

Relationship between uncertainty and segmentation accuracy

To demonstrate validity of the uncertainty metric, we investigated relationship between the estimated uncertainty and error metrics with the 20 labeled CTs. We performed 4-fold cross-validation where Bayesian U-Net was trained with randomly selected 15 CTs, and tested with the remaining 5 CTs using the same condition as the experiment above.

Fig. 2.6 shows a scatter plot of DC for the individual muscle structure as a function of its PSV. The 95% confidence ellipses clearly show the increased error (i.e., decreased DC) in accordance with the increased PSV. The correlation ratio was more than 0.5 for all muscles except for the obturator internus. Fig. 2.7 shows example uncertainty visualization with Bayesian U-Net. The high correlation between the two metrics suggested validity of using the uncertainty metric estimated from Bayesian U-Net as an indicator of the unobservable error metric in a real clinical situation.

Generalization ability to TCIA dataset

To demonstrate the generalization ability of Bayesian U-Net, we tested it with TCIA dataset. Figure 2.8(a) shows a scatter plot of DC as a function of PDC. Note that α and β in Eq. (2.6) were determined by linear regression of 20 data points of THA dataset. The mean absolute error between DC and PDC was 0.011±0.0084. Figure 2.8(b) shows representative results for two patients. Figure 2.8(c) and (d) show two patients with lower dice. We observed higher uncertainty in tumor regions where the segmentation error was large.

Quantitative evaluation was performed in the gluteus medius muscle. The average DC and ASD from 18 patients were 0.914±0.026, 2.927±4.997 mm, respectively, and their uncertainty was plotted in Fig. 2.6(b). When excluding four outlier patients, the average of DC and ASD was 0.925±0.014, 1.135±0.777 mm, which was comparable to the results on the Osaka university hospital THA
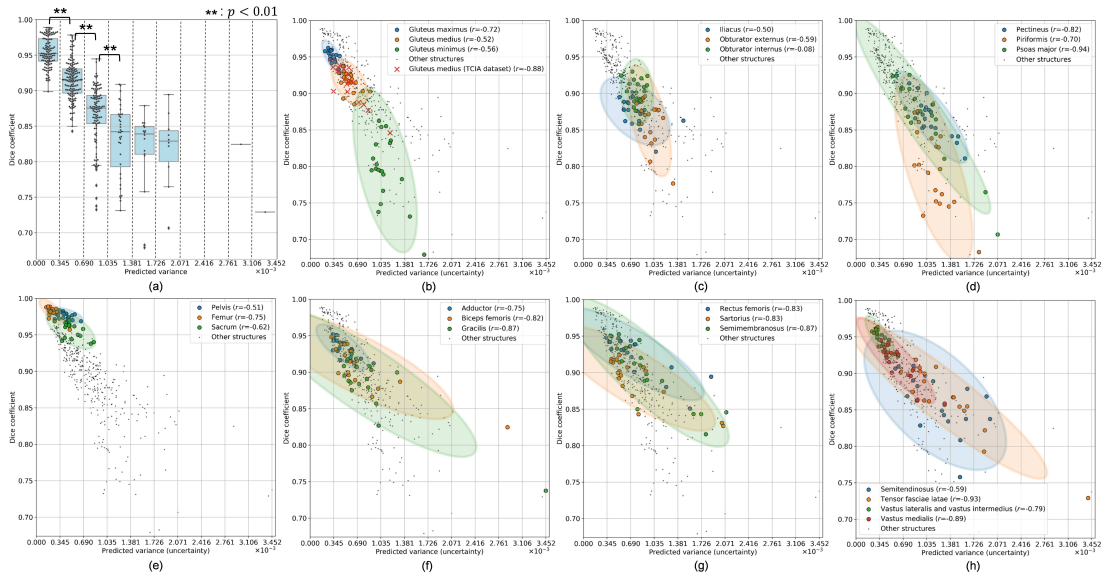
Figure 2.6: (a) Box and whisker plots of DC as a function of predictive structure-wise variance (PSV). We divided PSV into 10 bins with an equal width. Mann–Whitney U test was performed in adjacent bins. (b-h) The scatter plot, with 95 % confidence ellipses, of DC for individual muscle structures as a function of PSV. (b-d), (e) and (f-h) show the bone, hip and thigh muscle structures, respectively. The symbol "r" denotes Pearson's correlation coefficient.

dataset. The uncertainty and DC were plotted in Fig. 2.6(b), which resulted in a distribution similar to the gluteus medius of THA dataset suggesting generalization ability of the uncertainty metric between different datasets.

2.3.3 Bayesian active-learning

To investigate one of the application scenarios of the uncertainty estimates, we tested an active-learning method in a simulated environment using the 20 fully labeled clinical CTs. The experiment assumed that 15 CTs consisting of 95 % of unlabeled slices and 5 % of labeled slices are available. Then, from each CT, 5 % of the total number of slices from the unlabeled slices was added to the labeled data in one step, which we call one "*acquisition step*". We iterate the
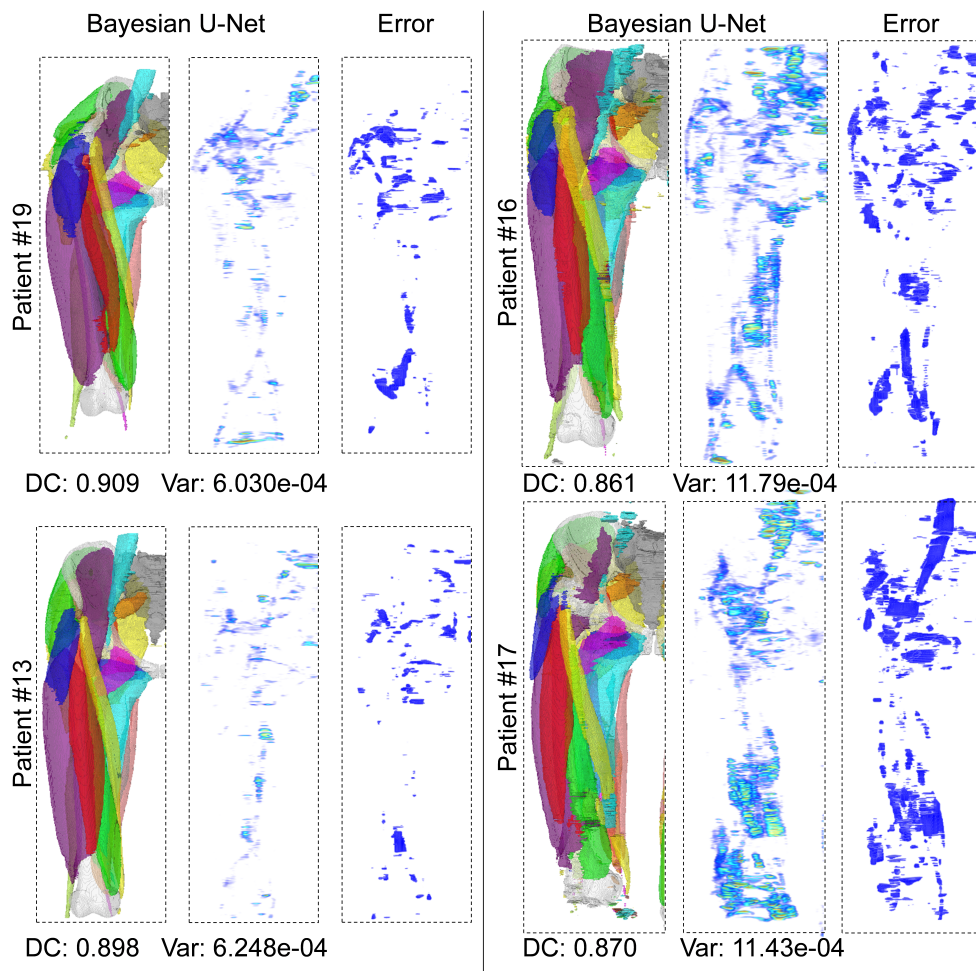
Figure 2.7: Visualization of predictive variance computed by Bayesian U-Net. The average Dice coefficient and predictive structure-wise variance (PSV) of musculoskeletal structures are denoted as DC and Var. A good agreement between the uncertain region and error region suggests validity of the uncertainty metric to predict unobservable error in a real clinical situation.

acquisition step 20 times. The remaining 5 CTs were used as the test dataset. In each acquisition step, Bayesian U-Net was initialized and trained using Adam (Kingma and Ba (2014)) for maximal 300 epochs at the learning rate of 0.0001 with the early stopping schema. Note that each axial CT slice was downsampled
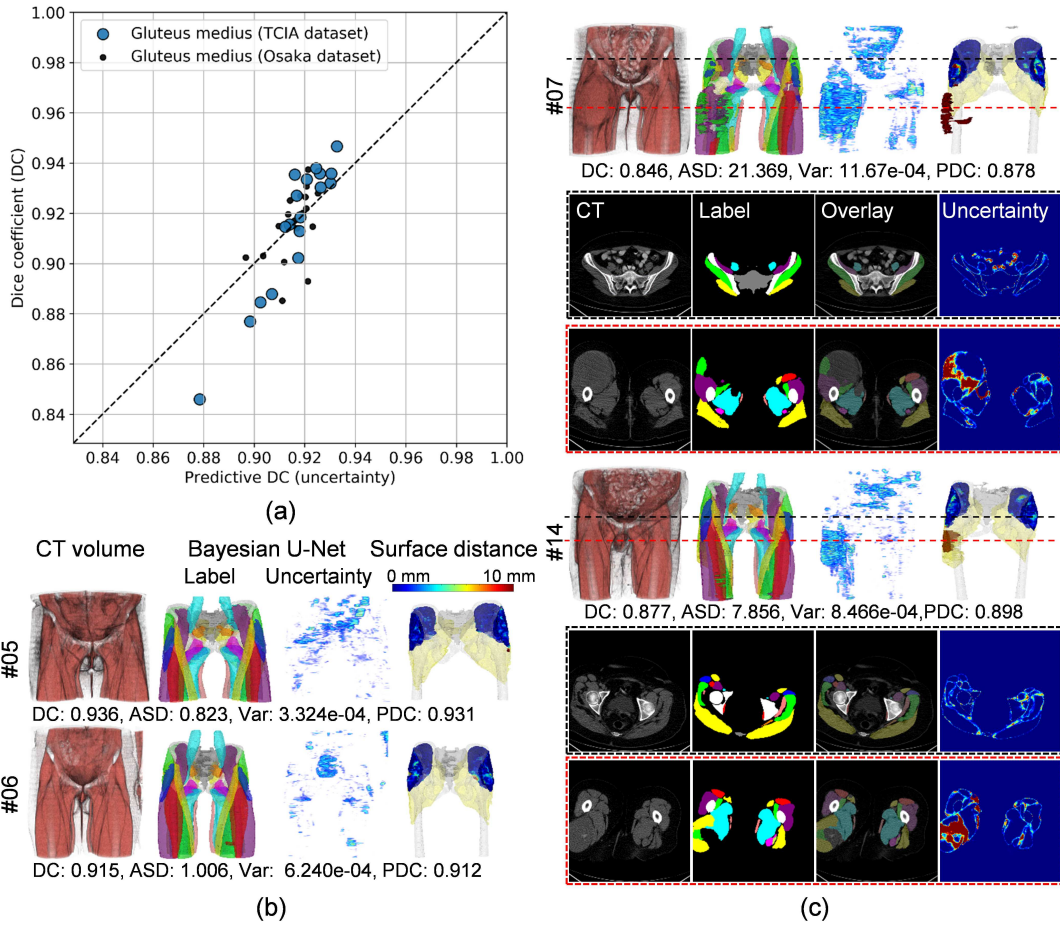
Figure 2.8: Visualization of generalization ability of Bayesian U-Net on the TCIA soft-tissue sarcoma dataset. (a) Scatter plot of DC as a function of predictive Dice coefficient (PDC), (b) Representative results for two patients, (c) Two patients with failed segmentation. We observed higher uncertainty in tumor regions in the patient #07 and #14 where segmentation error was large. In each patient, from the left to the right, the input CT volume, the predicted label and uncertainty, and the surface distance error of gluteus medius muscle. The predictive structure-wise variance of gluteus medius muscle and PDC are reported, respectively. In the gluteus medius muscle, the average of DC and ASD from 18 patients was 0.914±0.026 and 2.927±4.997 mm, respectively.

to 256×256 due to the limitation of training time in this experiment. The data augmentation was not performed in order to investigate behavior of the model purely depending on the number of training dataset.

In order to quantitatively evaluate the advantage of the semi-supervised approach, we defined a new metric that we call manual-annotation cost (MAC) as

$$\mathrm{MAC}(X, Y) = \frac{|Y^{manual}|}{|X \neq b|} \tag{2.8}$$

where X and Y are ground-truth and label added by the proposed method. $|Y^{manual}|$ denotes the number of manually annotated pixels in Y. The symbol b denotes the background region.

Uncertain slices were first selected from the unlabeled dataset, which follows the pixel selection step. We compared the segmentation accuracy at each acquisition step with the following three pixel selection methods. (1) Fully-manual selection (Yang et al. (2017)): The user annotates all pixels in the uncertain slices. (2) Random selection: The user annotates random pixels in the uncertain slices. (3) Semi-automatic selection (proposed method): The user annotates only uncertain pixels in the uncertain slices. In order for a fair comparison, we set the experimental condition so that the number of pixels annotated in (2) and (3) were equal.

Fig. 2.9(a) shows mean DC over all muscles and patients as a function of acquisition step (note that each acquisition step adds 5% of the total training dataset resulting in 100% after 20 steps). First, we observed a trend that the accuracy increases as training dataset increases for any selection method. The random selection method stopped the increase at around DC from 0.840 to 0.845, while the other two methods kept increasing. DC of the proposed method reached higher than random selection by about 0.03, which was close to the fully-manual selection method. Fig. 2.9(b) shows the box and whisker plots of MAC metric as a function of acquisition step. We observed that as training dataset increases, MAC value decreases, gradually. Compared with fully-manual selection method,
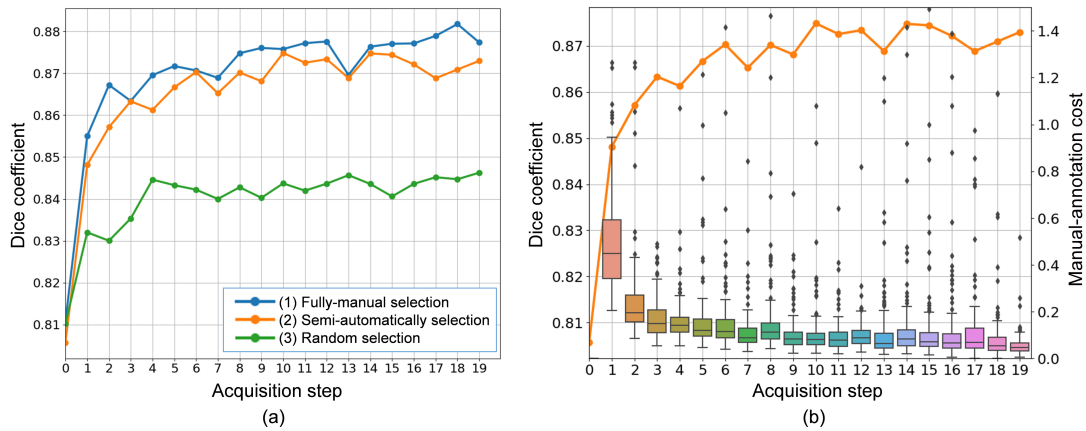
Figure 2.9: Results of proposed active-learning method. (a) The plot of mean of DC over individual structures and patients as a function of each acquisition step for different pixel selection. (b) The box and whisker plots of manual-annotation cost over as a function of each acquisition step.

the proposed method reduced about 55% of manual-annotation in the first step, about > 70% of that after third steps, in this experiment. Fig. 2.10 shows examples of manually annotated regions and its MAC value.

2.4 Discussion

In this work, we presented an application of deep neural network for automated segmentation of individual muscles from CT. The trained Bayesian U-Net takes only 5 minutes for segmentation of one CT volume, including about 500 slices, on a standard desktop workstation with one GPU, which would allows us the patient-specific musculoskeletal modeling in everyday clinical routine.

Our target application mainly focuses on personalization of the biomechanical simulation. Volumetric muscle modeling has been showing advantage for accurate prediction of muscle behavior using finite element modeling (Webb et al. (2014); Yamamura et al. (2014)) or a simpler approximation in shape deformation for real-time applications such as Murai et al. (2016). In addition, Otake

CT slice	Ground-truth	Query	CT slice	Ground-truth	Query

MAC : 0.501   MAC : 0.475

MAC : 0.252   MAC : 0.302

MAC : 0.156   MAC : 0.103
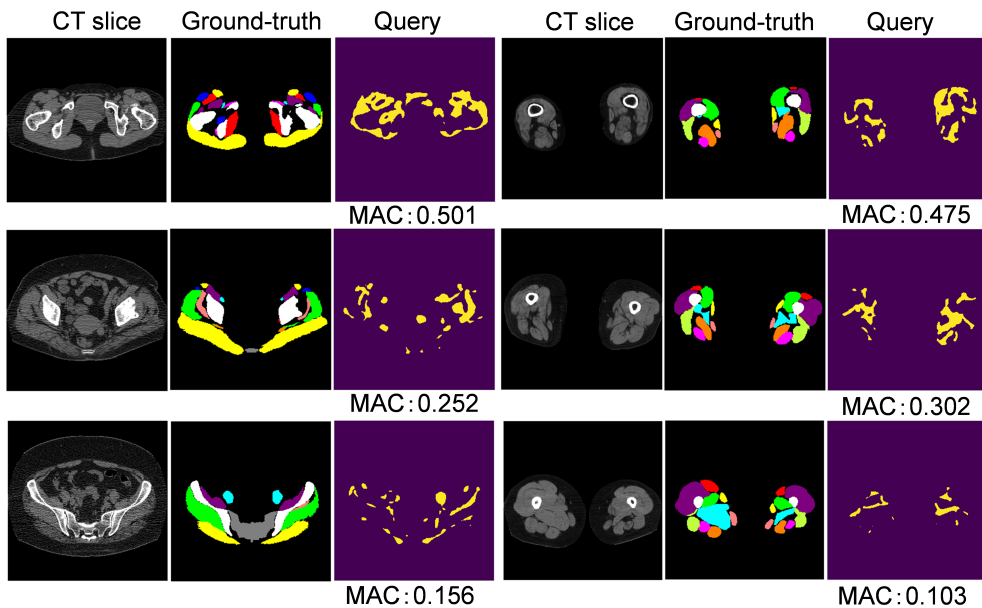
Figure 2.10: Examples of queries which denotes manually annotated pixel (colored by yellow) and its manual-annotation cost (MAC).

et al. (2018) demonstrated the potential for estimating patient-specific muscle fiber structure from a clinical CT assuming the segmentation of each muscle was provided. The proposed fast and automated segmentation method would enhance application of these accurate biomechanical modeling in a clinical routine as well as in studies using large-scale CT database for applications such as statistical analysis of human biomechanics for ergonomic designs.

As for the uncertainty metric, we found a high correlation with Dice coefficient, and visually observed in TCIA dataset that the tumor exhibits higher uncertainty and mis-segmentation actually happened in that region. These results suggest that the Bayesian U-Net trained by CTs of patients without tumor can also be used to detect anomaly such as the tumor. One muscle with low correlation ($r = 0.08$) was the obturator internus, which is a small muscle connecting internal surface of obturator membrane of the pelvis and medial surface of greater trochanter of the femur and traveling almost in parallel to the axial

plane (see Fig. 2.2). We believe these properties entailed a challenge in manual tracing and produced variability in the ground truth (so-called *aleatory variability*) which disturbed the measurement of the uncertainty due to limited data (so-called *epistemic* uncertainty) represented by the predictive variance in the MC dropout estimates.

The patient-specific anatomy of skeletal muscles has also been studied in clinical diagnosis and monitoring of muscle atrophy or muscle fatty degeneration caused by or associated with trauma, aging, disuse, malnutrition, diabetes and so on (Rasch et al. (2009); Uemura et al. (2016)) using MRI or CT. In those studies, muscles are delineated manually by single operator from the image either in 2D (Werf et al. (2018)) or in 3D (Handsfield et al. (2014)). The proposed 2D slice-based automated muscle segmentation is helpful in order to reduce the manual labor and inter-operator variability in these analyses. The advantage of 2D slice-based approach without using information of neighboring slices includes flexibility against variation in slice interval and FOV of the CT scan protocol.

While MRI has advantages in terms of absence of radiation exposure and superior soft tissue contrast, we targeted CT in this study because the scanning of CT is much faster. In both modalities, the scan time is determined by the trade-off between spatial resolution and the craniocaudal extent that can be acquired in a clinically acceptable time without causing motion artifact. The CT scanning protocol that we use for the lower extremity takes less than 30 seconds, while a typical MRI scan of the same range with the same spatial resolution would require more than 10 minutes. The fast scan is especially advantageous in orthopedic surgery, where the biomechanical simulation is most helpful, to obtain the entire muscle shapes from their origin to insertion in the thigh region. Furthermore, CT intensity (Hounsfield Unit) is a more stable quantitative measure than the intensity of MRI which is susceptive to variation of the scan protocol, resulting in inaccuracy in evaluation of muscle quality. Nonetheless, application of the proposed method to MR images would also be achievable, for example, by using an image synthesis algorithm such as CycleGAN (Hiasa et al. (2018b);

Zhang et al. (2017)) for synthesizing a CT-like image from the MR image.

One limitation in this study is the limited variation in the training and test dataset. THA dataset that we used contains only female who are subject to THA surgery, which limits the variation in size and fat content in muscles. Although TCIA dataset contains male dataset and larger variation in terms of pathology, the ground truth label is available only for the gluteus medius muscle due to limitation of manual trace labor. Evaluation with a large-scale fully annotated database is in our future work. Another limitation, specifically in the active-learning method, is computation time for multiple forward and backward propagation to update weights. Thus, it would not be suitable for application in real-time interactive annotation method such as Maier-Hein et al. (2016). The lack of large-scale database also limited the evaluation of the active-leaning method. Further investigation with a larger unlabeled CT database would be required to evaluate effectiveness of the proposed method in a more realistic clinical scenario.

Eaton-Rosen et al. (2018) investigated uncertainty metrics in binary segmentation of brain tumor, specifically for quantifying the uncertainty in volume measurement. Nair et al. (2018) also explored uncertainty in binary segmentation for lesion detection in multiple sclerosis. Our present work is distinct from these previous works in that we targeted a multi-class organ segmentation problem, demonstrated correlation between the uncertainty metric, namely MC sample variance, and dice coefficient of each class, and illustrated locality of uncertain pixels correlate with incorrectly segmented pixels as well as anatomical abnormality such as sarcoma in muscles. We also presented its potential use case in a simulated active learning framework for selection of query pixels, and demonstrated that the uncertainty-based selection saved 70% of manual labeling labor while keeping the accuracy comparable to the conventional all-pixel-annotation scheme.

2.5 Conclusion

We have demonstrated performance of 2D U-Net for segmentation of 19 muscles in lower extremity. Slice-by-slice (2D) segmentation without using information of neighboring slices allows flexibility in scan protocol (namely, slice interval, FOV), for example, it is amenable to images acquired with relatively old CT scans where the slice interval was large. Leave-one-out cross validation showed average DC of 0.891±0.016, ASD of 0.994±0.230 mm, which are significant improvement ($p < 0.01$) compared with the state-of-the-art hierarchical multi-atlas method.

3 Recovery of three-dimensional rib motion from dynamic chest radiograph

Dynamic chest radiography (2D x-ray video) is a low-dose and cost-effective functional imaging method with high temporal resolution. While the analysis of rib-cage motion has been shown to be effective for evaluating respiratory function, it has been limited to 2D. In this chapter, we aim at 3D rib-motion analysis for high temporal resolution while keeping the radiation dose at a level comparable to conventional examination.

3.1 Related work

The rib cage, lungs, and thoracic muscles cooperatively contribute to the pulmonary function. Since the rib cage mobility is a key factor in the respiratory motion partly in relation with spine deformity, analysis of the rib cage motion has received considerable attention and its importance was pointed out especially on the patients with chronic obstructive pulmonary disease (COPD) (Gilmartin and Gibson (1986)), kyphosis (Culham et al. (1994)), scoliosis (Tanaka et al. (2015)), and so on. Clinically available systems for accurate rib cage motion analysis are required for diagnosis of these patients. In the biomechanics research field, three-dimensional (3D) modeling of the rib cage and muscles has been addressed to analyze the rib cage and muscular biomechanics (Didier et al. (2009); Bruno

et al. (2015)).

Dynamic chest radiography based on an x-ray video system with a flat panel detector is expected to be a low-dose and cost-effective functional imaging method for evaluating pulmonary function (Tanaka (2016)). A recent report (Tanaka et al. (2015)) demonstrated the usefulness of rib-motion analysis using two-dimensional (2D) optical flows in the x-ray video to evaluate the respiratory function of scoliotic patients. Although time-varying information was effectively used, its analysis was limited to 2D.

The 3D rib motion was previously analyzed using multi-time-phase CT data acquired at inhale and exhale phases (Ito et al. (2011)). However, acquisition of CT at a larger number of time phases to increase temporal resolution requires a larger radiation dose, which limits its applicability in a routine clinical setting. In a broader context, 3D knee-motion analysis at higher temporal resolution was conducted by combining 3D MRI and x-ray video (Chen et al. (2014)), where the patient-specific rotational axis of the articular motion was estimated from multiple 3D MRI data of different knee-joint postures to ensure the robustness of 2D-3D registration of the x-ray video and MRI-derived bone models. The necessity of acquiring multiple 3D MRI for estimating the rotational axis is a drawback. Due to its long scan time and artifacts caused by patient motion, this approach is not applicable to anatomies that require large field-of-view (FOV) and move continuously such as the chest region. Computed tomography is commonly acquired for the chest domain in a routine clinical setting; therefore, to avoid an increase in radiation exposure, we attempt to use the single-time-phase CT data.

Static 3D rib analysis based on 2D-3D registration and reconstruction was also addressed in previous work (Dworzak et al. (2010); Grenier et al. (2013)). These studies assumed manual extraction and anatomical labeling of the contours or the center lines of the ribs in the x-ray images. Applying these techniques to each frame of the x-ray video would be time-consuming and labor task due to manually extraction.

LCN has been paid more attention by the human/computer vision and neural

network communities (Heeger (1992); Jarrett et al. (2009)) rather than medical imaging community, and its effectiveness in face recognition application was investigated (Le (2013)). Although there are a couple of early successful medical applications (Veldkamp and Karssemeijer (2000); Fleming et al. (2006)). Chen et al. (2013) carried out a pre-processing similar to the LCN in 2D-3D registration of carpal bones; however, the detailed formulation of the pre-processing was not given and its effect on the registration results was not investigated.

In the context of evaluation of 2D-3D registration, the quantitative evaluation in experiments using real images is generally not straightforward due to lack of the ground truth. Chen et al. (2014) qualitatively confirmed that the motion pattern of the knee joint estimated from real-image reasonably follows the well-known kinematics of the human knee joint. Villard et al. (2014) evaluated the estimated rib motion indirectly by computing the synchronicity of the motion pattern with diaphragm movement. In this paper, we quantitatively evaluated the rib motion in two steps: validity of the error metric, distance in the projection plane (mPCD), was confirmed by determining its correlation with the actual rotation angle error (Fig. 3.6(c)), and then by using the validated error metric, we evaluated the method on the real images using manually traced contours as the ground truth.

3.2 Methods

3.2.1 Overview

Figure 3.1 shows the workflow of the proposed method. Local contrast normalization is first applied to the x-ray video, which is used as an input of 2D-3D registration. The algorithm underlying the proposed 2D-3D registration method follows that of Otake et al. (2012, 2015) optimize rigid transformation parameters to maximize the similarity measure between the digitally reconstructed radiograph (DRR) and acquired actual radiograph. We enhanced the algorithm to estimate relative poses of multiple rigid objects connected by joints such as the
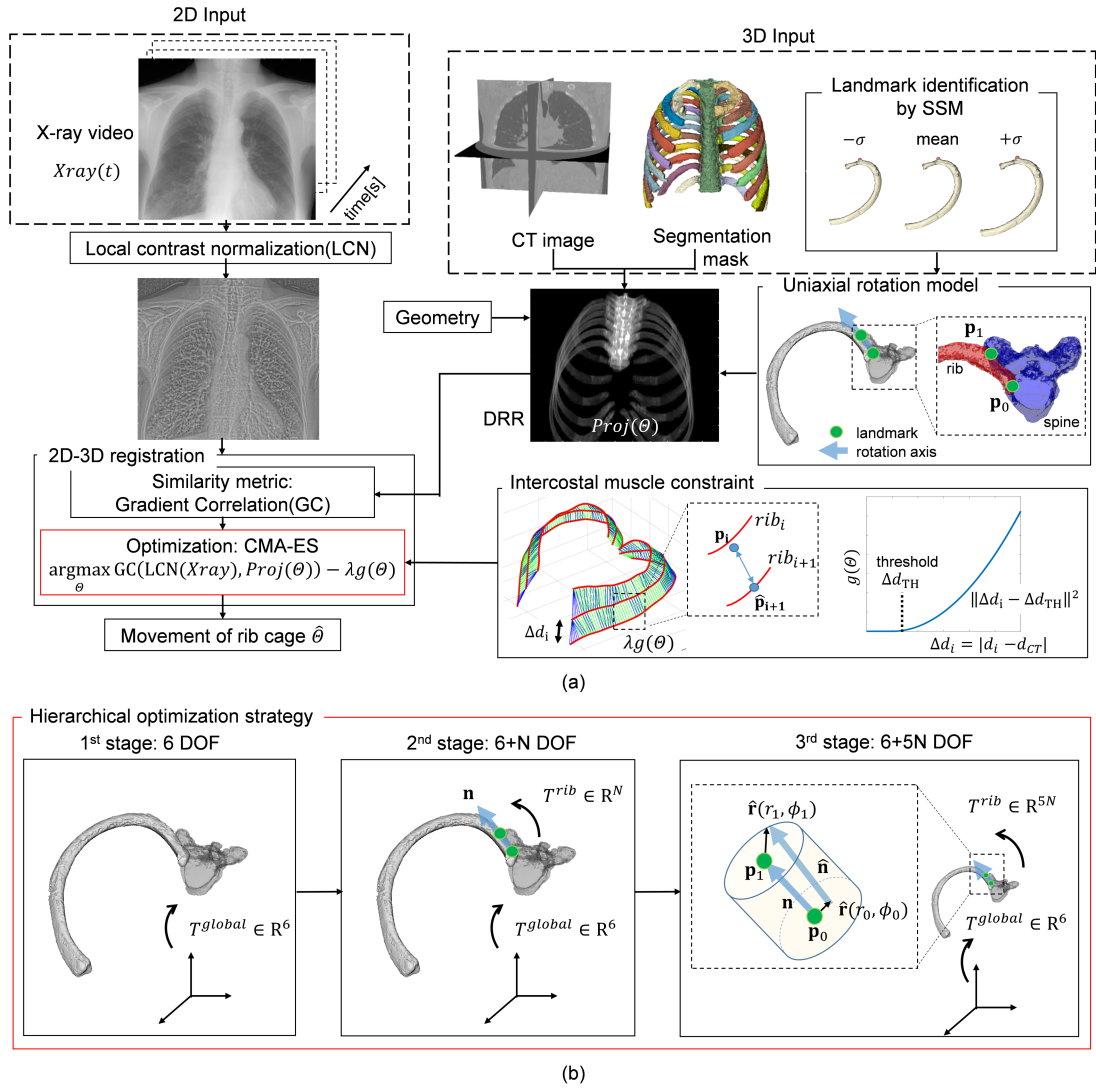
Figure 3.1: (a) Workflow of the proposed method. The optimization parameter Θ represents the rigid transformation of the spine, and rotation axis and rotation angle of N ribs (note the rotation axis parameters are not optimized in the first stage). $g(\Theta)$ is a regularization term that penalizes the cost according to the distance between the ribs which acts like the intercostal muscles and facilitates a robust and anatomically feasible estimation. (b) Hierarchical optimization strategy in the proposed method. The first stage optimizes the global rigid transformation of the rib cage. The second stage jointly optimizes the local rigid transformation of the rib-rotation angle around the rotation axis $\mathbf{n}$ for each rib. The final stage jointly optimizes the rotation axis $\mathbf{n}$ within a small cylindrical region.

rib cage. First, we segment each bone in the CT data. Then, using a statistical shape model (SSM), we automatically identify joints on each rib bone to define an initial estimate of the uniaxial rotation axis, which is described in detail in Section 3.2.4. We formulate the recovery of the rib motion as a problem of estimating a six-degree-of-freedom (DoF) global transformation of the spine and local transformations of each rib with respect to the spine. The motion of each rib is modeled with a uniaxial rotation joint. In the optimization, we introduce a constraint, which takes into account the fact that the adjacent ribs are connected by the intercostal muscle.

3.2.2 Preprocessing of CT

Segmentation of rib and spine in CT

We semi-automatically segment each rib bone and the spine in CT data using commercial interactive 3D-image-segmentation software, Synapse Vincent (Fujifilm, Tokyo, Japan). For the segmentation of the rib, we use the *bone separation module*, which requires the operator to select seed points interactively in each rib bone. For the segmentation of the spine, we used the *spine extraction module*, which allows us to extract spine semi-automatically. In our experiments, it took about 20 min for segmentation of the 2nd to 4th ribs and vertebrae of one case.

Estimation of rib landmarks using statistical shape model

The costotransverse and costovertebral joints on each rib, which define the initial estimate of the rotation axis, are automatically estimated using a statistical shape model (SSM). The workflow of the estimation process is illustrated in Fig. 3.2. The SSM of each rib, i.e., the average shape S_{ave} and variation modes, were computed using a method of SSM construction described in Yokota et al. (2013) (but hierarchization was not adapted). Then, the anatomical landmarks, i.e., the costovertebral and costotransverse joints, manually identified on the average shape
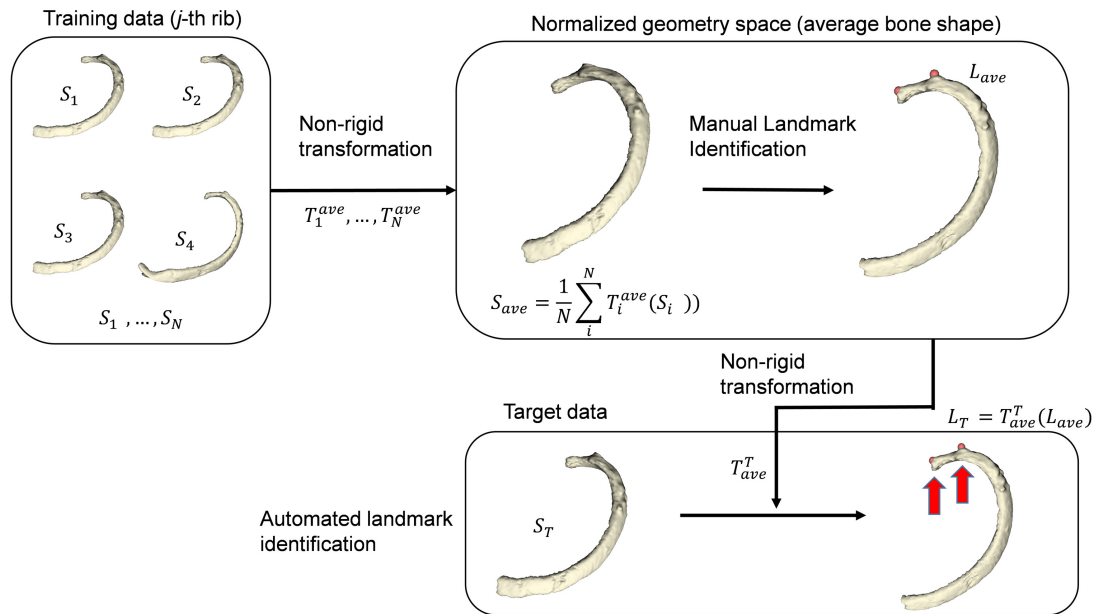
Figure 3.2: Workflow of computing the anatomical landmarks on the rib bone. The landmarks are manually identified on the average bone shape (L_{ave}) and automatically mapped to the target shape (L_T) by statistical shape model (SSM) fitting.

are automatically mapped on the target shape (L_T) by fitting the coefficients of the variation mode of the constructed SSM.

3.2.3 Preprocessing of radiography

Local contrast normalization is applied to the radiograph as a pre-process. It normalizes the intensity value within the neighborhood to zero mean and unit variance. Let x, y be vertical and horizontal indices of an interest pixel of input image $a \in \mathcal{R}^2$, and i, j be indices of the neighboring pixels. Now output image $\hat{a} \in \mathcal{R}^2$ is given by

$\hat{a}_{xy} = \frac{\bar{a}_{xy}}{\sigma_{xy}} = \frac{a_{xy} - \sum_{(i,j)} w_{ij} a_{i+x,j+y}}{\sqrt{\sum_{(i,j)} w_{ij} a_{i+x,j+y}^2}}$, where $\bar{a}_{xy}$ and σ_{xy} are the average and standard deviation of the pixels within the neighborhood window, respectively. The

weights w_{ij} are determined in such a way that their summation amounts to one ($\sum_{(i,j)} w_{ij} = 1$). In this study, all the weight values were uniform, i.e., all weights are $1/N$ where N is the number of pixels in the window, which is called the averaging filter. In the proposed method, LCN is used to enhance the contrast between the ribs and background soft-tissues such as the lungs and bronchia. In our experiments, the window size was empirically determined as 21 pixels. In the experiments using real images introduced in Section 3.3.2, we demonstrated how LCN improves the shape of the cost function and makes the optimization more robust.

3.2.4 Parameterization of rib motion

The rib cage consists of 12 pairs of ribs, which are connected to the thoracic vertebrae via joints. De Troyer et al. (2005) approximated rib motion using a uniaxial rotation about the axis connecting the costovertebral joint, which connects between the vertebral body and the head of rib, and the costotransverse joint, which connects between the transverse process and the tubercle of rib. The model was validated by Ito et al. (2011) by analyzing CTs at the inhale and exhale phases of one subject.

We formulate De Troyer's model as follows. Given the positions of the costovertebral joint ($\mathbf{p}_0$) and the costotransverse joint ($\mathbf{p}_1$) in the CT coordinate, the initial approximated rotation axis of the rib is defined by $\mathbf{n} = (\mathbf{p}_1 - \mathbf{p}_0)$. We denote a homogeneous transformation of the rib as matrix $rot(\mathbf{n}, \theta) \in \mathcal{R}^{4\times4}$, which represents a rotation θ around the axis $\mathbf{n}$. Note that the translation is not zero when line $\mathbf{n}$ does not pass through the origin. We parameterize $\mathbf{n}$ by using four parameters $(r_0, \phi_0, r_1, \phi_1)$, as shown in Fig. 3.1(b) (see the caption for the definitions of the four parameters). The homogeneous transform T^{rib} with respect to the spine $T^{global} \in \mathcal{R}^{4\times4}$ is defined as

$$T^{rib}(\theta; r_0, \phi_0, r_1, \phi_1) = rot(\hat{\mathbf{n}}, \theta) \cdot T^{global} \tag{3.1}$$

where,

$$\hat{\mathbf{n}} = \mathbf{n} + \hat{\mathbf{r}}(r_1, \phi_1) - \hat{\mathbf{r}}(r_0, \phi_0)$$
$$\hat{\mathbf{r}}(r, \phi) = r\{\cos \phi \ \mathbf{r} + \sin \phi (\mathbf{r} \times \mathbf{n})\}$$

and $\mathbf{r}^{\mathrm{T}}\mathbf{n} = 0$, and $|\mathbf{r}| = 1$. The T^{global} is represented with six parameters of translation and rotation $(t_x, t_y, t_z, r_x, r_y, r_z)$. The ZYX Euler angle is used to describe rotation. We considered multiple vertebrae as one rigid object in this study. The rationale of this assumption is discussed in Section 3.4.

3.2.5 DRR generation

The DRR-based method, which is also called the intensity-based method, is known to be more accurate and robust than the feature-based method, where the image features such as contours or landmark points are first extracted and the distances between these features are minimized, because the intensity-based method does not rely on the error-prone segmentation of the features and utilizes all the information in the images (Markelj et al. (2012)). The DRR is generated from a CT data using the ray-casting algorithm that computes line integrals of CT value along the passing ray.

In this paper, the x-ray imaging geometry is represented using the perspective projection model, as shown in Fig. 3.3. The origin of the detector-coordinate system is located at the center of the detector. The x and y axes are parallel to the detector coordinate, the z axis is parallel to their cross product, and the origin is at the position of the x-ray source. The location of the patient is (x, y, z) = (0,0, source-object distance (SOD)) and the distance from the x-ray source to the detector is denoted as the source-detector distance (SDD).

3.2.6 Similarity metric

The similarity metric used in the proposed 2D-3D registration method is gradient correlation (GC) (Penney et al. (1998)). GC is the normalized cross correlation
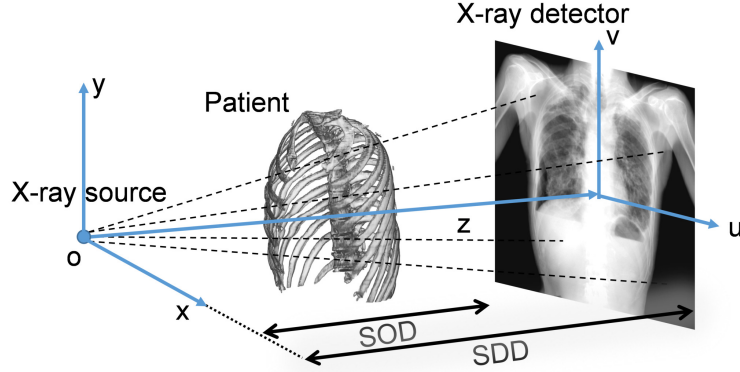
Figure 3.3: Definition of projection geometry used in this study. Parameters associated with x-ray source and computed tomography (CT) and detector-coordinate systems are shown.

(NCC) between gradients of the source and target images, which is formulated as

$$\text{GC}(A, B) = \frac{1}{2}\{\text{NCC}(\nabla_x A, \nabla_x B) + \text{NCC}(\nabla_y A, \nabla_y B)\} \tag{3.2}$$

where

$$\text{NCC}(A, B) = \frac{\sum_{(i,j)}(A - \bar{A})(B - \bar{B})}{\sqrt{\sum_{(i,j)}(A - \bar{A})^2}\sqrt{\sum_{(i,j)}(B - \bar{B})^2}}$$

, the horizontal and vertical directions are denoted as ∇_x and ∇_y, and the source and target images are denoted as A and B, respectively.

3.2.7 Initialization of patient position

In our experiments, the geometric calibration of the imaging system was not carried out, and SOD and SDD were set to 1.8 and 2.0 m, respectively, which are approximate values used in a typical clinical protocol of chest radiography. Better geometric calibrations generally produce better registration results; however, obtaining an accurate calibration every time using a specially designed calibration phantom is difficult in a routine clinical setting. Through the experiments, we

found that this simple assumption in the geometry still provided a reasonably accurate initialization of patient position, resulting in a robust registration. We further discuss the initialization in Section 3.4.

3.2.8 Regularization and optimization

To improve the robustness of optimization, we introduce the following penalized objective function that takes into account the effects of the intercostal muscle that connects the neighboring ribs and enforces the interlocking motion of the rib cage (i.e., the movement of each rib bone is not independent but is constrained by the relative position with respect to its neighboring ribs).

$$\hat{\Theta} = \arg\max_{\Theta} \text{GC}(\text{LCN}(Xray), Proj(\Theta)) - \lambda g(\Theta) \tag{3.3}$$

where

$$g(\Theta) = \begin{cases} \max_i (\Delta d_i - \Delta d_{TH})^2 & (\max_i(\Delta d_i) > \Delta d_{TH}) \\ 0 & (otherwise) \end{cases}$$

The term Θ denotes the parameters to optimize, which represent the transformation of all the rib bones, $Proj(\Theta)$ is the simulated projection image (i.e., DRR) of the rib bones whose positions are parameterized by Θ, $LCN(Xray)$ represents the actual x-ray image after the LCN pre-process, and $g(\Theta)$ is the regularization term that penalizes the objective function according to the distance between the neighboring ribs. The distance between the i-th rib and its neighboring rib in a reference-phase CT (maximum exhale phase in our experiments) is defined as Δd_i, which we compute as

$$\Delta d_i = \frac{1}{|p_i|} \sum_{\mathbf{p}_i \in p_i} \left\{ \|T^{rib,i}\mathbf{p}_i - T^{rib,i+1}\hat{\mathbf{p}}_{i+1}\| - \|\mathbf{p}_i - \hat{\mathbf{p}}_{i+1}\| \right\}$$

where

$$\hat{\mathbf{p}}_{i+1} = \arg\min_{\mathbf{p}_{i+1} \in p_{i+1}} \|\mathbf{p}_i - \mathbf{p}_{i+1}\|$$

The term $\mathbf{p}_i$ is a point set representing the center line of the i-th rib bone computed from its segmentation mask using the skeletonization algorithm based on the distance field and fast matching method (Van Uitert and Bitter (2007)).

By enforcing the constraint on the distance between adjacent ribs so that it does not largely deviate from its initial distance, the penalty term helps maintain the anatomically plausible rib-cage shape, preventing mis-registration by incorrectly matching one rib in the CT to the other ribs in the projection image. The parameter λ is a weight used to balance the similarity-metric term and the penalty term. In this study, λ and Δd_{TH} were empirically determined as 1×10^{-4} and 5 mm, respectively.

A stochastic optimization algorithm called covariance matrix adaptation evolution strategy (CMA-ES) (Hansen (2006)) is used for optimization. This highly parallelizable evolution strategy is especially beneficial on GPU implementation, as detailed in a previous study (Otake et al. (2012)). The proposed optimization strategy consists of three stages. The first stage optimizes the global rigid transformation of the rib cage (6 DoFs for a rigid object consisting of all vertebrae and ribs). The second stage uses the result of the first stage as an initial estimate to jointly optimize the local rigid transformation of the rib-rotation angle θ around the rotation axis $\mathbf{n}$ for each rib ($6 + N$ DoF for N ribs) assuming the relative position of the rotation axis with respect to the vertebrae is fixed. The final stage uses the result of the second stage as an initial estimate to further jointly optimize the rotation axis $\mathbf{n}$ within a small cylindrical region (i.e., $\mathbf{n}$ is perturbed inside the cylinder with radii of r_0 and r_1 at its base, as shown in Fig. 3.1(b).). The two-level multi-resolution pyramid (Otake et al. (2013)) was used in each stage to improve robustness against local optima. The optimization parameters used in our experiments are summarized in Table 3.1.

Table 3.1: Summary of optimization parameters.

Optimization stage	First stage: Global transformation search	Second stage: Rib angle search	Final stage: Rib axis search
Dimension of solution space	6	6+N	6+5N
Population size (1st-level, 2nd-level in Gaussian resolution pyramid)	(200, 100)	(200, 100)	(200, 100)
Number of multi-starts	30	10	10
Regularization Lambda	-	1×10^{-4}	1×10^{-4}
Δd_{TH} [mm]	-	5	5
Upper/lower bound of search space Translation (x, y, z) [mm]	$(\pm 50, \pm 100, \pm 150)$	$(\pm 50, \pm 50, \pm 50)$	$(\pm 10, \pm 10, \pm 10)$
Rotation (x, y, z) [deg]	$(\pm 15, \pm 15, \pm 15)$	$(\pm 15, \pm 15, \pm 15)$	$(\pm 5, \pm 5, \pm 5)$
Rib angle [deg]	-	± 15	± 5
Rib axis (r, ϕ) [mm, deg]	-	-	$(\pm 15, \pm 180)$

3.2.9 Error metric and ground truth

One of the common error metrics used in evaluating 2D-3D registration is the mean projection distance (mPD) (van de Kraats et al. (2005)), which is the average distance in the projection plane between anatomical landmarks defined on the target bone in 3D (and projected onto the 2D plane with the computed transformation) and its corresponding ground truth point in the 2D x-ray image. The mPD metric requires identification of the corresponding anatomical landmarks in the x-ray image, which is quite difficult and error-prone in the case of rib bones because of their smooth surfaces with a small number of feature points. Therefore, in this study, we evaluated the error using the contour lines in the projection plane similar to a previous study (Lamecker et al. (2006)), which we call the mean projected contour distance (mPCD). Given the true contour s_i of an i-th rib projected onto the x-ray image and a contour s_i' in the DRR of a rib at an estimated pose, mPCD is computed as

$$\text{mPCD} = \frac{1}{|s_i|} \sum_{\mathbf{x} \in s_i} \min_{\mathbf{x}' \in s_i'} \|\mathbf{x} - \mathbf{x}'\|^2. \tag{3.4}$$

Example registration results with different mPCD values are illustrated in Fig. 3.4. The ground truth in the real-image experiments was defined manually by tracing the contours of the rib in the x-ray images.

3.3 Results

3.3.1 Simulation experiments

In the simulation experiments, we used DRRs of the two-time-phase CTs, acquired at maximum inhale and maximum exhale phases, included in the EMPIRE10 data set (Murphy et al. (2011)) as the target image. The detailed specifications of the data set are listed in Table 3.2. Among 30 cases in the EMPIRE10 data set, we selected 6 that include both maximum inhale and maximum exhale phases, as shown in Fig. 3.5. Figure 3.5 also illustrates the rotation angle of each
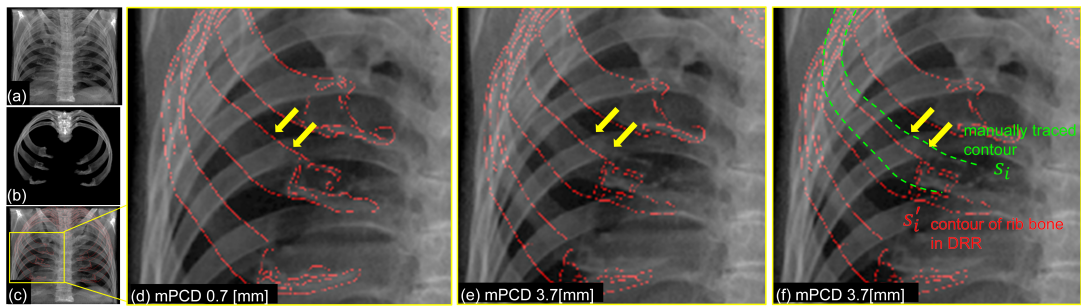
Figure 3.4: Error metric used in this study, mean projected contour distance (mPCD), and example registration results with different mPCD values. (a) X-ray image, (b) digitally reconstructed radiograph (DRR) at the estimated pose, (c) overlaid with the DRR edges (in red), (d)(e) enlarged view of the trials with different mPCD values. The yellow arrows indicate the true contour of the rib on the x-ray image. The result with mPCD of 0.7 mm (d) exhibits almost no visually recognizable difference between the lines while 3.7 mm (e) shows a clear discrepancy. (f) mPCD was defined as the distance between the manually traced contour (green dashed line) of each rib bone on the x-ray image and the automatically detected contour on the DRR (red dashed line).

47

rib between the two phases, which was computed by applying the 3D-3D surface registration, the iterative closest point (ICP) algorithm (Besl et al. (1992)), on the rib surfaces that were segmented from the CTs. Note that there were no recognizable deformations of spine curvature between the two time-phases because the subject stayed in the supine position during both scans. Therefore, the experimental setup was suitable for evaluating the effectiveness of the proposed rib-motion model excluding one potential error factor, which is spine deformation. To simulate the quantum noise appearing in real x-ray images, the Poisson random noise computed based on the exposure level at a uniform 1.985×10^5 photons per detector element for an un-attenuated beam was applied to the target DRR (Prince and Links (2014)). Note that the photon counts were computed from the un-attenuated beam in the real x-ray image acquired with our clinical protocol, which is explained in detail in the next section.

In the proposed method, we estimate the rotation axis as well as the rotation angle of the ribs. Thus, in the following explanation of our experiments, we use the term *true rotation axis* to compare the registration accuracy with an ideal situation in which the rotation axis is accurately known. The *true rotation axis* was computed via ICP registration as explained above. We denote the translation and rotation from the exhale phase to the inhale phase as t_{GT} and R_{GT}, respectively, and the translation to the *true rotation axis* as t_C. The rotation matrix $R_G T$ is represented as the axis-angle representation based on the Rodrigues' rotation formula, and the axis vector was used as the direction of the *true rotation axis.*

The evaluation of the proposed uniaxial joint model and rotation-axis search was conducted by comparing the registration error in the following three scenarios. (1) Angle search with the true axis using two-time-phase CT: Optimization of the rotation angle using the *true rotation axis* (i.e., the simplest case in which one DoF rotation around the axis provides the registration that perfectly matches the rib bone in the two phases). (2) Angle search using a single-time-phase CT: Optimization of the rotation angle using the axes defined by the anatomical landmarks, which we call *anatomical axis* (i.e., the simple one-DoF optimization

Table 3.2: Specification of data set used in experiments

Simulation experiments (8 cases included in EMPIRE10)		
CT (source)	Dimension (avg.)	$438 \times 329 \times 399$ [pixel]
	Pixel spacing	$0.63 \sim 0.77$ [mm]
	Slice spacing	0.7 [mm]
	Breathing phase	Exhalation
DRR (target)	Dimension	1344×1344 [pixel]
	Pixel spacing	0.32×0.32 [mm]
	Breath phase	Inhalation
Real image experiments (8 cases)		
CT (source)	Dimension (avg.)	$512 \times 512 \times 120$ [pixel]
	Pixel spacing	0.68×0.68 [mm]
	Slice spacing	2.5 [mm]
	Posture	Supine
	Breathing phase	Exhalation
X-ray video (target)	Dimension	1344×1344 [pixel]
	Pixel spacing	0.32×0.32 [mm]
	Number of frames	36
	Frame rate	3 [fps]
	Posture	Standing
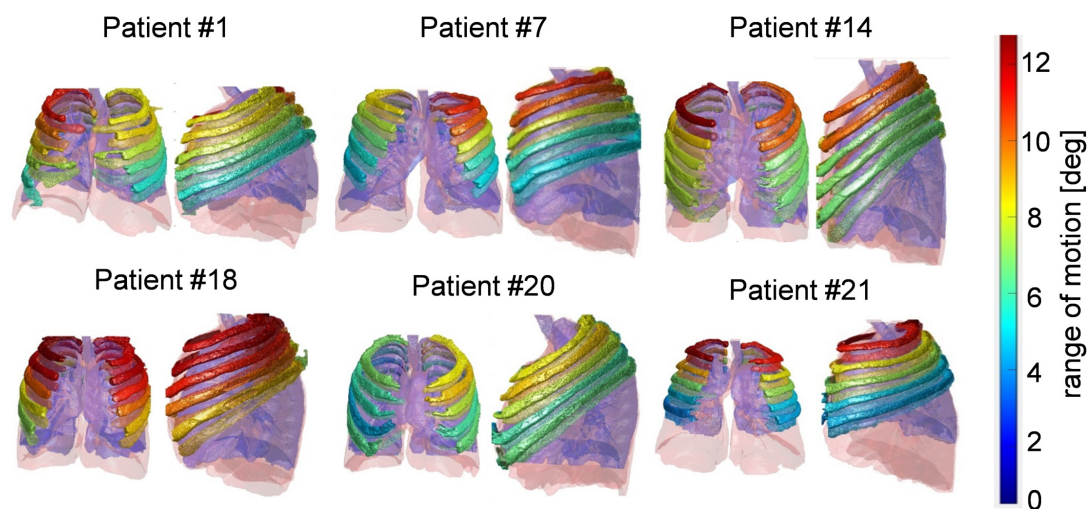
Figure 3.5: Rib cage and lungs of six cases in EMPIRE10 data set that were used in the evaluation experiments. CT images at inhale and exhale phases were analyzed. Color of each rib indicates the rotation angle between inhale and exhale phases (see the colormap on the right). The opaque and transparent ribs show the inhale and exhale phases, respectively (they overlap each other in some cases). The lungs at inhale and exhale phases are shown in red and blue. The larger rotation angle was observed at the ribs of the superior levels.

with a realistic scenario in the case in which we have only single-time-phase CT and the *true rotation axis* is not available). (3) Angle+axis search using a single-time-phase CT (proposed method): Optimization of one rotation angle and four rotation-axis parameters representing its direction using the *anatomical axis* as an initial estimate (i.e., a method in which the optimization is most involved though accurate).

The results of the simulation experiments with six ribs (2nd, 3rd and 4th ribs on the left and right sides) of six patients are shown in Fig. 3.6 (i.e., $n = 36$ data points in total). Since the CMA-ES produces different optimization results even with the same initialization due to the stochasticity in generation of the population, 20 trials from the same initialization were carried out and their average is shown in the figure. Scenario (3), the proposed method with rotation-axis search, provided statistically significant improvement (double asterisk (**) indicates that the p-value was less than 0.01) by the Welch's test in mPCD (mean $\pm$ std: 0.742 ± 0.281 mm) and the rotation angle error (1.495 ± 0.993 deg) compared to scenario (2), which is the angle search with the *anatomical axis*. The error is not supposed to be zero even with the *true rotation axis*, i.e. scenario (1), because of the following two reasons: effect of soft tissues and the axis estimation error. The main potential error source is the soft tissues in the simulated image that was introduced to enhance the realism of the experiment. The soft tissues created false local optima, causing the degradation of the accuracy. Moreover, what we call the *true rotation axis* might have error due to the error in ICP registration of the two-time-phase CT.

The scatter plot in Fig. 3.6(c) shows the relationship between mPCD (the error metric in 2D projection) and the rotation-angle error (the error metric in 3D space). All registration trials with an mPCD value smaller than 1.0 mm (left side of the vertical dotted line) were with the rotation-angle error smaller than 3.0 deg (lower side of the horizontal dotted line), suggesting the effectiveness of mPCD, which is observable in the 2D projection image, as an error metric to detect registration failure.

Table 3.3: Results of the simulation experiments. The mean and standard deviation of mPCD and rotation angle error are reported for each scenario. For the statistical significance, refer to Fig. 3.6.

mPCD [mm]				
	Non-uniaxial model	Uniaxial model		
		True axis	Anatomical axis	
			w/o rotation axis search	w/ rotation axis search
w/o penalty	5.844 ± 2.433	0.698 ± 0.326	1.148 ± 0.743	0.757 ± 0.288
w/ penalty	5.355 ± 2.043	0.664 ± 0.284	1.151 ± 0.736	0.742 ± 0.281

Rotation angle error [deg]				
	Non-uniaxial model	Uniaxial model		
		True axis	Anatomical axis	
			w/o rotation axis search	w/ rotation axis search
w/o penalty	11.139 ± 4.124	0.863 ± 0.654	2.246 ± 1.839	1.539 ± 0.993
w/ penalty	10.161 ± 3.454	0.840 ± 0.686	2.237 ± 1.808	1.495 ± 0.993

Moreover, in order to evaluate the effectiveness of the proposed uniaxial joint model, we conducted the registration with non-uniaxial joints, resulting in a rotation-angle error of 11.139 ± 4.124 deg and mPCD of 5.844 ± 2.433 mm. The overall result in the simulation experiments is shown in Table 3.3.

To evaluate the influence of the Poisson random noise in the simulation experiments, we evaluated the registration accuracy without the noise using the *true rotation axis* and found a slight accuracy improvement (0.664 ± 0.284 mm, 0.840 ± 0.686 deg to 0.661 ± 0.300 mm, 0.818 ± 0.693 deg, for mPCD and rotation-angle error, respectively).

These results indicate the effectiveness of the rib-motion model constrained by a uniaxial joint as well as the importance of accurate estimation of the rotation axis. The highest accuracy was achieved when the *true rotation axis* computed

from the two-time-phase CT was used. The rotation-angle error significantly increased in the scenario using the *anatomical axis*, which has been regarded as the rotation axis of ribs since De Troyer's model (De Troyer et al. (2005)) was introduced, while the error decreased to the level close to that with the *true rotation axis* when the rotation axis was jointly searched along with the rotation angle of the ribs in the optimization. In practice, the *true rotation axis* is only available when we acquire a two-time-phase CT, which requires higher radiation dose. The proposed rotation-axis-search approach with a rough initial estimate by De Troyer's anatomical landmarks can be considered as a practical alternative in routine clinical practice or annual screening examinations in which a lower-dose modality is preferable. Note that, in this study, we automatically obtained the initial estimate of the axis from an individual single-time-phase CT using a statistical modeling of the anatomical landmarks.

3.3.2 Real-image experiments

In the real-image experiments, we used x-ray videos acquired when the patients were in standing position while breathing and CTs of each patient acquired in the supine position at the inhale phase of eight cases, as detailed in Table 3.2. The study protocol was approved by our institutional review board (#598-1 at Kanazawa University, Japan). Each x-ray video consisted of 30 frames acquired over one breathing cycle in 10 sec starting from the maximum exhale phase. The entrance-surface dose for the 30 frames, measured in air without backscattering, was approximately 0.4 mGy, which was less than that in the lateral chest radiography with the protocol suggested by the International Atomic Energy Agency (IAEA), which is 1.5 mGy (Tanaka et al. (2015)). The experiments with the real images allowed us to analyze the robustness of the proposed method against the properties in a real x-ray-imaging device that were not considered in the simulation experiments, yet might adversely affect registration accuracy, such as the polyenergetic x-ray spectrum, x-ray scatter, and finite focal spot size. Note that
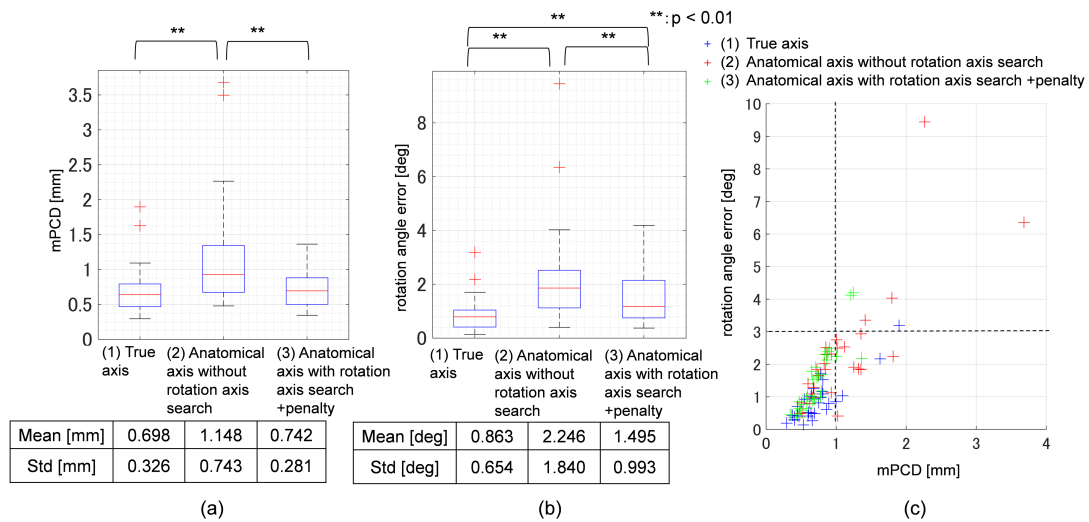
Figure 3.6: Results of simulation experiments. (a-b) box and whisker plots for different search scenarios. (a) is used mPCD and (b) is used rotation-angle error as error metric. Boxes denote the 1st/3rd quartiles, the median is marked with the horizontal line within each box, and outliers are marked with crosses. (c) The scatter plot showing the rotation angle error as a function of mPCD. The correlation between the two metrics suggested validity of using mPCD, the error metric observable in the projection image, as an indicator of the unobservable 3D rotation-angle error.

the slice spacing of the CT used in our clinical protocol was larger than that of the EMPIRE10 data set used in the simulation experiments. We applied the spline interpolation in the out-of-plane direction and created a volume with 0.625 mm slice spacing before the pre-process to suppress the aliasing effect appearing in the DRR.

For quantitative evaluation of the registration accuracy, four frames at 2.5 sec intervals were selected for each patient, and the true contour of each rib was manually traced to calculate mPCD. As is the case with any real patient studies, the ground truth data of the actual rib-rotation angle in the 3D space were not available. Therefore, we assessed the registration accuracy by the contour error appearing in the 2D projection plane, which our simulation experiments suggested to be correlated with the 3D rib-rotation angle.

First, we investigated the effect of LCN on the cost function and on robustness of the 2D-3D registration with real x-ray images acquired with the low-dose protocol. The comparison of the results with and without LCN is shown in Fig. 3.7. The scatter plots show mPCD as a function of the similarity measure, GC, of 30 trials of the 6 ribs for each patient. Note that one cross represents an average over 100 function evaluations (approximately 5×10^4 function evaluations were conducted in one trial).

In the analysis of numerical optimization, the goodness of the cost function was generally measured with the shape of the plot of the cost value as a function of the optimization parameters, which is often called the *landscape* (Talbi (2009)). In particular, the dispersion metric (Lunacek and Whitley (2006)) was proposed to quantitatively determine whether the local optima in the cost function are dispersed over the parameter domain or rather concentrated at one point. The cost function with concentrated local optima is considered better for optimization than that with dispersed local optima. Our analysis illustrated in Fig. 3.7 is based on a similar idea, though we used the error metric (i.e., mPCD) as the x-axis rather than the optimization parameter to verify the validity of our similarity metric as well as the dispersion of local optima. In the registration problem, the

global optimum of the cost function does not necessarily yield the minimum error in cases, for example, 1) the choice of similarity metric is not appropriate and 2) there is a mismatch between the image-simulation process (i.e., DRR generation in our case) and the actual image-acquisition process. As shown in Fig. 3.7, we first confirmed that the global maximum of the similarity metric actually yielded the minimum mPCD, suggesting that the effect of 1) and 2) mentioned above is reasonably small and the similarity metric we used worked as intended. We define the shape of the function that relates the similarity to error metric as similarity metric *landscape*, following customary expressions in the field of mathematical optimization. The *landscape* allows an intuitive analysis of suitability of the similarity metric for the target registration problem. We then observed that the similarity metric *landscape* with LCN is clearly better than that without LCN, showing a smaller number of local optima concentrated to the smallest mPCD value. The mean and standard deviation of the final mPCD values also clearly improved with LCN. Also note that in the case of patient #4, in which the image contrast over the entire image was rather uniform resulting in a relatively clear boundary of the rib bones even without LCN, the *landscape* looks better than the other cases. This indicates that LCN correctly worked for the cases in which the contrast was not uniform in the original x-ray image and the similarity metric was not useful to stably find the correct alignment due to the confusing edges around the true rib region.

Finally, quantitative evaluation of LCN and the rotation-axis search were conducted by comparing the registration error for the four scenarios: With/without LCN and the rotation-axis search. The results with six ribs of eight patients are shown in Fig. 3.8 (i.e., $n = 48$ data points in total). The average of the four frames is shown in the plot. The case with both LCN and rotation-axis search yielded statistically significant improvement of Welch's test ($p < 0.01$) in mPCD (1.255 ± 0.615 mm) compared with all other cases.

Table 3.4 shows the overall result in the real-image experiments. mPCD was improved from 1.548 ± 0.810 mm to 1.255 ± 0.615 mm in these experiments.
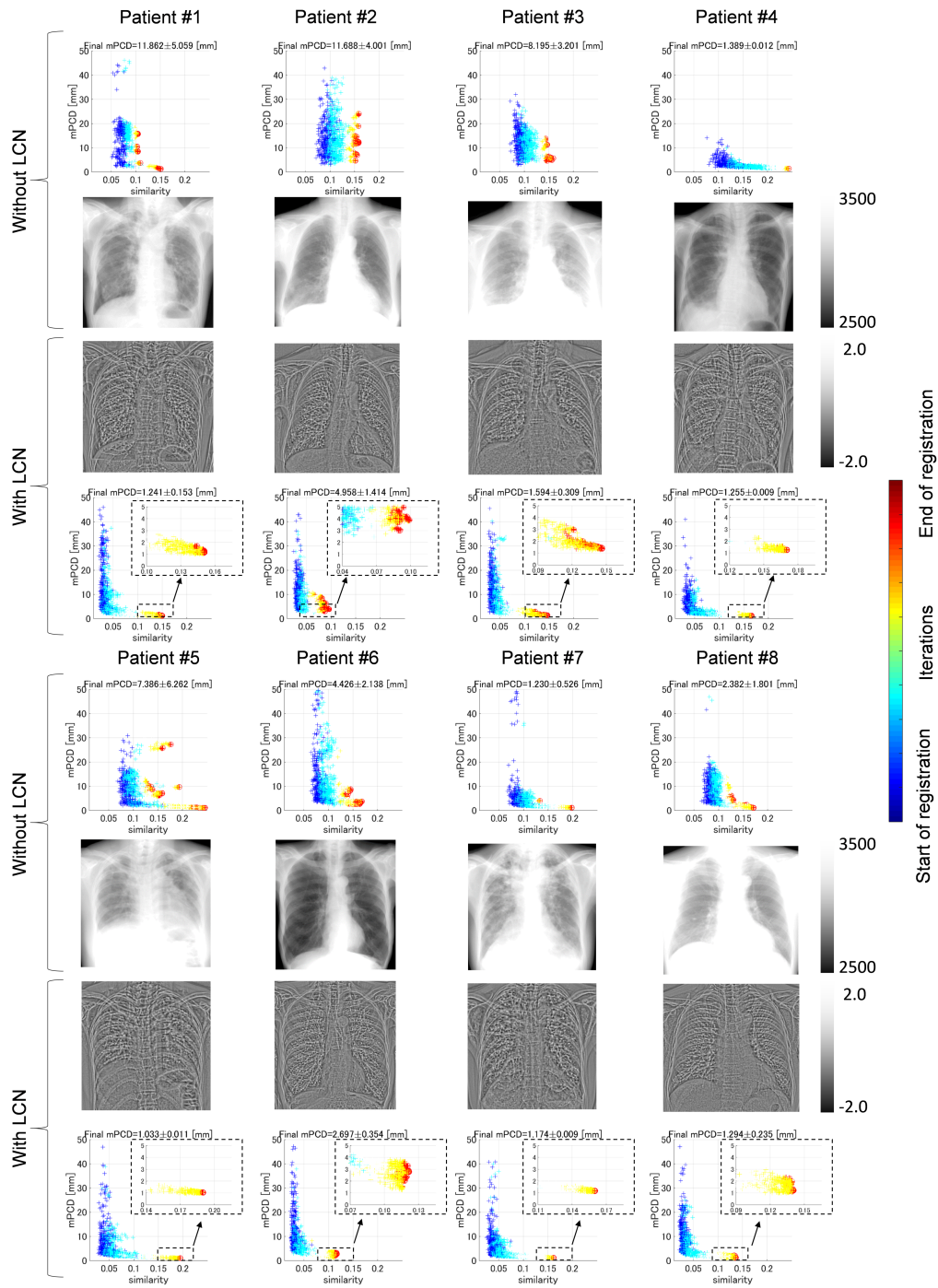
Figure 3.7: Analysis of similarity metric *landscape* to investigate the effectiveness of LCN. The mPCD is plotted as a function of the similarity measure (GC) in the top and bottom rows of each patient. The color of each cross indicates the progression of iterations in the optimization process. The red circle indicates the converged solution.

57

Table 3.4: Results of the real-image experiments. The mean and standard deviation of mPCD are reported for each scenario. For the statistical significance, refer to Fig. 3.8.

mPCD [mm]			
		Uniaxial model with anatomical axis	
		w/o rotation axis search	w/ rotation axis search
w/o LCN	w/o penalty	6.358 ± 6.636	6.436 ± 6.998
	w/ penalty	2.708 ± 1.815	2.607 ± 1.798
w/ LCN	w/o penalty	1.628 ± 0.851	1.548 ± 0.810
	w/ penalty	1.392 ± 0.711	1.255 ± 0.615

In order to evaluate the sensitivity of LCN to the window size, registration accuracy was evaluated with window size of 11, 21, and 31, all of which provided a visually reasonable filtered image. The error with the rotation axis search with penalty (i.e., the proposed method) was 1.280 ± 0.627 mm, 1.255 ± 0.615 mm, and 1.272 ± 0.662 mm, respectively, which suggested robustness against the window size.

Although the proposed method improved robustness significantly, we still observed failures, which can be classified into two major modes: failures 1) due to insufficient contrast-to-noise ratio in the image even after LCN and 2) due to an irregular joint rotation that does not fit to our uniaxial motion model. Patient #2 had a low contrast-to-noise ratio; thus the gradient image was noisy, which created many false local optima in the cost-function space and led to failures (mPCD > 4 mm) in some trials. In case of patient #6, the final registration error was larger than others (mPCD > 2.5 mm) though the spine was aligned correctly based on visual verification, which may suggest that the rib motion did not follow the assumption of the uniaxial joint model. Even for those two patients (#2 and #6) that did not result in a clinically acceptable accuracy, Fig. 3.7 suggests a considerable improvement in robustness by applying LCN.

	Without local contrast normalization		With local contrast normalization	
	Without rotation axis search	With rotation axis search +penalty	Without rotation axis search	With rotation axis search +penalty
Mean [mm]	6.358	2.607	1.628	1.255
Std [mm]	6.636	1.798	0.851	0.615
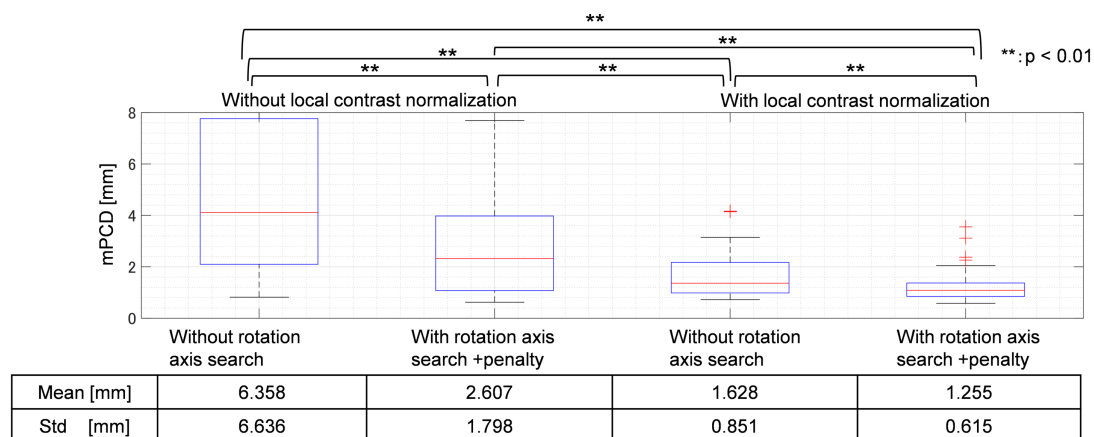
Figure 3.8: Quantitative evaluation results of the real-image experiments with 6 ribs (2nd, 3rd and 4th ribs of both sides) of 8 patients (i.e., 48 data points in total). The mean and standard deviation for each scenario are listed in the table below. The error was significantly decreased by adding the rotation axis search and penalty term, and further decreased by adding the preprocessing with LCN.

Figure 3.9 shows an example visualization of the registration results in the real-image experiment. The manually traced contour (lines in green) and estimated contour (red) are overlaid to visually confirm the registration accuracy. Figure 3.10 shows the 3D visualization of the reconstructed motion color-coded by the range of motion of each rib, similar to Fig. 3.5, though the colors in Fig. 3.10 were computed from a single-time-phase CT and an x-ray video while those in Fig. 3.5 were from a two-time-phase CT. In this paper, we define the range of motion as the difference between the maximum and minimum rib-rotation angles within the 10 sec in the video. The plots in Fig. 3.10(d) show the motion pattern of the 2nd to 4th ribs as a function of time for one breathing cycle. Based on the error analysis in the simulation experiments, which is 1.495 ± 0.993 deg with the rotation-angle search, and the range of motion, 7.115 ± 2.018 deg, the proposed method would provide a reasonable estimate of the trend in the motion that is associated with a clinically important symptom of diseases such as the left-right asymmetric motion in patient #5 and the abrupt swing of the ribs that deviates

59
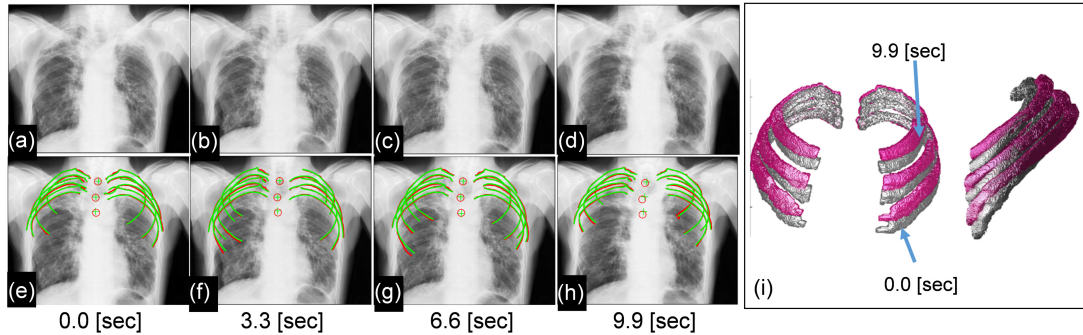
Figure 3.9: Visualization of the registration results at 4 representative frames of patient #1 in the real image experiment. (a-d) The original x-ray image, (e-h) overlaid with the contours of the ribs (green: manually identified ground truth, red: estimated) and the center of vertebrae (green cross: ground truth, red circle: estimated)

from the normal breathing cycle (indicated with the red arrows in patient #1 of Fig. 3.10(d)).

The proposed method was implemented on MATLAB ver. R2015b (Math-Works, Natick, MA), which included function calls for rendering DRRs and computing the similarity metric on the GPU with CUDA 8.0 (nVidia, Santa Clara, CA). The average number of function evaluations, including DRR rendering and computation of similarity measure, conducted in one registration trial and the total computation time were approximately 5×10^4 and 5 min without the rotation-axis search and 2×10^5 and 20 min with rotation-axis search. Note that the real-time application is not our main focus in this paper. Although previous work such as (Gendrin et al. (2012)) used a non-evolution strategy for optimization and demonstrated faster computation time than our current study, the evolution strategy such as CMA-ES provides a much higher robustness as shown in (Gill et al. (2012); Otake et al. (2013)) at the sacrifice of longer computation time due to a large number of evaluation of cost function in each generation. In this study, our focus is robustness against local-minima rather than computation time, thus we selected CMA-ES.
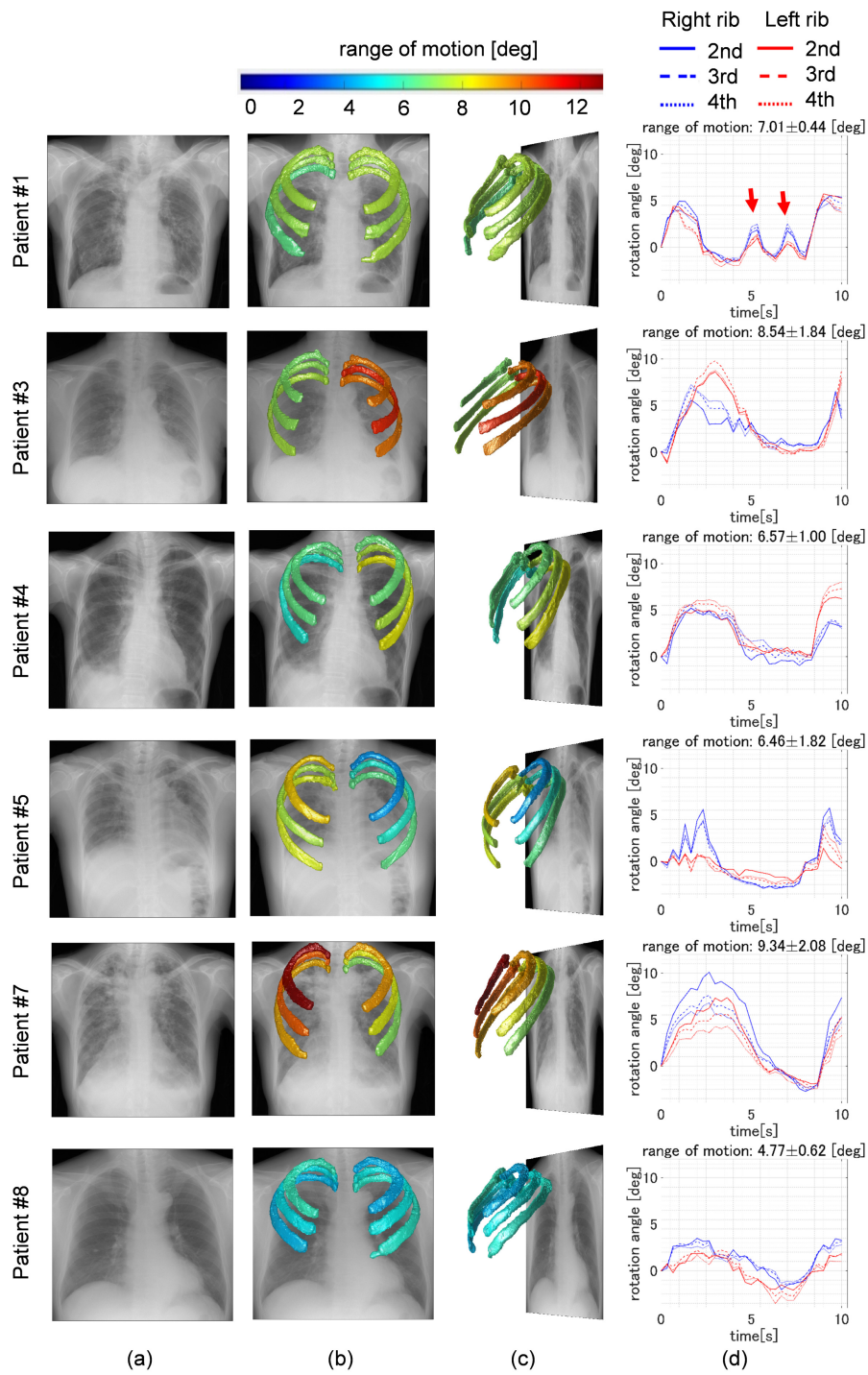
Figure 3.10: Visualization of the registration result of all cases in the real image experiment. (a) The original x-ray image, (b,c) overlaid with the 3D rib model at the estimated pose. (d) The rib rotation angle around the rotation axis as a function of time.

3.4 Discussion

3.4.1 Effectiveness of the proposed method

One of the advantages of the analysis using the x-ray video over another approach that acquires multiple-phase CTs, for example so-called 4D CT, is significant reduction in radiation exposure. In a clinical scenario in which periodical follow-up screening examinations are required, the difference in radiation exposure is even larger since the proposed method requires only one CT image at the first examination and can use it with lower-dose x-ray videos in repetitive examinations as long as the shape of the bones remains unchanged.

We evaluated the proposed method through simulation experiments using the two-time-phase CT (i.e., inhale and exhale phases) and real-image experiments using x-ray videos acquired with a protocol used in a routine clinical setting. In both simulation and real-image experiments, the optimization of the rotation axis yielded statistically significant improvement ($p < 0.01$) compared to the scenario without rotation-axis search. A uniaxial rib-motion model together with the optimization of the rotation axis to improve robustness in optimization while maintaining a high degree of accuracy. We also confirmed the effectiveness of LCN in the real-image experiments. The LCN in the pre-processing of x-ray video to better condition the cost function space.

In this study, for the first time to the best of our knowledge, we demonstrated that LCN actually changed the cost-function space in a way that is suitable for optimization, as clearly shown in Fig. 3.7.

3.4.2 Application to other tasks

Initialization of registrations is a critical aspect that affects on success of registration. The x-ray videos we used in this study were obtained in a standard clinical protocol of standing chest x-ray using a fixed source and detector. Therefore, we initialized the registration using the prior knowledge about the position of the

patient relative to the detector (i.e., SDD = 2.0 m, SOD = 1.8 m) and found that this simple initialization without using an accurate geometric calibration resulted in a reasonably high success rate and low registration error in the real image experiments. However, Villard et al. (2014) described the potential use of 2D-3D registration of the ribs in an intraoperative navigation using images acquired from a mobile x-ray C-arm for estimating liver deformation co-occurring with rib motion. In such an application scenario, the patient position relative to the detector is different from our setup. An application-specific initialization considering the setup of each imaging device would be necessary. The setup is generally consistent in a specific type of surgery and a simple assumption similar to the one we used in this study may be applied to other setups.

3.4.3 Limitations

One limitation of this study was the lack of evaluation with patients with scoliotic spines. In scoliotic patients, the rotation axis of the ribs significantly differs from healthy subjects, or the movement may not follow the uniaxial joint model. Further clinical evaluations are needed to assess the clinical applicability of the proposed method. A more sophisticated motion model considering the scoliosis in such patients is also necessary.

Another potential limitation is lack of deformation of vertebrae in the model. As described in Section 3.2.4, we considered our target vertebrae (T2-T4) as one rigid object. One potential rationale for this assumption came from the study by Beillas et al. (2009), in which they investigated the relationship between the body posture and spine curvature in nine subjects using an MRI scanner (Fonar Upright, Melville, NY) that allows acquisition in both standing and supine positions. They found that the difference in the T4-T12 thoracic kyphosis angle (i.e., forward rounding curve of the thoracic vertebrae) between standing and supine positions was 16 ± 4 deg. Assuming that the spine is curved at a regular interval at each intervertebral disc, the kyphosis angle per disc is 2.3 deg. Therefore, in

this study, we considered that the change in the curvature from T2 to T4 is small enough to be modeled as one rigid object. However, we acknowledge that further investigation of the effect of this assumption in diseased patients is necessary. A more sophisticated spine model such as the statistical shape and pose model proposed by Rasoulian et al. (2013), may be incorporated to improve registration accuracy.

In proposed method, segmentation of each rib bone is required manual interaction (i.e., seed placements for each rib). One possible method to automate the segmentation would be to use a machine learning approach.

3.4.4 Future work

Our future work includes incorporation of the knowledge on temporal continuity as the regularization method similar to that proposed by Chen et al. (2013). By imposing the temporal continuity constraint in the optimization of time-sequential frames, a smoother motion, which is generally more plausible in terms of physical principle, can be obtained, although it sacrifices detectability of abrupt motion which is critical in the diagnosis of certain diseases. This trade-off requires investigation. Another future work is automated detection of registration failure, which would reduce human interaction and potentially improve robustness by repeating the registration trial from a different initialization in case of failure. We plan to apply a CNN to failure detection. A CNN would be able to be trained to predict the error metric in the projection plane, mPCD, directly from the registration results.

3.5 Conclusion

We proposed a robust 2D-3D registration method for 3D rib-motion recovery from dynamic radiography. Specifically, the contributions of this paper are the introduction of 1) a uniaxial rib-motion model together with the optimization

of the rotation axis to improve robustness in optimization while maintaining a high degree of accuracy, and 2) the LCN in the pre-processing of x-ray video to better condition the cost function space. We evaluated the proposed method through simulation experiments using the two-time-phase CT (i.e., inhale and exhale phases) and real-image experiments using x-ray videos acquired with a protocol used in a routine clinical setting. In both simulation and real-image experiments, the optimization of the rotation axis yielded statistically significant improvement ($p < 0.01$) compared to the scenario without rotation-axis search. We also confirmed the effectiveness of LCN in the real-image experiments.

4 Cross-modality image synthesis from unpaired data using CycleGAN

Computed tomography (CT) is commonly used in orthopedic procedures. Magnetic resonance imaging (MRI) is used along with CT to identify muscle structures and diagnose osteonecrosis due to its superior soft tissue contrast (Cvitanic et al. (2004)). However, MRI has poor contrast for bone structures. It would be helpful if a corresponding CT were available, as bone boundaries are more clearly seen and CT has standardized (i.e., Hounsfield) units. Considering radiation exposure in CT, it is preferable if we can delineate boundaries of both muscle and bones in MRI. Therefore, our focus in this chapter is MR-to-CT synthesis.

4.1 Related work

Image synthesis has been extensively studied using the patch-based learning (Torrado-Carvajal et al. (2016)) as well as deep learning, specifically, convolutional neural networks (CNN) (Zhao et al. (2017)) and generative adversarial networks (GAN) (Kamnitsas et al. (2017)). The conventional approaches required the paired training data, i.e., images of the same patient from multiple modalities that are registered, which limited the application. A method recently proposed by Zhu et al. (2017), called CycleGAN, utilizes the unpaired training data by appreciating the cycle consistency loss function. While CycleGAN has

already applied to MR-to-CT synthesis (Wolterink et al. (2017)), all these previous approaches in medical image application targeted CT and MRI of the head in which the scan protocol (i.e., field-of-view (FOV) and the head orientation within the FOV) is relatively consistent resulting in a small variation in the two image distributions even without registration, thus a small number of training data set (20 to 30) allowed a reasonable accuracy. On the other hand, our target anatomy, the hip region, has larger variation in the anatomy as well as their pose (i.e., joint angle change and deformation of muscles).

Applications of image synthesis include segmentation. Some previous studies aimed at segmentation of musculoskeletal structures in MRI (Gilles and Magnenat-Thalmann (2010); Ranzini et al. (2017)), but the issues in these studies were the requirement for multiple sequences and devices. Another challenge in segmentation of MRI is that there is no standard unit as in CT. Therefore, manually traced label data are necessary for training of each sequence and each imaging device. Thus, MR-to-CT synthesis realizes modality independent segmentation (Hamarneh et al. (2008)).

4.2 Methods

4.2.1 Datasets

The datasets we used in this study are MRI dataset consisting of 302 unlabeled volumes and CT dataset consisting of 613 unlabeled, and 20 labeled volumes which are associated with manual segmentation labels of 19 muscles around hip and thigh, pelvis, femur and sacrum bones. Patients with metallic artifact due to implant in the volume were excluded. As an evaluation dataset, we also used other three sets of paired MR and CT volumes, and 10 MR volumes associated with manual segmentation labels of gluteus medius and minimus muscles, pelvis and femur bones, as a ground truth. MR volumes were scanned in the coronal plane for diagnosis of osteonecrosis by a 1.0T MR imaging system. The T1-
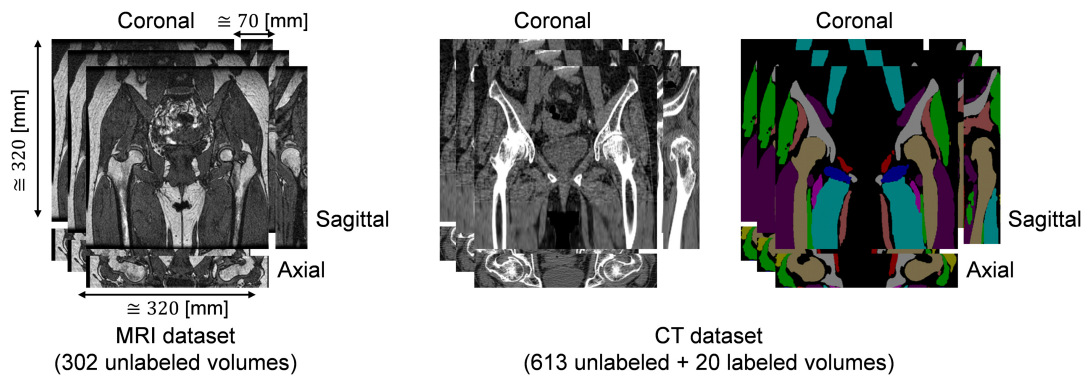
Figure 4.1: Training datasets used in this study. MRI dataset consists of 302 unlabeled volumes and CT dataset consists of 613 unlabeled and 20 labeled volumes. N4ITK intensity inhomogeneity correction (Tustison et al. (2010)) was applied to all MRI volumes. Two datasets have similar field-of-view, although these are not registered.

weighted volumes were obtained by 3D spoiled gradient recalled echo sequence (SPGR) with a repetition time (TR) of 7.9 ms, echo time (TE) of 3.08 ms, and flip angle of 30. The field of view was 320 mm, and the matrix size was 256×256. The slab thickness was 76 mm, and the slice thickness was 2 mm without an inter-slice gap. CT volumes were scanned in the axial plane for diagnosis of the patients subjected to total hip arthroplasty (THA) surgery. The field of view was 360×360 mm and the matrix size was 512×512. The slice thickness was 2.0 mm for the region including pelvis and proximal femur, 6.0 mm for the femoral shaft region, and 1.0 mm for the distal femur region. In this study, the CT volumes were cropped and resliced so that the FOV resembles that of MRI volumes, as shown in Figure 4.1, and then resized to 256×256.

4.2.2 Image synthesis using CycleGAN with gradient-consistency loss

The underlying algorithm of the proposed MR-to-CT synthesis follows that of Zhu et al. (2017) which allows to translate an image from CT domain to MR domain without pairwise aligned CT and MR training images of the same patient. The workflow of the proposed method is shown in Figure 4.2. The networks G_{CT} and G_{MR} are generators to translate real MR and CT images to synthesized CT and MR images, respectivery. The networks D_{CT} and D_{MR} are discriminators to distinguish between real and synthesized images. While discriminators try to distinguish synthesized images by maximizing adversarial losses $\mathcal{L}_{CT}$ and $\mathcal{L}_{MR}$, defined as

$$\mathcal{L}_{CT} = \sum_{x \in I_{CT}} \log D_{CT}(x) + \sum_{y \in I_{MR}} \log(1 - D_{CT}(G_{CT}(y))), \qquad (4.1)$$

$$\mathcal{L}_{MR} = \sum_{y \in I_{MR}} \log D_{MR}(y) + \sum_{x \in I_{CT}} \log(1 - D_{MR}(G_{MR}(x))), \qquad (4.2)$$

generators try to synthesize images which is indistinguishable from the target domain by minimizing these losses. Where x and y are images from domains I_{CT} and I_{MR}. However, networks with large capacity have potential to converge to the one that translate the same set of images from source domain to any random permutation of images in the target domain. Thus, adversarial losses alone cannot guarantee that the learned generator can translate an individual input to a desired corresponding output. Therefore, the loss function is regularized by cycle consistency, which is defined by the difference between real and reconstructed image, which is the inverse mapping of the synthesized image (Zhu et al. (2017)). The cycle consistency loss $\mathcal{L}_{Cycle}$ is defined as

$$\mathcal{L}_{Cycle} = \sum_{x \in I_{CT}} |G_{CT}(G_{MR}(x)) - x| + \sum_{y \in I_{MR}} |G_{MR}(G_{CT}(y)) - y| \quad (4.3)$$

We extended the CycleGAN approach by explicitly adding the gradient consistency loss between real and synthesized images to improve the accuracy at the boundaries. The gradient correlation (GC) (Penney et al. (1998)) has been used

69
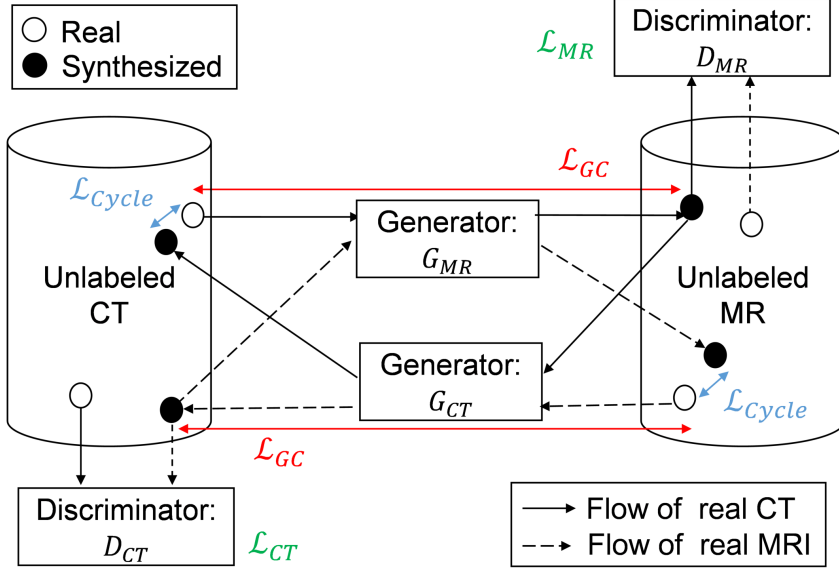
Figure 4.2: Workflow of the proposed method. G_{CT} and G_{MR} are generator networks that translate MR to CT images, and CT to MR images, respectively. D_{CT} and D_{MR} are discriminator networks to distinguish between real and synthesized images. The cycle consistency loss $\mathcal{L}_{Cycle}$ is a regularization term defined by the difference between real and reconstructed image. To improve the accuracy at the edges, loss function is regularized by gradient consistency loss $\mathcal{L}_{GC}$.

as a similarity metric in the medical image registration, which is defined in Eq. (3.2). We formulate the gradient-consistency loss $\mathcal{L}_{GC}$ as

$$\mathcal{L}_{GC} = \frac{1}{2}\{ \sum_{x \in I_{CT}} (1 - GC(x, G_{MR}(x))) + \sum_{y \in I_{MR}} (1 - GC(y, G_{CT}(y)))\} \quad (4.4)$$

Finally, our objective function is defined as:

$$\mathcal{L}_{total} = \mathcal{L}_{CT} + \mathcal{L}_{MR} + \lambda_{Cycle}\mathcal{L}_{Cycle} + \lambda_{GC}\mathcal{L}_{GC} \quad (4.5)$$

where λ_{Cycle} and λ_{GC} are weights to balance each loss. Then, we solve:

$$\hat{G}_{MR}, \hat{G}_{CT} = \arg \min_{G_{CT}, G_{MR}} \max_{D_{CT}, D_{MR}} \mathcal{L}_{total} \quad (4.6)$$

In this paper, we used 2D CNN with 9 residual blocks for generator, similar to the one proposed in (Johnson et al. (2016)). For discriminators, we used 70×70 PatchGAN (Isola et al. (2017)). We replaced the Eq. (1) and Eq. (2) by least-squares loss as in (Mao et al. (2016)). These settings follows (Zhu et al. (2017); Wolterink et al. (2017)). The CycleGAN was trained using Adam (Kingma and Ba (2014)) for the first 1×10^5 iterations at fixed learning rate of 0.0002, and the last 1×10^5 iterations at learing rate which linearly reducing to zero. The balancing weights were empirically determined as $\lambda_{Cycle} = 3$ and $\lambda_{GC} = 0.3$. CT and MR volumes are normalized such that intensity of [-150, 350] HU and [0, 100] are mapped to [0, 255], respectively.

4.3 Results

4.3.1 Quantitative evaluation on image synthesis

To evaluate image synthesis, we investigated dependency of the accuracy on the number of training data and with or without the GC loss. The CycleGAN was trained with datasets of different sizes, i) 20 MR and 20 CT volumes, ii) 302 MR and 613 CT volumes, and both with and without GC loss. We conducted two experiments. The first experiment used three sets of paired MR and CT volumes of the same patient for test data. Because availability of paired MR and CT volumes was limited, we conducted the second experiment in which unpaired 10 MR and 20 CT volumes were used.

In the first experiment, we evaluated synthesized CT by means of mean absolute error (MAE) and peak-signal-to-noise ratio (PSNR) [dB] between synthesized CT and ground truth CT, both of which were normalized as mentioned in Section 4.2. The ground truth CT here is a CT registered to the MR of the same patient. CT and MR volumes were aligned using landmark-based registration as initialization, and then aligned using rigid and non-rigid registration. The results of MAE and PSNR are shown in Table 4.1. PSNR is calculated as $PSNR = 20 \log_{10} \frac{255}{\sqrt{MSE}}$,

Table 4.1: Mean absolute error (MAE) and Peak-signal-to-noise ratio (PSNR) between synthesized and real CT volumes.

		20 volumes		>300 volumes	
		w/o GC	/w GC	w/o GC	/w GC
MAE	Patient #1	30.121	30.276	26.899	26.388
	Patient #2	26.927	26.911	22.319	21.593
	Patient #3	33.651	32.155	29.630	28.643
	Average ± SD	30.233 ± 2.177	29.781 ± 1.777	26.283 ± 1.367	25.541 ± 1.129
PSNR	Patient #1	14.797	14.742	15.643	15.848
	Patient #2	15.734	15.628	17.255	17.598
	Patient #3	14.510	14.820	15.674	15.950
	Average ± SD	15.014 ± 0.330	15.063 ± 0.380	16.190 ± 0.273	16.465 ± 0.296

where MSE is mean squared error. The average of MAE decreased and PSNR increased according to the increase of training data size and inclusion of GC loss, respectively. Fig 4.3 shows representative results.

In the second experiment, we tested with unpaired 10 MR and 20 CT volumes. Mutual information (MI) between synthesized CT and original MR was used for evaluation when the paired ground truth was not available. The quantitative results are show in Fig.4.4(a). The left side is the box and whisker plots of the mean of each slice of MI between real CT and synthesized MR (i.e., 20 data points in total). The right side is the mean of MI between real MR and synthesized CT (i.e., 10 data points in total). The result shows that the larger number of training data yielded statistically significant improvement ($p < 0.01$) according to the paired t-test in MI. The GC loss also leads to an increase in MI between MR and synthesized CT ($p < 0.01$). Fig.4.4(b) and Fig.4.5 show examples of the visualization of real MR and synthesized CT volumes. As indicated by arrows, we can see that synthesized volumes with GC loss preserved the shape near the femoral head and adductor muscles.
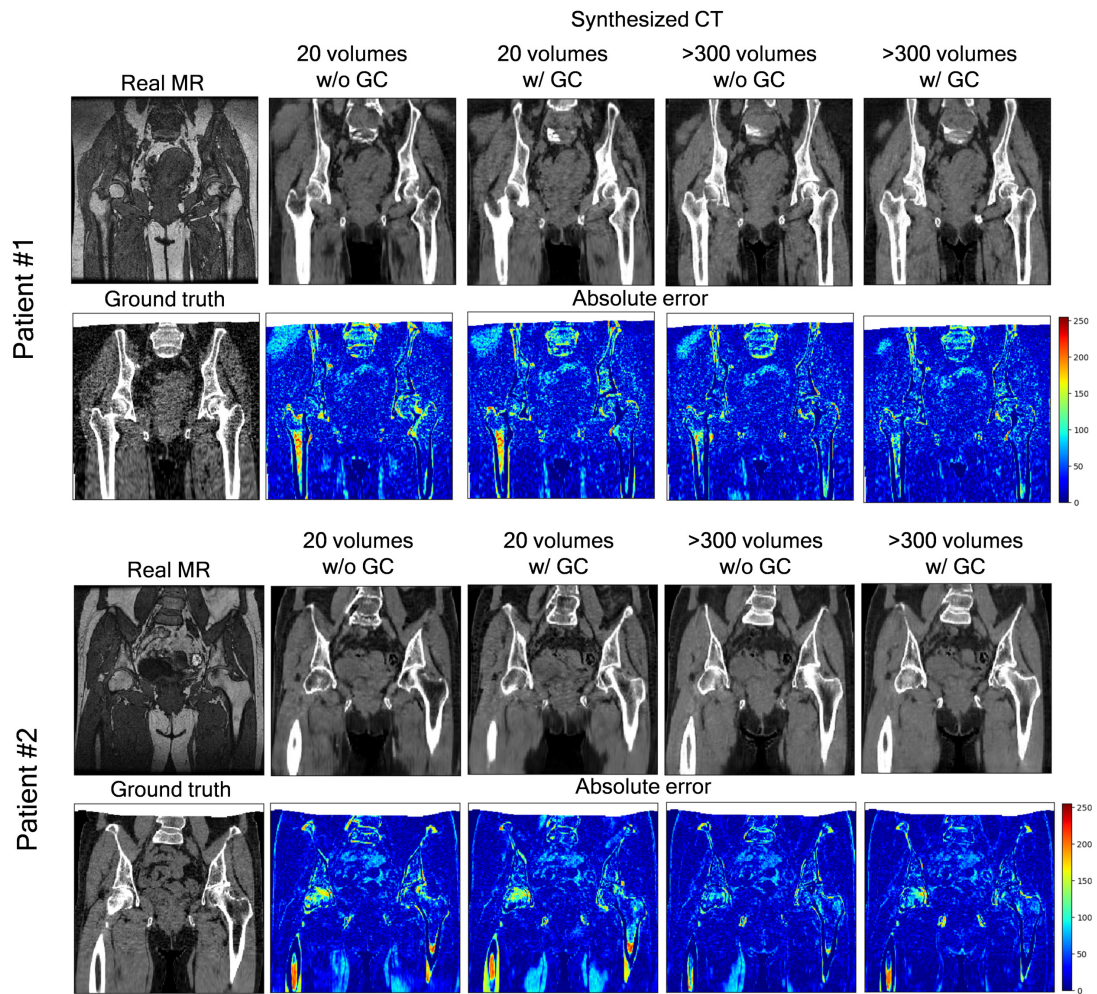
Figure 4.3: Representative results of the absolute error between the ground truth paired CT and synthesized CT from two patients. Since the FOV of MR and CT volumes are slightly different, there is no corresponding region near the top edge of the ground truth volumes (filled with white color). This area was not used for evaluation.

4.3.2 Quantitative evaluation on segmentation

To demonstrate the applicability of image synthesis in segmentation task, we evaluated the segmentation accuracy. Twenty labeled CT datasets were used to train the segmentation network. Then, we evaluated the segmentation accuracy
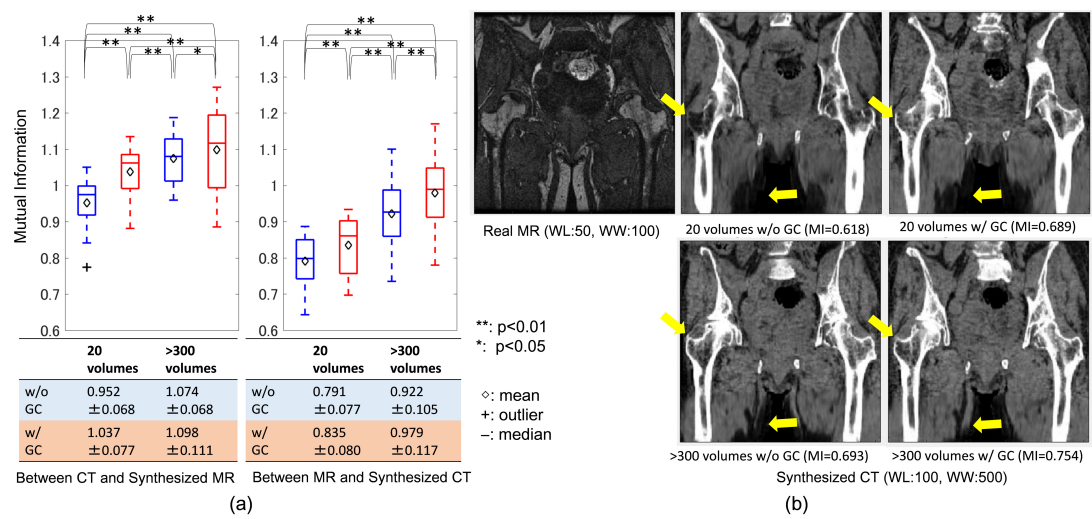
Figure 4.4: Evaluation of similarity between the real and synthesized volumes. (a) quantitative comparison of mutual information on different training data size with and without the gradient-consistency loss. (b) representative result of one patient.
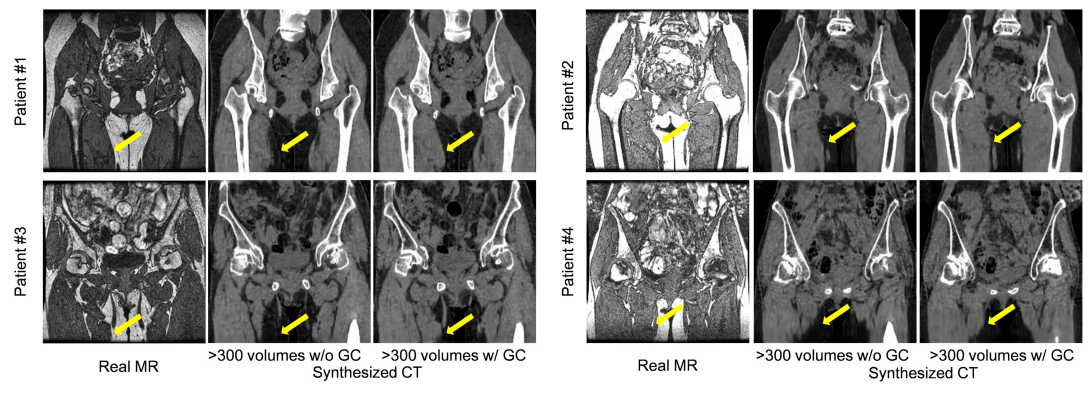
Figure 4.5: Representative results of translation from real MR to synthesized CT of four patients with and without the gradient consistency loss. As indicated by arrows, synthesized volumes with gradient consistency loss helped to preserve the shape near the adductor muscles.

with 10 MR volumes with manual segmentation labels of the gluteus medius and minimus muscles and femur.

We employed the 2D U-Net proposed by Ronneberger et al. (2015) as segmentation network, which is widely used in medical image analysis and demonstrated high performance with a limited number of labeled volumes. In MRI, muscle boundaries are clearer while bone boundaries are clearer in CT. To incorporate the advantage of both CT and MR, we modified the 2D U-Net to take the two-channel input of both CT and synthesized MR images. We trained on 2D U-Net using Adam (Kingma and Ba (2014)) for 1×10^5 iterations at learning rate of 0.0001. At the test phase, a pair of MR and synthesized CT was used as two-channel input.

The results with 4 musculoskeletal structures for 10 patients are shown in Fig.4.6 (i.e., 10 data points in total on each plot). The result shows that the larger number of training data yielded statistically significant improvement in DC on pelvis ($p < 0.01$), femur ($p < 0.01$), glutes medius ($p < 0.01$) and glutes minimus regions ($p < 0.05$) of paired t-test. The GC loss also leads to an increase in DC on the glutes minimus regions ($p < 0.01$). The average DC in the cases trained with more than 300 cases and GC loss was 0.808±0.036 (pelvis), 0.883±0.029 (femur), 0.804±0.040 (gluteus medius) and 0.669±0.054 (gluteus minimus), respectively. Fig.4.7 shows example visualization of real MR, synthesized CT, and esimated label for one patient. The result with GC loss has smoother segmentation not only in the gluteus minimus but also near the adductor muscles.

4.4 Discussion and Conclusion

In this study, we proposed an image synthesis method which extended the Cycle-GAN approach by adding the GC loss to improve the accuracy at the boundaries. Specifically, the contributions of this paper are 1) introduction of GC loss in CycleGAN, and 2) quantitative and qualitative evaluation of the dependency of both image synthesis accuracy and segmentation accuracy on a large number of
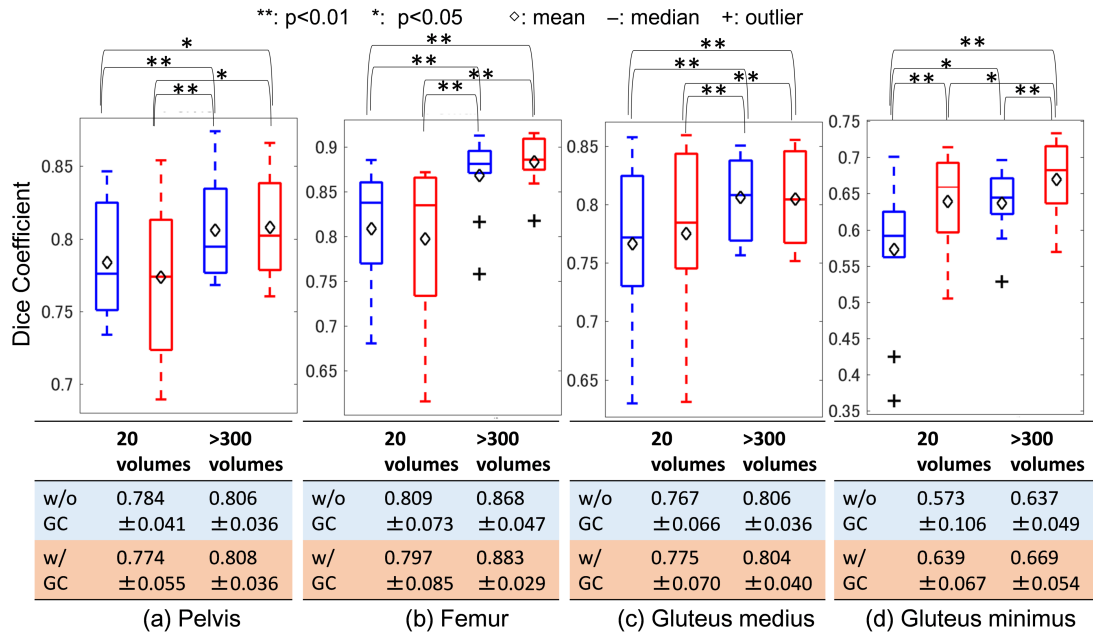
Figure 4.6: Evaluation of segmentation accuracy on different training data size in CycleGAN with and without the gradient-consistency loss. Segmentation of (a) pelvis, (b) femur, (c) gluteus medius and (d) gluteus minimus muscle in MR volumes were performed using MR-to-CT synthesis.

training data.

As a comparison against a single modality training, we performed 5-fold cross validation of MR segmentation using 10 labeled MR volumes (i.e., trained with 8 MR volumes and tested on remaining 2 MR volumes) using U-Net segmentation network. The DC was 0.815±0.046 (pelvis), 0.921±0.023 (femur), 0.825±0.029 (gluteus medius) and 0.752±0.045 (gluteus minimus), respectively. We found the gap of accuracy between modality independent and dependent segmentation. A potential improvement of modality independent segmentation is to construct an end-to-end network that performs image synthesis and segmentation (Huo et al. (2017)).

Since CycleGAN translate CT and MRI bi-directionally, it is also possible to segment musculoskeletal structures directly from real MRI by U-Net trained
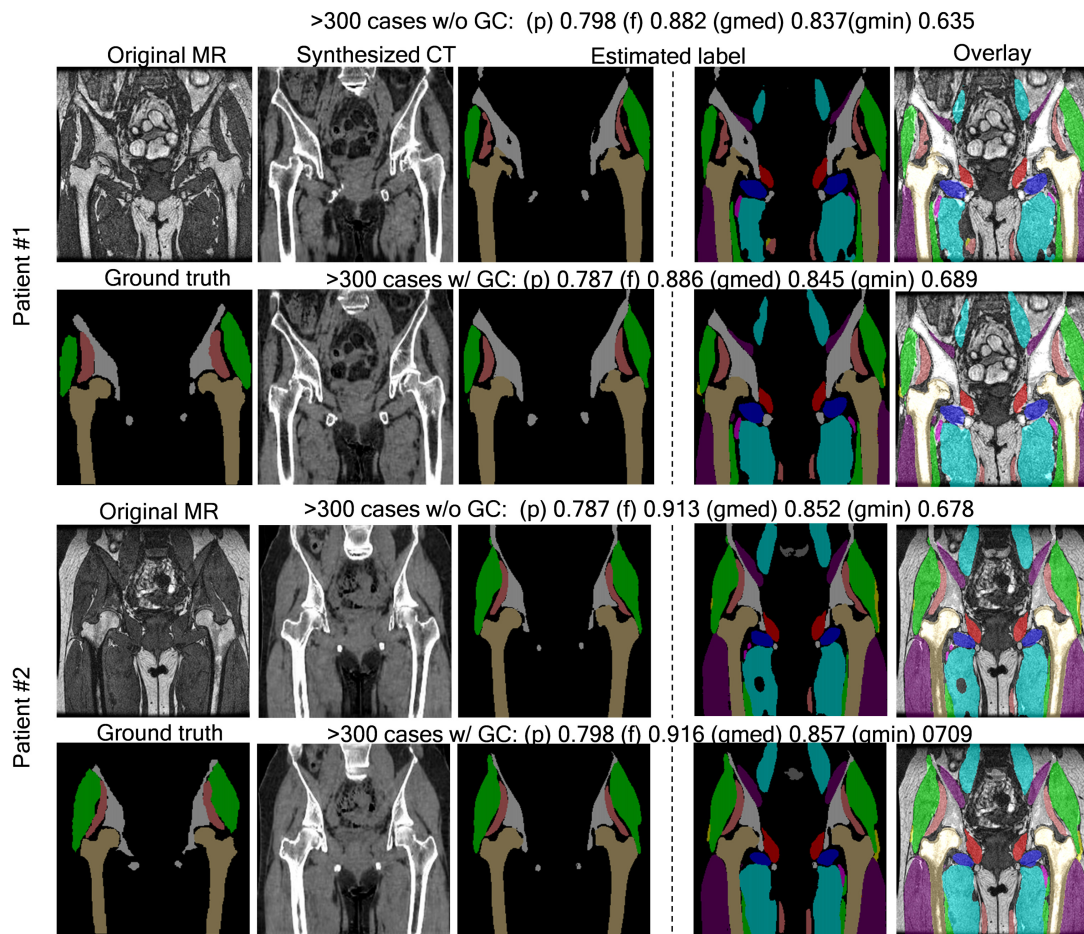
Figure 4.7: Representative results of segmentation from one patient. The ground truth label is consist of 4 musculoskeletal structures in MRI. Although we evaluated only on 4 structures because ground truth were not available for the other structures on MRI, all 22 estimated labels are shown for qualitative evaluation. In the right-most column, all estimated labels are overlayed on the real MRI. p, f, gmed, gmin denote DC of pelvis, femur, gluteus medius, and gluteus minimus, respectively.

with the synthesized MRI. Another approach for segmentation directly from real MRI is active-learning method. For active-learning, we apply the Bayesian deep networks pre-trained in Chapter 2 on synthesized CT, to obtain uncertainty as

77

well as predicted label for real MRI. Using the uncertainty, we can efficiently expand MRI labeled dataset.

One limitation in this study is that we excluded the patients with implants, while our target cohort (i.e., THA patients) sometime has implant on one side, for example, in case of the planning of secondary surgery.

Our future work also includes development of a method that effectively incorporates information in unlabeled CT and MR volumes to improve segmentation accuracy (Zhang et al. (2017)).

5 Conclusion

In this thesis, we tackled on automatic recovery of patient-patient anatomy and kinematics from multi-modal clinical images, especially three challenges: automatic recovery of 1) musculoskeletal anatomy, 2) musculoskeletal kinematics, and 3) from different modalities.

In Chapter 2, we proposed an automatic segmentation method for hip and thigh muscles from CT images using CNN. We also proposed active-learning method using Bayesian deep networks to tackle a bottleneck in the construction of labeled training dataset. We evaluated the proposed method using the 20 clinical CTs.

In Chapter 3, we proposed a 2D-3D registration method for recovery of three-dimensional rib motion from x-ray video. We enhanced previous algorithm to estimate relative poses of multiple rigid objects connected by joints such as the rib cage. We evaluated the proposed method through simulation experiments using 6 two-time-phase CTs (i.e., inhale and exhale phases) and real-image experiments using 8 x-ray videos acquired with a protocol used in a routine clinical setting.

In Chapter 4, we proposed an image synthesis method to realize the modality independent segmentation. We extended the CycleGAN approach by adding the gradient consistency loss to encourage edge alignment between images in the two domains. We evaluated the proposed method using CT dataset consists of 613 unlabeled and 20 labeled images and MRI dataset consists of 302 unlabeled and 10 labeled MR images.

In the scope of our future work includes both image synthesis and registration between structural information (e.g., CT and MRI) and functional information (e.g., Scintigraphy).

Acknowledgements

References

Andrews, S., Hamarneh, G., 2015. The generalized log-ratio transformation: learning shape and adjacency priors for simultaneous thigh muscle segmentation. IEEE transactions on medical imaging 34 (9), 1773–1787.

Beillas, P., Lafon, Y., Smith, F. W., 2009. The effects of posture and subject-to-subject variations on the position, shape and volume of abdominal and thoracic organs. Stapp car crash journal 53, 127.

Besl, P. J., McKay, N. D., et al., 1992. A method for registration of 3-d shapes. IEEE Transactions on pattern analysis and machine intelligence 14 (2), 239–256.

Beyer, B., Sholukha, V., Dugailly, P. M., Rooze, M., Moiseev, F., Feipel, V., Jan, S. V. S., 2014. In vivo thorax 3d modelling from costovertebral joint complex kinematics. Clinical Biomechanics 29 (4), 434–438.

Bruno, A. G., Bouxsein, M. L., Anderson, D. E., 2015. Development and validation of a musculoskeletal model of the fully articulated thoracolumbar spine and rib cage. Journal of biomechanical engineering 137 (8), 081003.

Chen, H.-C., Wu, C.-H., Wang, C.-K., Lin, C.-J., Sun, Y.-N., 2014. A joint-constraint model-based system for reconstructing total knee motion. IEEE Transactions on Biomedical Engineering 61 (1), 171–181.

Chen, X., Graham, J., Hutchinson, C., Muir, L., 2013. Automatic inference and

measurement of 3d carpal bone kinematics from single view fluoroscopic sequences. IEEE transactions on medical imaging 32 (2), 317–328.

Chèze, Laurence, Moissenet, Florent, Dumas, Raphaël, 2015. State of the art and current limits of musculo-skeletal models for clinical applications. Mov Sport Sci/Sci Mot (90), 7–17.
URL https://doi.org/10.1051/sm/2012026

Culham, E. G., Jimenez, H. A., King, C. E., 1994. Thoracic kyphosis, rib mobility, and lung volumes in normal women and women with osteoporosis. Spine 19 (11), 1250–1255.

Cvitanic, O., Henzie, G., Skezas, N., Lyons, J., Minter, J., 2004. MRI diagnosis of tears of the hip abductor tendons (gluteus medius and gluteus minimus). American Journal of Roentgenology 182 (1), 137–143.

De Troyer, A., Kirkwood, P. A., Wilson, T. A., 2005. Respiratory action of the intercostal muscles. Physiological Reviews 85 (2), 717–756.

Delp, S. L., Anderson, F. C., Arnold, A. S., Loan, P., Habib, A., John, C. T., Guendelman, E., Thelen, D. G., 2007. Opensim: open-source software to create and analyze dynamic simulations of movement. IEEE transactions on biomedical engineering 54 (11), 1940–1950.

Dice, L. R., 1945. Measures of the amount of ecologic association between species. Ecology 26 (3), 297–302.

Didier, A.-L., Villard, P.-F., Saadé, J., Moreau, J.-M., Beuve, M., Shariat, B., 2009. A chest wall model based on rib kinematics. In: Visualisation, 2009. VIZ'09. Second International Conference in. IEEE, pp. 159–164.

Dworzak, J., Lamecker, H., von Berg, J., Klinder, T., Lorenz, C., Kainmüller, D., Seim, H., Hege, H.-C., Zachow, S., 2010. 3d reconstruction of the human rib

cage from 2d projection images using a statistical shape model. International journal of computer assisted radiology and surgery 5 (2), 111–124.

Eaton-Rosen, Z., Bragman, F., Bisdas, S., Ourselin, S., Cardoso, M. J., 2018. Towards safe deep learning: accurately quantifying biomarker uncertainty in neural network predictions. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 691–699.

Fleming, A. D., Philip, S., Goatman, K. A., Olson, J. A., Sharp, P. F., 2006. Automated microaneurysm detection using local contrast normalization and local vessel detection. IEEE Transactions on Medical Imaging 25 (9), 1223–1232.

Gal, Y., Ghahramani, Z., 2016. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In: international conference on machine learning. pp. 1050–1059.

Gendrin, C., Furtado, H., Weber, C., Bloch, C., Figl, M., Pawiro, S. A., Bergmann, H., Stock, M., Fichtinger, G., Georg, D., et al., 2012. Monitoring tumor motion by real time 2d/3d registration during radiotherapy. Radiotherapy and oncology 102 (2), 274–280.

Gill, S., Abolmaesumi, P., Fichtinger, G., Boisvert, J., Pichora, D., Borshneck, D., Mousavi, P., 2012. Biomechanically constrained groupwise ultrasound to ct registration of the lumbar spine. Medical image analysis 16 (3), 662–674.

Gilles, B., Magnenat-Thalmann, N., 2010. Musculoskeletal MRI segmentation using multi-resolution simplex meshes with medial representations. Medical image analysis 14 (3), 291–302.

Gilmartin, J., Gibson, G., 1986. Mechanisms of paradoxical rib cage motion in patients with chronic obstructive pulmonary disease. American Review of Respiratory Disease 134 (4), 683–687.

Grenier, S., Parent, S., Cheriet, F., 2013. Personalized 3d reconstruction of the rib cage for clinical assessment of trunk deformities. Medical engineering & physics 35 (11), 1651–1658.

Hamarneh, G., Jassi, P., Tang, L., 2008. Simulation of ground-truth validation data via physically-and statistically-based warps. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 459–467.

Handsfield, G. G., Meyer, C. H., Hart, J. M., Abel, M. F., Blemker, S. S., 2014. Relationships of 35 lower limb muscles to height and body mass quantified using mri. Journal of biomechanics 47 (3), 631–638.

Hansen, N., 2006. The cma evolution strategy: a comparing review. In: Towards a new evolutionary computation. Springer, pp. 75–102.

He, K., Zhang, X., Ren, S., Sun, J., 2015. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In: Proceedings of the IEEE international conference on computer vision. pp. 1026–1034.

Heeger, D. J., 1992. Normalization of cell responses in cat striate cortex. Visual neuroscience 9 (02), 181–197.

Hiasa, Y., Otake, Y., Nakatani, S., Harada, H., Kanaji, S., Kakeji, Y., Sato, Y., 2018a. Surgical tools segmentation in laparoscopic images using convolutional neural networks with uncertainty estimation and semi-supervised learning. In: CARS 2018—Computer Assisted Radiology and Surgery Proceedings of the 32nd International Congress and Exhibition. pp. 14–15.

Hiasa, Y., Otake, Y., Takao, M., Matsuoka, T., Takashima, K., Carass, A., Prince, J. L., Sugano, N., Sato, Y., 2018b. Cross-modality image synthesis from unpaired data using cyclegan. In: International Workshop on Simulation and Synthesis in Medical Imaging. Springer, pp. 31–41.

Huo, Y., Xu, Z., Bao, S., Assad, A., Abramson, R. G., Landman, B. A., 2017. Adversarial synthesis learning enables segmentation without target modality ground truth. arXiv preprint arXiv:1712.07695.

Isola, P., Zhu, J.-Y., Zhou, T., Efros, A. A., 2017. Image-to-image translation with conditional adversarial networks. arXiv preprint.

Ito, H., Koshizuka, S., Akihiro, H., Keiichi, N., 2011. Rib cage motion model construction based on patient-specific ct images between inhalation and exhalation. Medical Imaging Technology 29 (4), 208–214.

Jarrett, K., Kavukcuoglu, K., LeCun, Y., et al., 2009. What is the best multi-stage architecture for object recognition? In: Computer Vision, 2009 IEEE 12th International Conference on. IEEE, pp. 2146–2153.

Johnson, J., Alahi, A., Fei-Fei, L., 2016. Perceptual losses for real-time style transfer and super-resolution. In: European Conference on Computer Vision. Springer, pp. 694–711.

Kamnitsas, K., Baumgartner, C., Ledig, C., Newcombe, V., Simpson, J., Kane, A., Menon, D., Nori, A., Criminisi, A., Rueckert, D., et al., 2017. Unsupervised domain adaptation in brain lesion segmentation with adversarial networks. In: International Conference on Information Processing in Medical Imaging. Springer, pp. 597–609.

Karlsson, A., Rosander, J., Romu, T., Tallberg, J., Grönqvist, A., Borga, M., Dahlqvist Leinhard, O., 2015. Automatic and quantitative assessment of regional muscle volume by multi-atlas segmentation using whole-body water–fat mri. Journal of Magnetic Resonance Imaging 41 (6), 1558–1569.

Kendall, A., Badrinarayanan, V., Cipolla, R., 2015. Bayesian segnet: Model uncertainty in deep convolutional encoder-decoder architectures for scene understanding. arXiv preprint arXiv:1511.02680.

Kendall, A., Gal, Y., 2017. What uncertainties do we need in bayesian deep learning for computer vision? In: Advances in neural information processing systems. pp. 5574–5584.

Kingma, D. P., Ba, J., 2014. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.

Lamecker, H., Wenckebach, T. H., Hege, H.-C., 2006. Atlas-based 3d-shape reconstruction from x-ray images. In: Pattern Recognition, 2006. ICPR 2006. 18th International Conference on. Vol. 1. IEEE, pp. 371–374.

Le, Q. V., 2013. Building high-level features using large scale unsupervised learning. In: Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on. IEEE, pp. 8595–8598.

Lee, H., Troschel, F. M., Tajmir, S., Fuchs, G., Mario, J., Fintelmann, F. J., Do, S., 2017. Pixel-level deep segmentation: artificial intelligence quantifies muscle on computed tomography for body morphometric analysis. Journal of digital imaging 30 (4), 487–498.

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

Long, J., Shelhamer, E., Darrell, T., 2015. Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 3431–3440.

Lunacek, M., Whitley, D., 2006. The dispersion metric and the cma evolution strategy. In: Proceedings of the 8th annual conference on Genetic and evolutionary computation. ACM, pp. 477–484.

Maier-Hein, L., Ross, T., Gröhl, J., Glocker, B., Bodenstedt, S., Stock, C., Heim, E., Götz, M., Wirkert, S., Kenngott, H., et al., 2016. Crowd-algorithm collab-

oration for large-scale endoscopic image annotation with confidence. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 616–623.

Mao, X., Li, Q., Xie, H., Lau, R. Y., Wang, Z., 2016. Multi-class generative adversarial networks with the L2 loss function. CoRR, abs/1611.04076 2.

Markelj, P., Tomaževič, D., Likar, B., Pernuš, F., 2012. A review of 3d/2d registration methods for image-guided interventions. Medical image analysis 16 (3), 642–661.

Moissenet, F., Modenese, L., Dumas, R., 2017. Alterations of musculoskeletal models for a more accurate estimation of lower limb joint contact forces during normal gait: A systematic review. JOURNAL OF BIOMECHANICS 63, 8–20. URL http://dx.doi.org/10.1016/j.jbiomech.2017.08.025

Murai, A., Endo, Y., Tada, M., 2016. Anatomographic volumetric skin-musculoskeletal model and its kinematic deformation with surface-based ssd. IEEE Robotics and Automation Letters 1 (2), 1103–1109.

Murphy, K., Van Ginneken, B., Reinhardt, J. M., Kabus, S., Ding, K., Deng, X., Cao, K., Du, K., Christensen, G. E., Garcia, V., et al., 2011. Evaluation of registration methods on thoracic ct: the empire10 challenge. IEEE transactions on medical imaging 30 (11), 1901–1920.

Nair, T., Precup, D., Arnold, D. L., Arbel, T., 2018. Exploring uncertainty measures in deep networks for multiple sclerosis lesion detection and segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 655–663.

Otake, Y., Armand, M., Armiger, R. S., Kutzer, M. D., Basafa, E., Kazanzides, P., Taylor, R. H., 2012. Intraoperative image-based multiview 2d/3d registration for image-guided orthopaedic surgery: incorporation of fiducial-based

c-arm tracking and gpu-acceleration. IEEE transactions on medical imaging 31 (4), 948–962.

Otake, Y., Takao, M., Fukuda, N., Takagi, S., Yamamura, N., Sugano, N., Sato, Y., 2018. Registration-based patient-specific musculoskeletal modeling using high fidelity cadaveric template model. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 703–710.

Otake, Y., Wang, A. S., Uneri, A., Kleinszig, G., Vogt, S., Aygun, N., Shengfu, L. L., Wolinsky, J.-P., Gokaslan, Z. L., Siewerdsen, J. H., 2015. 3d–2d registration in mobile radiographs: algorithm development and preliminary clinical evaluation. Physics in medicine and biology 60 (5), 2075.

Otake, Y., Wang, A. S., Webster Stayman, J., Uneri, A., Kleinszig, G., Vogt, S., Khanna, A. J., Gokaslan, Z. L., Siewerdsen, J. H., 2013. Robust 3d-2d image registration: application to spine interventions and vertebral labeling in the presence of anatomical deformation. Physics in medicine and biology 58 (23), 8535–8553.

Penney, G. P., Weese, J., Little, J. A., Desmedt, P., Hill, D. L., et al., 1998. A comparison of similarity measures for use in 2-d-3-d medical image registration. IEEE transactions on medical imaging 17 (4), 586–595.

Pflesser, B., Petersik, A., Pommert, A., Riemer, M., Schubert, R., Tiede, U., Hohne, K., Schumacher, U., Richter, E., 2001. Exploring the visible human's inner organs with the voxel-man 3d navigator. Studies in health technology and informatics, 379–385.

Prince, J. L., Links, J. M., 2014. Medical imaging signals and systems (2nd Edition). Pearson Prentice Hall Upper Saddle River, New Jersey.

Rajagopal, A., Dembia, C. L., DeMers, M. S., Delp, D. D., Hicks, J. L., Delp, S. L., 2016. Full-body musculoskeletal model for muscle-driven simulation of human gait. IEEE Trans. Biomed. Engineering 63 (10), 2068–2079.

Ranzini, M. B. M., Ebner, M., Cardoso, M. J., Fotiadou, A., Vercauteren, T., Henckel, J., Hart, A., Ourselin, S., Modat, M., 2017. Joint multimodal segmentation of clinical CT and MR from hip arthroplasty patients. In: International Workshop and Challenge on Computational Methods and Clinical Applications in Musculoskeletal Imaging. Springer, pp. 72–84.

Rasch, A., Byström, A., Dalen, N., Martinez-Carranza, N., Berg, H., 2009. Persisting muscle atrophy two years after replacement of the hip. The Journal of bone and joint surgery. British volume 91 (5), 583–588.

Rasoulian, A., Rohling, R., Abolmaesumi, P., 2013. Lumbar spine segmentation using a statistical multi-vertebrae anatomical shape+ pose model. IEEE transactions on medical imaging 32 (10), 1890–1900.

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

Seth, A., Hicks, J. L., Uchida, T. K., Habib, A., Dembia, C. L., Dunne, J. J., Ong, C. F., DeMers, M. S., Rajagopal, A., Millard, M., et al., 2018. Opensim: Simulating musculoskeletal dynamics and neuromuscular control to study human and animal movement. PLoS computational biology 14 (7), e1006223.

Shu, L., Yamamoto, K., Yao, J., Saraswat, P., Liu, Y., Mitsuishi, M., Sugita, N., 2018. A subject-specific finite element musculoskeletal framework for mechanics analysis of a total knee replacement. Journal of biomechanics 77, 146–154.

Spitzer, V. M., Whitlock, D. G., 1998. The visible human dataset: the anatomical platform for human simulation. The Anatomical Record: An Official Publication of the American Association of Anatomists 253 (2), 49–57.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R., 2014. Dropout: a simple way to prevent neural networks from overfitting. The Journal of Machine Learning Research 15 (1), 1929–1958.

Styner, M., Lee, J., Chin, B., Chin, M., Commowick, O., Tran, H., Markovic-Plese, S., Jewells, V., Warfield, S., 2008. 3d segmentation in the clinic: A grand challenge ii: Ms lesion segmentation. Midas Journal 2008, 1–6.

Taddei, F., Martelli, S., Valente, G., Leardini, A., Benedetti, M. G., Manfrini, M., Viceconti, M., 2012. Femoral loads during gait in a patient with massive skeletal reconstruction. Clinical Biomechanics 27 (3), 273–280.

Talbi, E.-G., 2009. Metaheuristics: from design to implementation. Vol. 74. John Wiley & Sons.

Tanaka, R., 2016. Dynamic chest radiography: flat-panel detector (fpd) based functional x-ray imaging. Radiological physics and technology 9 (2), 139–153.

Tanaka, R., Sanada, S., Sakuta, K., Kawashima, H., 2015. Quantitative analysis of rib kinematics based on dynamic chest bone images: preliminary results. Journal of Medical Imaging 2 (2), 024002–024002.

Torrado-Carvajal, A., Herraiz, J. L., Alcain, E., Montemayor, A. S., Garcia-Cañamaque, L., Hernandez-Tamames, J. A., Rozenholc, Y., Malpica, N., 2016. Fast patch-based pseudo-CT synthesis from T1-weighted MR images for PET/MR attenuation correction in brain studies. Journal of Nuclear Medicine 57 (1), 136–143.

Tustison, N. J., Avants, B. B., Cook, P. A., Zheng, Y., Egan, A., Yushkevich, P. A., Gee, J. C., 2010. N4ITK: improved N3 bias correction. IEEE transactions on medical imaging 29 (6), 1310–1320.

Uemura, K., Takao, M., Sakai, T., Nishii, T., Sugano, N., 2016. Volume increases of the gluteus maximus, gluteus medius, and thigh muscles after hip arthroplasty. The Journal of arthroplasty 31 (4), 906–912.

Ulbrich, Erika J, N. D. L. O. D. M. M. F. M. A., 2018. Whole-body adipose tissue and lean muscle volumes and their distribution across gender and age:

Mr-derived normative values in a normal-weight swiss population. Magnetic resonance in medicine 79, 449–458.

Valindria, V. V., Lavdas, I., Bai, W., Kamnitsas, K., Aboagye, E. O., Rockall, A. G., Rueckert, D., Glocker, B., 2017. Reverse classification accuracy: Predicting segmentation performance in the absence of ground truth. IEEE Transactions on Medical Imaging 36 (8), 1597–1606.

Vallières, M., Freeman, C. R., Skamene, S. R., El Naqa, I., 2015. A radiomics model from joint fdg-pet and mri texture features for the prediction of lung metastases in soft-tissue sarcomas of the extremities. Physics in Medicine & Biology 60 (14), 5471.

van de Kraats, E. B., Penney, G. P., Tomazevic, D., Van Walsum, T., Niessen, W. J., 2005. Standardized evaluation methodology for 2-d-3-d registration. IEEE Transactions on Medical Imaging 24 (9), 1177–1189.

Van Uitert, R., Bitter, I., 2007. Subvoxel precise skeletons of volumetric data based on fast marching methods. Medical physics 34 (2), 627–638.

Veldkamp, W. J., Karssemeijer, N., 2000. Normalization of local contrast in mammograms. IEEE Transactions on Medical Imaging 19 (7), 731–738.

Villard, P.-F., Escamilla, P., Kerrien, E., Gorges, S., Trousset, Y., Berger, M.-O., 2014. Preliminary study of rib articulated model based on dynamic fluoroscopy images. In: SPIE Medical Imaging. p. 90361Y.

Webb, J. D., Blemker, S. S., Delp, S. L., 2014. 3d finite element models of shoulder muscles for computing lines of actions and moment arms. Computer methods in biomechanics and biomedical engineering 17 (8), 829–837.

Werf, A., Langius, J., Schueren, M., Nurmohamed, S., Pant, K., Blauwhoff-Buskermolen, S., Wierdsma, N., 2018. Percentiles for skeletal muscle index, area and radiation attenuation based on computed tomography imaging in a

healthy caucasian population. European journal of clinical nutrition 72 (2), 288.

Wolterink, J. M., Dinkla, A. M., Savenije, M. H., Seevinck, P. R., van den Berg, C. A., Išgum, I., 2017. Deep MR to CT synthesis using unpaired data. In: International Workshop on Simulation and Synthesis in Medical Imaging. Springer, pp. 14–23.

Yamamura, N., Alves, J. L., Oda, T., Kinugasa, R., Takagi, S., 2014. Effect of tendon stiffness on the generated force at the achilles tendon-3d finite element simulation of a human triceps surae muscle during isometric contraction. Journal of Biomechanical Science and Engineering 9 (3), 13–00294.

Yang, L., Zhang, Y., Chen, J., Zhang, S., Chen, D. Z., 2017. Suggestive annotation: A deep active learning framework for biomedical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 399–407.

Yokota, F., Okada, T., Takao, M., Sugano, N., Tada, Y., Tomiyama, N., Sato, Y., 2013. Automated ct segmentation of diseased hip using hierarchical and conditional statistical shape models. InInternational Conference on Medical Image Computing and Computer-Assisted Intervention, 190–197.

Yokota, F., Otake, Y., Takao, M., Ogawa, T., Okada, T., Sugano, N., Sato, Y., 2018. Automated muscle segmentation from ct images of the hip and thigh using a hierarchical multi-atlas method. International journal of computer assisted radiology and surgery, 1–10.

Zhang, Y., Yang, L., Chen, J., Fredericksen, M., Hughes, D. P., Chen, D. Z., 2017. Deep adversarial networks for biomedical image segmentation utilizing unannotated images. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 408–416.

Zhao, C., Carass, A., Lee, J., He, Y., Prince, J. L., 2017. Whole brain segmentation and labeling from CT using synthetic MR images. In: International Workshop on Machine Learning in Medical Imaging. Springer, pp. 291–298.

Zhu, J.-Y., Park, T., Isola, P., Efros, A. A., 2017. Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 2223–2232.

Publication List

Journal

[1a] M. Nishi, S. Kanaji, Y. Otake, H. Harada, M. Yamamoto, T. Oshikiri, T. Nakamura, S. Suzuki, Y. Suzuki, <u>Y. Hiasa</u>, Y. Sato and Y. Kakeji, "Quantitative comparison of operative skill using 2-and 3-dimensional monitors during laparoscopic phantom tasks", Surgery, vol. 161, no. 5, pp. 1334-1340, 2017.

[2a] <u>Y. Hiasa</u>, Y. Otake, R. Tanaka, S. Sanada and Y. Sato, "Recovery of 3D rib motion from dynamic chest radiography and CT data using local contrast normalization and articular motion model", Medical Image Analysis, vol. 51, pp. 144-156, 2019.

International Conference (Refereed)

[1b] <u>Y. Hiasa</u>, Y. Otake, R. Tanaka, F. Yokota, S. Sanada and Y. Sato, "Constrained piecewise rigid 2D-3D registration for patient-specific analysis of rib cage motion using X-ray video", CARS 2016-Computer Assisted Radiology and Surgery Proceedings of the 30th International Congress and Exhibition, vol. 11, pp. 52-53, 2016.

[2b] K. Koyama, Y. Otake, K. Uemura, M. Takao, T. Ogawa, <u>Y. Hiasa</u>, F. Yokota, N. Sugano and Y. Sato, An automated pipeline of bone segmentation and 2D-3D registration on hip radiology graphs: Towards large-scale population study on hip joint biomechanics, The 16th Annual Meeting of the International Society for Computer Assisted Orthopaedic Surgery (CAOS), June, 2016.

[3b] K. Koyama, Y. Otake, K. Uemura, M. Takao, T. Ogawa, <u>Y. Hiasa</u>, F. Yokota, N. Sugano and Y. Sato, Analysis of Pelvis and Femur Alignment in

the Standing Position using Automated 2D-3D Registration: Toward Large-scale Population Analysis, The International Forum on Medical Imaging in Asia (IFMIA), January, 2017.

[4b] <u>Y. Hiasa</u>, 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, vol. 11037, pp. 31-41, 2018.

[5b] <u>Y. Hiasa</u>, Y. Otake, S. Nakatani, H. Harada, S. Kanaji, Y. Kakeji and Y. Sato, "Surgical tools segmentation in laparoscopic images using convolutional neural networks with uncertainty estimation and semi-supervised learning", CARS 2018-Computer Assisted Radiology and Surgery Proceedings of the 32nd International Congress and Exhibition, vol. 13, pp. 14-15, 2018.

[6b] A. Jodeiri, Y. Otake, R. A. Zoroofi, <u>Y. Hiasa</u>, M. Takao, K. Uemura, N. Sugano and Y. Sato, "Estimation of Pelvic Sagital Inclanation from Anteroposterior Radiograph Using Convolutional Neural Networks: Proof-of-Concept Study", The 18th Annual Meeting of the International Society for Computer Assisted Orthopaedic Surgery (CAOS), vol. 2, pp. 114-118, 2018. [Best Technical Paper Award]

[7b] M. Sakamoto, <u>Y. Hiasa</u>, Y. Otake, M. Takao, Y. Suzuki, N. Sugano and Y. Sato, "Automated Segmentation of Hip and Thigh Muscles in Metal Artifact Contaminated CT using CNN", The International Forum on Medical Imaging in Asia (IFMIA), 2019 (in press). [Best Paper Award]

Domestic Conference

[1c] <u>日朝祐太</u>, 大竹義人, 田中利恵, 真田茂, 佐藤嘉伸, "拘束条件付き2D-3Dレジストレーションを用いたX線動画像からの胸郭動態の計測", 電子情報通信学会, 医用画像研究会, vol. 115, no. 301, pp. 61-65, 2015年.

[2c] 児山昂生, 大竹義人, 上村圭亮, 日朝祐太, 横田太, 高尾正樹, 小川剛, 菅野伸彦, 佐藤嘉伸, "大規模患者データベースにおける骨盤傾斜角測定を目的とした2D-3Dレジストレーション手法の検討", 電子情報通信学会, 医用画像研究会, vol. 115, no. 401, pp. 331-336, 2016年.

[3c] 日朝祐太, 大竹義人, 児山昂生, 上村圭亮, 高尾正樹, 小川剛, 菅野伸彦, 佐藤嘉伸, "畳み込みニューラルネットワークを用いた2D-3Dレジストレーションの自動失敗検出 〜大規模股関節患者データベースにおける全自動股関節動態解析に向けて〜", 電子情報通信学会, 医用画像研究会, vol. 116, no. 160, pp. 43-48, 2016年.

[4c] 日朝祐太, 鈴木裕紀, Austin Reiter, 大竹義人, 西将康, 原田仁, 金治新悟, 掛地吉弘, 佐藤嘉伸, "Segmentation of Surgical Instruments from RGB-D Endoscopic Images using Convolutional Neural Networks: Preliminary Experiments towards Quantitative Skill Assessment", 生体医用画像研究会, 第3回若手発表会 抄録集, pp. 7, 2016年.

[5c] 日朝祐太, 大竹義人, 田中利恵, 真田茂, 佐藤嘉伸, "一軸性関節による肋骨運動の制約付き2D-3D位置合わせ精度の評価", 電子情報通信学会, 医用画像研究会, vol. 116, no. 393, pp. 7-12, 2017年.

[6c] 宮本康平, 大竹義人, 横田太, 日朝祐太, Min Suk Chung, 高尾正樹, 菅野伸彦, 佐藤嘉伸, " CT画像を用いた股関節周辺骨格筋の線維走行解析", 電子情報通信学会, 医用画像研究会, vol. 116, no. 393, pp. 127-131, 2017年.

[7c] Yusuke Tenma, Yuta Hiasa, Yoshito Otake, Shingo Abe, Hirokazu Kato, Tsuyoshi Murase, Yoshinobu Sato, "Kinematic analysis of forearm rotation by 2D-3D registration using biplane fluoroscopy", 生体医用画像研究会, 第4回若手発表会 抄録集, pp. 15, 2017年.

[8c] Satoshi Nakatani, Yuta Hiasa, Yoshito Otake, Hitoshi Harada, Shingo Kanaji, Yoshihiro Kakeji, Yoshinobu Sato, "Segmentation of stereo endo-

scopic video images using super-pixel and disparity map in laparoscopic surgery", 生体医用画像研究会, 第4回若手発表会 抄録集, pp. 1, 2017年.

[9c] 日朝祐太, 中谷聡志, 大竹義人, 原田仁, 金治新悟, 掛地吉弘, 佐藤嘉伸, "スーパーピクセルと畳み込みニューラルネットワークを用いた腹腔鏡下手術における術具領域セグメンテーション", 第36回日本医用画像工学会大会, 2017年.

[10c] 小野真理子, 鈴木裕紀, 日朝祐太, 堀雅敏, 大竹義人, 富山憲幸, 佐藤嘉伸, "畳み込みニューラルネットワークを用いた腹部造影 CT 画像からの腎動脈枝の強調", 第26回日本コンピュータ外科学会, vol. 19, no. 4, pp. 265, 2017年.

[11c] 日朝祐太, 大竹義人, 横田太, 高尾正樹, 小川剛, 菅野伸彦, 佐藤嘉伸, "(ショートペーパー) 畳み込みニューラルネットワークを用いたCT画像からの股関節および大腿部の筋骨格領域自動抽出", 電子情報通信学会, 医用画像研究会, vol. 117, no. 281, pp. 3-4, 2017年.

[12c] 椛島基嵩, 日朝祐太, 大竹義人, 田中利恵, 真田茂, 佐藤嘉伸, "胸部X線画像を用いた畳み込みニューラルネットワークによる肋骨2D-3D 位置合わせの自動初期化", 電子情報通信学会, 医用画像研究会, vol. 117, no. 281, pp. 5-8, 2017年.

[13c] 天満勇介, 日朝祐太, 大竹義人, 阿部真悟, 村瀬剛, 佐藤嘉伸, "2方向X線透視動画像を用いた2D-3Dレジストレーションによる亜脱臼動作を含む前腕回旋動態解析", 生体医用画像研究会, 第5回若手発表会 抄録集, pp. 7, 2018年.

[14c] 鈴木裕紀, 小野真理子, 日朝祐太, 堀雅敏, 大竹義人, 富山憲幸, 佐藤嘉伸, "畳み込みニューラルネットワークと解剖学的知識を用いた腹部造影CT画像からの腎動脈枝の自動抽出〜ネットワーク構造の比較〜", 生体医用画像研究会, 第5回若手発表会 抄録集, pp. 4, 2018年.

[15c] 末長和馬, 鍵山善之, 中西裕紀, 日朝祐太, 横田太, 大竹義人, 高尾正樹, 菅野伸彦, 佐藤嘉伸, "人工股関節カップ自動手術計画システムにおける術後データ解析", 第12回日本CAOS研究会, 2018年.

[16c] 松岡拓未, 日朝祐太, 大竹義人, 高尾正樹, 高嶋和磨, Jerry L. Prince, 菅野伸彦, 佐藤嘉伸, "CycleGANによる異種モダリティ画像生成を用いた股関節MRIの筋骨格セグメンテーション", 第37回日本医用画像工学会大会, 2018年.

[17c] 椛島基嵩, 大竹義人, 塩出亮哉, 日朝祐太, 村瀬剛, 佐藤嘉伸, "深層学習を用いたX線投影像のみによる橈骨遠位端2D-3D再構成", 第27回コンピュータ外科学会大会, pp. 268, 2018年.

[18c] 松岡拓未, 日朝祐太, 大竹義人, 高尾正樹, 高嶋和磨, Jerry L. Prince, 菅野伸彦, 佐藤嘉伸, "CycleGANを用いたCT-マルチパラメトリックMR画像変換", 第27回コンピュータ外科学会大会, pp. 250, 2018年.

[19c] 阪本充輝, 日朝祐太, 大竹義人, 高尾正樹, 鈴木裕紀, 菅野伸彦, 佐藤嘉伸, "人工股関節全置換術の術前・術後CTからの筋骨格セグメンテーションとその手術結果解析への応用", 第27回コンピュータ外科学会大会, pp. 375, 2018年.

[20c] 日朝祐太, 大竹義人, 松岡拓未, 高尾正樹, 上村圭亮, Ata Jodeiri, Reza A. Zoroofi, 菅野伸彦, 佐藤嘉伸, "敵対的生成ネットワークによる実X線画像からの疑似X線画像生成", 第27回コンピュータ外科学会大会, pp. 270, 2018年.

[21c] 日朝祐太, 大竹義人, 高尾正樹, 菅野伸彦, 佐藤嘉伸, "CT画像からの筋骨格自動抽出〜能動学習を用いた畳み込みニューラルネットワークに関する検討〜", 第27回コンピュータ外科学会大会, pp. 376, 2018年.

[22c] 阿部真悟, 天満勇介, 日朝祐太, 岡久仁洋, 塩出亮哉, 大竹義人, 菅本一臣, 佐藤嘉伸, 村瀬剛, "Biplane Fluoroscopic Intensity Based 2D3D Registra-

tion法による前腕回旋動態解析", 第27回コンピュータ外科学会大会, pp. 267, 2018年.

[23c] 椛島基嵩, <u>日朝祐太</u>, 大竹義人, 塩出亮哉, 村瀬剛, 佐藤嘉伸, "畳み込みニューラルネットワークによるＸ線画像を用いた前腕骨遠位端三次元形状の再構成", 電子情報通信学会, 医用画像研究会, vol. 118, no. 412, pp. 235-238, 2019年.

Articles

[1d] 大竹義人, <u>日朝祐太</u>, 松岡拓未, 高尾正樹, Jerry L. Prince, 菅野伸彦, 佐藤嘉伸, "敵対的生成ネットワーク（GAN）による異種モダリティ画像生成", INNERVISION, vol. 37, no. 7, pp. 36-39, 2018年.

[2d] 大竹義人, <u>日朝祐太</u>, 高尾正樹, 菅野伸彦, 佐藤嘉伸, "医用画像処理における深層学習の応用と展望", 医療機器学, vol. 88, no. 4, pp. 490-496, 2018年.