Deep learning based high-resolution reconstruction of trabecular bone microstructures from low-resolution CT scans using GAN-CIRCLE

Guha, Indranil, Nadeem, Syed Ahmed, You, Chenyu, Zhang, Xiaoliu, Levy, Steven, et al.

Indranil Guha, Syed Ahmed Nadeem, Chenyu You, Xiaoliu Zhang, Steven M. Levy, Ge Wang, James C. Torner, Punam K. Saha, "Deep learning based high-resolution reconstruction of trabecular bone microstructures from low-resolution CT scans using GAN-CIRCLE," Proc. SPIE 11317, Medical Imaging 2020: Biomedical Applications in Molecular, Structural, and Functional Imaging, 113170U (28 February 2020); doi: 10.1117/12.2549318

SPIE.

Event: SPIE Medical Imaging, 2020, Houston, Texas, United States

Deep Learning Based High-Resolution Reconstruction of Trabecular Bone Microstructures from Low-Resolution CT Scans using GAN-CIRCLE

Indranil Guha,[a*] Syed Ahmed Nadeem,[a] Chenyu You,[c] Xiaoliu Zhang,[a] Steven M Levy,[d] Ge Wang,[e] James C Torner,[f] Punam K Saha[a,b]

[a]Department of Electrical and Computer Engineering, College of Engineering, University of Iowa, Iowa City, IA 52242
[b]Department of Radiology, Carver College of Medicine, University of Iowa, Iowa City, IA 52242
[c]Department of Computer Science, Yale University, New Haven, CT 05620
[d]Department of Preventive and Community Dentistry, College of Dentistry, University of Iowa, Iowa City, IA 52242
[e]Biomedical Imaging Center, BME/CBIS, Rensselaer Polytechnic Institute, Troy, New York, NY 12180
[f]Department of Epidemiology, University of Iowa, Iowa City, IA 52242

ABSTRACT

Osteoporosis is a common age-related disease characterized by reduced bone density and increased fracture-risk. Microstructural quality of trabecular bone (Tb), commonly found at axial skeletal sites and at the end of long bones, is an important determinant of bone-strength and fracture-risk. High-resolution emerging CT scanners enable *in vivo* measurement of Tb microstructures at peripheral sites. However, resolution-dependence of microstructural measures and wide resolution-discrepancies among various CT scanners together with rapid upgrades in technology warrant data harmonization in CT-based cross-sectional and longitudinal bone studies. This paper presents a deep learning-based method for high-resolution reconstruction of Tb microstructures from low-resolution CT scans using GAN-CIRCLE. A network was developed and evaluated using post-registered ankle CT scans of nineteen volunteers on both low- and high-resolution CT scanners. 9,000 matching pairs of low- and high-resolution patches of size 64×64 were randomly harvested from ten volunteers for training and validation. Another 5,000 matching pairs of patches from nine other volunteers were used for evaluation. Quantitative comparison shows that predicted high-resolution scans have significantly improved structural similarity index ($p < 0.01$) with true high-resolution scans as compared to the same metric for low-resolution data. Different Tb microstructural measures such as thickness, spacing, and network area density are also computed from low- and predicted high-resolution images, and compared with the values derived from true high-resolution scans. Thickness and network area measures from predicted images showed higher agreement with true high-resolution CT (CCC = [0.95, 0.91]) derived values than the same measures from low-resolution images (CCC = [0.72, 0.88]).

Keywords: GAN-CIRCLE, high-resolution reconstruction, deep learning, osteoporosis, trabecular bone, microstructure.

1. INTRODUCTION

Efficacy of *in vivo* computed tomography (CT) imaging modalities in most medical imaging applications depend on their spatial resolution, a determinant of the microstructural quality of the captured Tb network, outlining the scope of biological and clinical questions that can be addressed. Spatial resolution varies not only among different *in vivo* imaging modalities but also among different scanners of same modality, and low spatial resolution along with other modality specific factors such as signal to noise ratio (SNR), exposure time and radiation dose degrade the quality of the reconstructed CT image. Such degradation in CT image quality from high- to low-resolution scanners causes inconsistency in microstructural measurements in multi-site and longitudinal studies, where scanner mismatch across different study locations are evident.

*indranil-guha@uiowa.edu

Medical Imaging 2020: Biomedical Applications in Molecular, Structural, and Functional Imaging, edited by Andrzej Krol, Barjor S. Gimi, Proc. of SPIE Vol. 11317, 113170U · © 2020 SPIE · CCC code: 1605-7422/20/$21 · doi: 10.1117/12.2549318

Although, sophisticated hardware components such as high pitch detector, x-ray tube with fine focal spot etc. are capable of capturing high-resolution (HR) images, they contribute to high CT-machine cost, high radiation dose, and slow scan speed. Hence, HR image reconstruction using computational methods aiming to recover HR image from its low-resolution (LR) counterpart is an emerging research area. Several computational algorithms for HR image reconstruction are available in the literature which can be divided into two broad categories: (1) model-based reconstruction techniques, and (2) learning-based methods.

Model-based algorithms regulate the reconstruction process based on some prior assumption of the degradation scheme. [1-3] On the other hand, learning-based methods accomplish the task by learning the non-linear mapping from low- to high-resolution images.[4-7] Recent progress in deep learning (DL) has established the capability of convolutional neural networks (CNN) in extracting HR features from LR images at low computational cost. Numerous DL based methods have shown promising results in HR reconstruction for different *in vivo* CT modalities.[8-14] Chen *et al.*[8] proposed a deep densely connected network for super resolution (SR) reconstruction of magnetic resonance images. Chaudhury *et al.*[9] proposed a three-dimensional (3D) CNN entitled "DeepResolve" that learns residual-based transformations between paired high- and low-resolution magnetic resolution image slices. Park *et al.*[10] developed a CNN based SR CT image reconstruction technique that learns end-to-end mapping between paired low- and high- resolution image slices using the modified U-Net. Yu *et al.*[11] combined a single-slice CT SR network (s-CTSRN) with skip connections and a multi-slice CT SR network (m-CTSRN) to reconstruct SR CT images. Here, s-CTSRN improves the high frequency details extraction while m-CTSRN maintains the coherence between neighboring CT slices. Several generative adversarial network (GAN) based SR reconstruction techniques have also shown promising results.[15-21] Ledig *et al.* proposed a perceptual loss function as the weighted sum of adversarial and content loss, and used that as an objective for SR image reconstruction using GAN (SRGAN).[15] Wang *et al.* used Residual-in-Residual Dense Block (RRDB) without batch normalization as the building block of SRGAN to eliminate noisy artifacts in reconstructed SR images using SRGAN.[16] Wolterink *et al*. successfully enhanced the image quality by training an unsupervised GAN to learn the non-linear mapping from LR to HR images.[17] You *et al.* trained a CT SR GAN entitled "GAN-CIRCLE" constrained by the identical, residual, and cycle consistency loss using paired two-dimensional (2D) HR and LR patches, where LR CT images were generated by adding noise to the down sampled HR CT images.[18]

In this paper, we investigate high resolution reconstruction techniques in the context of musculoskeletal imaging especially related to osteoporosis.[22,23] Osteoporosis is a common age-related disease characterized by reduced bone mineral density (BMD) and enhanced fracture-risk. Approximately, 40% of women and 13% of men suffer osteoporotic fractures in their lifetime. Osteoporotic hip fractures reduce life expectancy by 10-20%,[24] and increased life expectancy will increase fracture incidence to 6.3 million by 2050.[25] Dual-energy X-ray absorptiometry (DXA)-measured areal BMD is the clinical standard for diagnosis of osteoporosis. However, it is generally agreed that about 60% of bone's mechanical competence is explained by variation in BMD,[26] and there are compelling evidences from histologic studies demonstrating the roles of bone microstructural degeneration in determining bone strength and fracture-risk.[27-29] Thus, valid and reliable measures of Tb microstructural features, associated with bone strength and fracture-risk, are of clinical significance.

Various topologic and geometric methods are available in literature to measure Tb micro-structure.[30-33] Vesterby *et al.* [30] conceived a stereologic parameter, called star volume, which is the average volume of an object region that can be seen from a point inside that region unobscured in all directions. Hahn *et al.* [31] introduced the "trabecular bone pattern factor" which captures Tb connectivity in terms of the convexity property of the Tb surface defined as the ratio of the differences in perimeter and area under dilation. Hildebrand *et al.* [32] developed a 3-D structure model index, a function of global plate-to-rod ratio, based on the observation that the rate of volume change with respect to half thickness (or the radius) for plate-like elements is different from that for rod-like elements. Feldkamp *et al.* [33] showed that the makeup of TB networks can be expressed in terms of topological entities such as the 3-D Euler number. Saha and his colleagues have pioneered unique algorithms,[34-41] characterizing individual trabecular plates and rods, which have been adopted in a large number of studies.[34-57]

Several 3D Tb imaging modalities, including magnetic resonance imaging (MRI)[26,42,58,59] and high-resolution peripheral quantitative computed tomography (HR-pQCT),[60-62] have been popularly applied in bone studies. Emerging multi-detector-row CT (MDCT) scanners enable segmentation and measurement of Tb microstructures and overcome the major deficits of MRI and HR-pQCT modalities related to slow scan speed, limited field-of-view and failure to provide quantitative BMD. It has been demonstrated that microstructural measures are dependent on image resolution and other features.[63] A pertinent challenge with MDCT-based bone imaging emerges due to wide discrepancies in spatial resolution, and other imaging and reconstruction features from different vendors, and rapid upgrades in technology. It introduces a

challenge related to data harmonization in large multi-site or longitudinal studies that typically involves data collection using different scanners. Often, research teams encounter a situation in longitudinal studies that a new and upgraded scanner substitutes the previous scanner in the middle of the study, which can result in an incomplete process for data acquisition and analysis, or a waste of previous data collected using the older machine. This results in a need for data harmonization, which will add immense value to the bone research community, enabling them to use data collected from multiple scanners, facilitating study design and improving fidelity of longitudinal studies that often face unavoidable situations of data collection from different scanners. In this paper, we present a deep learning-based optimized method for HR reconstruction of Tb microstructures from LR CT scans using GAN-CIRCLE,[18,19] and evaluate its performance in terms of image quality as well as clinically significant microstructural measures. Specifically, we compute different Tb microstructural measures such as thickness, spacing, and network area density from the LR and predicted HR CT scans, and examine their agreement with the reference values derived from true HR CT scans. DL network architecture, data acquisition protocols, post-processing steps, computation of microstructural measures, and statistical methods used for evaluation are elaborated in section 2. Experimental results are described in Section 3, and the conclusions are drawn in Section 4.

2. METHODOLOGY

In this section, we describe the deep learning network architecture developed for HR reconstruction of Tb image from the LR CT scan, followed by a brief description of the CT scan protocols on low- and high-resolution MDCT scanners as well as the steps for extracting matching pairs of LR and HR patches for training, validation, and testing. Finally, Tb microstructural measures examined in this paper are defined, and techniques adopted for statistical evaluation of these measures are discussed.

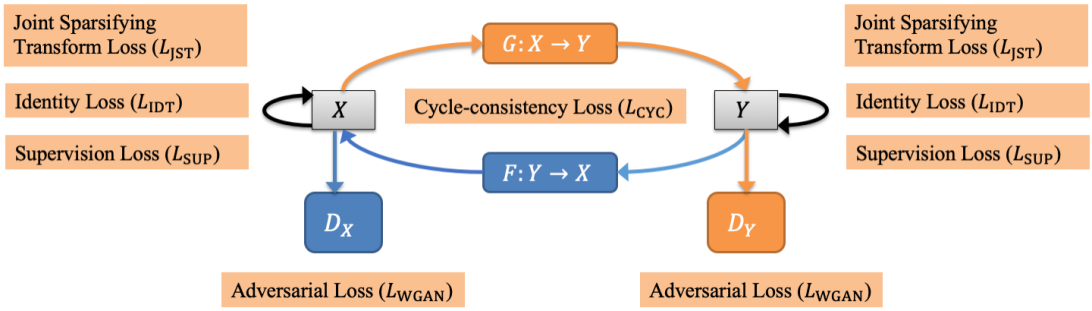

Figure 1. Basic modules of the GAN-CIRCLE used for HR CT image reconstruction. Here, X is the set of low-resolution CT scans, and Y is its high-resolution counterparts. The network consists of two basic GAN modules (Generator: G, Discriminator: D_Y) and (Generator: F, Discriminator: D_X), which are responsible for low-to-high and high-to-low resolution image reconstruction, respectively. Different loss functions are synergistically coupled to train the network, while constrained under the regularization terms related to cycle-consistency and identity loss to avoid over-fitting.

2.1 Deep Learning Network Architecture

A new deep learning network is designed and developed for HR reconstruction of Tb images capturing Tb microstructures from their LR CT scans using the basic principle of GAN-CIRCLE (Figure 1). GAN-CIRCLE is designed with two generator mappings $G: X \rightarrow Y$ and $F: Y \rightarrow X$, where X is the set of LR CT scans and Y is the set of HR counterparts. The overall goal of this network is to learn the nonlinear process of the low-to-high resolution generator ($G: X \rightarrow Y$), while synergistically regularizing with cycle-consistency, adversarial, and several other loss functions. The quality of the generator G is directly controlled using a supervision loss function L_{SUP} and, a discriminator D_Y and the associated adversarial loss function L_{WGAN}. The idea is that, during the learning process, both generator G and discriminator D_Y synergistically improve, simultaneously reducing both supervision and adversarial losses. To regularize the learning processes, a cycle-consistency check was added through a counter generator $F: Y \rightarrow X$ from HR to LR and a cycle-consistent loss function L_{CYC}. Similar to the case of generator G, the learning process of F is controlled using the supervision loss function L_{SUP} and the adversarial loss function L_{WGAN} associated with another discriminator D_X. An identity loss function L_{IDT} is used to regularize the training process which ensures that the network does not generate a significantly different output when a true HR patch is fed as an input to the LR-to-HR generator G. Finally, a joint

sparsifying transformation loss function L_{JST} is used for simultaneously sparsifying predicted images and reducing noise, while preserving anatomical features by minimizing the difference from the true HR image. So, during training the network tries to minimize the following objective function:

$$L_{\mathrm{GAN-CIRCLE}} = L_{\mathrm{WGAN}}(D_Y, G) + L_{\mathrm{WGAN}}(D_X, F) + \lambda_1 L_{\mathrm{CYC}}(G, F) + \lambda_2 L_{\mathrm{IDT}}(G, F) + \lambda_3 L_{\mathrm{JST}}(G) + L_{\mathrm{SUP}}(G, F),$$

where, $\lambda_1, \lambda_2, and\ \lambda_3$ are the weights for different losses.[18]

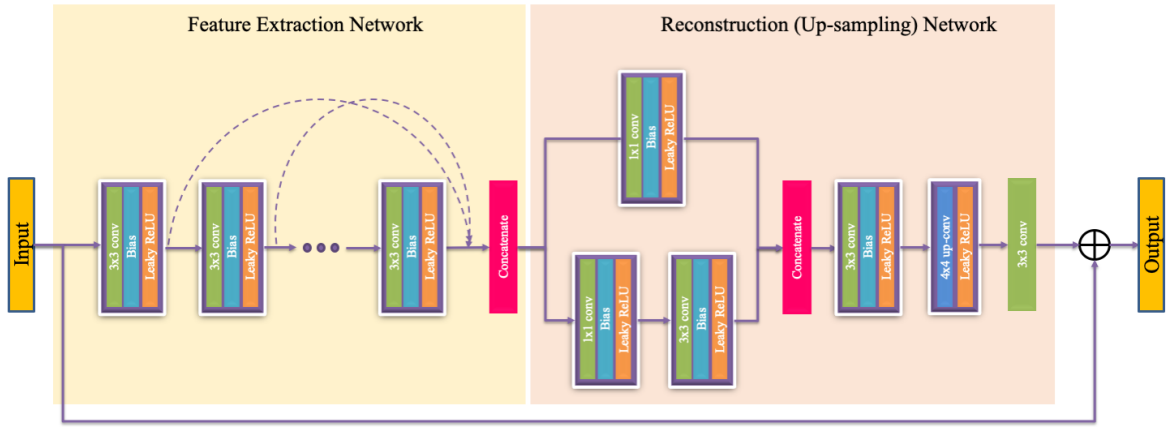

Figure 2. Architecture of a generator. Stride within each convolution layer is 1 except for the first layer where the stride is 2.

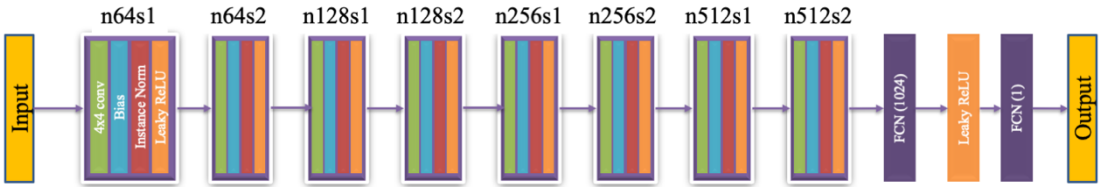

Figure 3. Architecture of a discriminator. Here, n denotes the number of kernels in a convolutional layer, and s denotes the stride.

Network architectures of the generators G and F, and the discriminators D_X and D_Y are shown in Figure 2 and Figure 3 respectively. The Generator network is consisting of two sub-networks, a feature extraction network, and a reconstruction or up-sampling network. In the feature extraction network 12 non-linear SR blocks are concatenated, where each block consists of 3×3 convolution kernels, bias, leaky rectified linear unit (Leaky ReLU), and a dropout layer. Leaky ReLU is defined as follows: $\max(0, x) - \alpha \max(0, -x)$, where α denotes the slope of the function. At each convolutional layer same number of filters as suggested by You *et al.*[18] is used. Stride within each convolution layer is set to 1 except for the first layer where the stride is 2; stride is the number of pixels with which the kernel slides horizontally or vertically. To capture the local as well as global features, outputs of the hidden layers in the feature extraction network are concatenated using a skip connection before feeding into the reconstruction network. The skip connection is used to prevent overfitting and training saturation. Reconstruction network follows a network in network architecture, where two parallel branches of CNNs are concatenated before up-sampling. Two SR blocks with 1×1 convolution kernel is used in the two parallel branches to reduce dimensionality while increasing the non-linearity of the network. After concatenating the outputs of the parallel branches, the image is up-sampled by a factor of 2 and finally, all the feature maps are fused to generate a residual image in the last convolution layer. For supervised training, residual image is combined with the input image to obtain the HR output image. Discriminator network is composed of 8 layers of convolution, bias, instance norm, and Leaky ReLU, followed by two fully-connected (FCN) layers with 1024 units and 1 unit. Each of the 8 convolution layers use a 4×4 convolution kernel with 64, 64, 128, 128, 256, 256, 512, and 512 filters respectively.

2.2 Dataset Description

Matching HR and LR ankle CT scans of human volunteers acquired on two different CT scanners with significantly different spatial resolution were used for training, validation, and testing of the GAN-CIRCLE-based HR reconstructor described in the previous section. Specifically, nineteen healthy volunteers (age: 26.2 ± 4.5 Y; 10 F) were recruited and the distal tibia of their left legs were scanned on two MDCT scanners. The study was conducted around the transition period of the MDCT scanner upgrade at the University of Iowa Comprehensive Lung Imaging Center (I-CLIC). The first MDCT distal scan on each volunteer was performed on a LR Siemens FLASH scanner, and then they were recalled and rescanned on a HR Siemens FORCE scanner after upgrade. The average time gap between the LR and HR scans was 44.6 ± 2.7 days, with the minimum and maximum gaps of 40 and 48 days, respectively. The human study was approved by The University of Iowa Institutional Review Board and all participants provided written informed consent. The CT scan protocols on the two scanners are described here.

FLASH scanner: Single X-ray source spiral acquisition at 120 kV, 200 effective mAs, 1 sec rotation speed, pitch factor: 1.0, total effective dose equivalent: 170 μSv ≈ 20 days of environmental radiation in the U.S. Images were reconstructed at 200 μm slice-spacing using a normal cone beam method with a special U70u kernel.

FORCE scanner: Single X-ray source spiral acquisition at 120 kV, 100 effective mAs, 1 sec rotation speed, pitch factor: 1.0, total effective dose equivalent: 50 μSv ≈ 5 days of environmental radiation in the U.S. Images were reconstructed at 200 μm slice-spacing and 150 μm pixel-size using Siemens's special kernel Ur77u with Edge Technology.

For both scanners, Siemens z-UHR scan mode was applied, enabling Siemens double z sampling technology and achieving high structural resolution.

2.3 Data Processing, Training, and Validation

A Gammex RMI 467 Tissue Characterization Phantom (Gammex RMI, Middleton, WI) was scanned to calibrate CT Hounsfield numbers into BMD. First, both LR and HR CT scans were converted into BMD images using corresponding calibration phantom scans; then LR images were interpolated at 150 μm isotropic voxel size; and finally, HR images were registered to corresponding LR images. HR images were registered to the LR image in two-steps: first, cortical bone and Tb network of the HR image was manually registered to the LR image using ITK-SNAP registration toolkit.[64] In the second step, a rigid transformation, initialized by the transformation matrix from the manual registration step, was applied on the registered HR image for fine tuning. To improve the registration accuracy, region of interest (ROI) for registration cost function was defined as the distal tibia with a soft boundary. For training and testing purposes, pairs of low- and high-resolution matching patches of size 64×64 were randomly harvested from 30% peeled ROIs of registered LR and HR BMD images, and scaled to the unit interval of [0 1].

For training, weights of the convolution layers were initialized using the techniques proposed by He *et al.*[65] In the training process $\lambda_1, \lambda_2,$ and λ_3 were set to 1, 0.5, and 0.001 respectively. Slope α of the leaky ReLU was set to 0.1 and drop out was applied to each convolution layer with $p = 0.8$. The network was trained for 500 epochs and took almost a day to train using Adam optimizer[66] with $\beta_1 = 0.5$ and $\beta_2 = 0.9$ and a learning rate of 1×10^{-4}.

2.4 Computation of Tb Microstructural Measures

Table 1. List of CT-derived trabecular bone measures examined in this paper. Nomenclatures of trabecular bone measures used by Bouxsein *et al.*[67] and Chen *et al.*[63] are followed here wherever possible.

Parameter (unit)	Description
Tb.NA (mm^2/mm^3)	Tb network area density, i.e., the average area of the medial surface of segmented bone per unit ROI
Tb.Th (μm)	Mean trabecular thickness computed by star-line analysis[68]
Tb.Sp (μm)	Mean trabecular spacing, i.e., the space between trabecular microstructures computed by star-line analysis[68]

Tb measures examined in this paper are listed in Table 1. Each BMD image was processed through the following image-processing steps to compute different trabecular bone measures – (1) fuzzy skeletonization[69] and computation of trabecular network area density (Tb.NA) measure; (2) star-line analysis for computation of trabecular thickness (Tb.Th) and

trabecular spacing (Tb.Sp) measures.[68] Note that all Tb measures were computed on 2D slices of LR, true HR and predicted HR images.

2.5 Statistical Analysis

For quantitative evaluation, 2D measurements of Tb thickness, spacing, and network area were stacked to reconstruct a 3D representation of the measures. Summary statistics of each Tb measure from LR, true and predicted HR scans was computed in terms of mean and standard deviation over 100 randomly selected spherical ROIs of 12-pixel radius, fully confined within the 30% peeled ROI of each scan. Finally, linear and concordance correlation of the mean values of individual Tb measure from LR and predicted HR images with the reference mean values derived from true HR CT data was examined for each specimen.

3. EXPERIMENTS AND RESULTS

GAN-CIRCLE based HR Tb image reconstructor was trained and validated using matching low- and high-resolution CT scans of ten human volunteers, and the network was qualitatively as well as quantitatively evaluated on matching low- and high-resolution CT scans of nine other volunteers. Finally, improvement in both image quality and clinically significant Tb microstructural measures from LR to predicted HR scans were assessed w.r.t. the reference measures from true HR image.

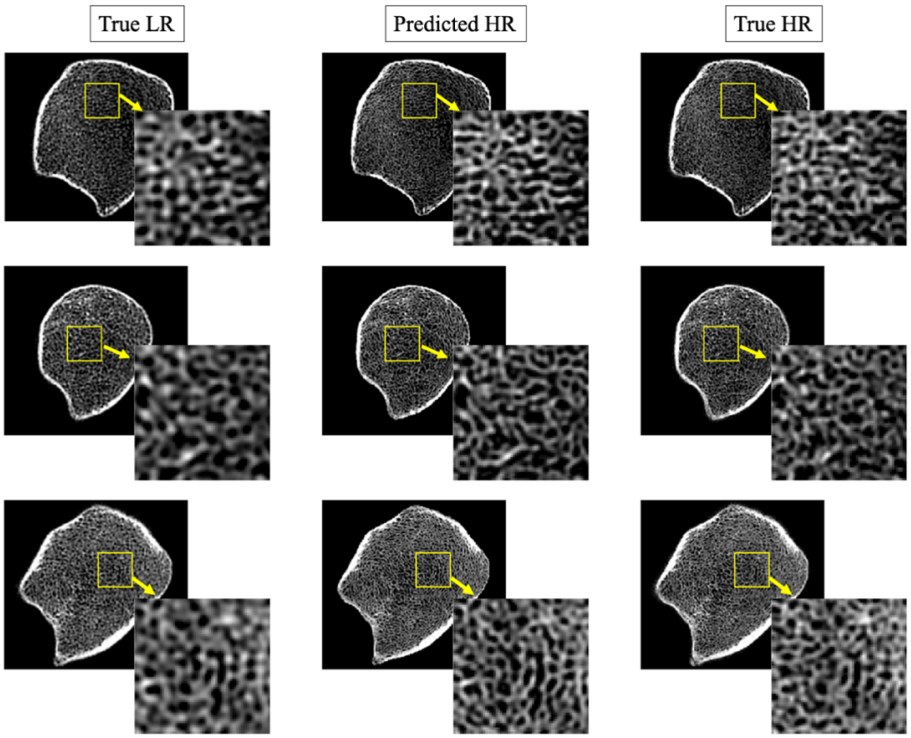

Figure 4. Results of HR prediction from LR trabecular bone CT scans using GAN-CIRCLE. Left-to-right columns: matching 2D image slices from true LR, predicted HR, and true HR CT data. LR and HR CT images of distal tibia were acquired on a Siemens FLASH and FORCE scanners, respectively. Three different rows represent images from three different human subjects.

Specifically, nine thousand matching pairs of low- and high-resolution patches from registered BMD images of ten volunteers were randomly harvested for training and validation of the HR Tb microstructure reconstructor. The set of 9,000 training samples was randomly split into 4:1 ratio for training and validation purposes. A different set of 5,000 pairs of matching LR and HR patches from the registered BMD images of nine other volunteers were used for testing and evaluation.

Results of deep learning-based HR reconstruction are illustrated in Figure 4. Figure 4 presents predicted HR images (middle row) from three LR images (left row); matching true HR images are shown in the right column for visual comparison. HR reconstruction of a full image slice was obtained by independently reconstructing non-overlapping 64×64 patches and stitching them together. No block effects are visible in the predicted HR images despite independent patch reconstruction. This observation suggests that cycle-consistency and other regularization constraints of the GAN-CIRCLE can avoid random patch-bias artifacts. To display the performance of the method at the level of Tb microstructures, one random patch is zoomed in for each example. It is observed by visually comparing matching zoomed in patches that the GAN-CIRCLE successfully performs a non-linear deblurring and filtering to reconstruct HR Tb microstructural features and resolution from blurred and noisy LR data.

Table 2. Results of linear and concordance correlation analyses for different Tb measures derived from LR, HR, and predicted HR images for individual subjects. Each cell shows mean ± SD of observed values of corresponding statistical metrics for different subjects.

	LR vs. True HR		**Predicted HR vs. True HR**	
Tb Measures	**Linear Correlation (r)**	**CCC**	**Linear Correlation (r)**	**CCC**
Tb.Th (μm)	0.95 ± 0.02	0.66 ± 0.12	0.96 ± 0.01	0.95 ± 0.03
Tb.Sp (μm)	0.96 ± 0.02	0.93 ± 0.05	0.97 ± 0.02	0.95 ± 0.04
Tb.NA (mm^2/mm^3)	0.88 ± 0.07	0.83 ± 0.11	0.89 ± 0.07	0.88 ± 0.08

The performance of the network was quantitatively evaluated using structural similarity (SSIM) index[70]— a widely used method to estimate perceived quality of images and videos. For quantitative analysis, only 4,000 patches entirely lying inside the Tb were used, since the objective of this work was to predict HR microstructures from LR images. SSIM was computed between LR and true HR, as well as between predicted HR and true HR patches, and a paired t-test was conducted between these two sets of SSIM values. The results of the paired t-test suggest that the reconstruction network significantly improves ($p < 0.01$) the structural similarity with the true HR Tb microstructures.

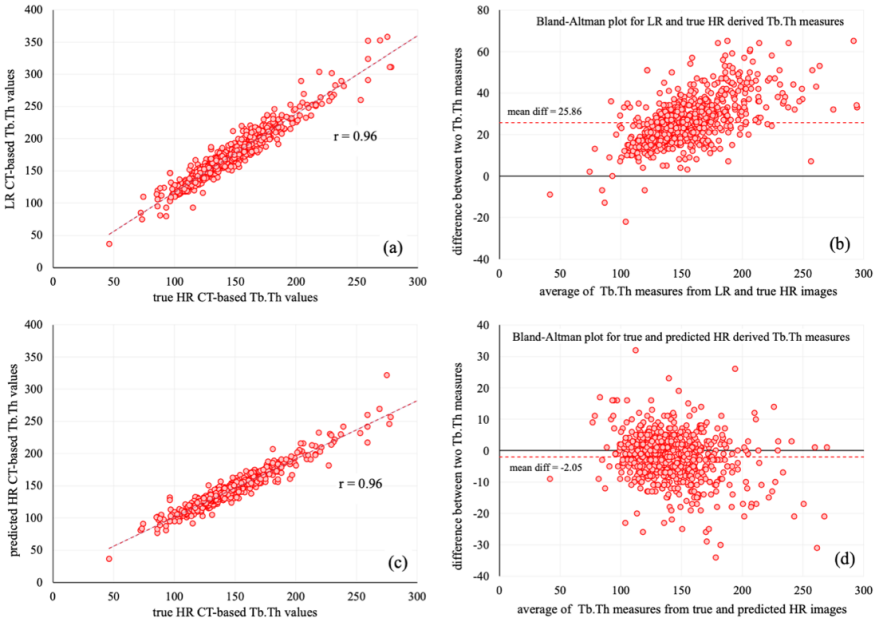

Figure 5. Linear correlation and Bland-Altman plots of Tb.Th measures from LR, HR, and predicted HR images of nine test subjects. (a) Linear correlation analysis among Tb.Th values from LR and true HR images. (b) Bland-Altman plot of measure difference between LR and true HR images. (c,d) Same as (a,b) but for true HR and predicted HR images.

For each test subject, agreement in the microstructural measurements from LR and predicted HR images with the reference measurements from true HR data was examined in terms of linear (r) and concordance correlation coefficients (CCC).

Mean and standard deviation (SD) of the r-values and CCCs for Tb.Th, Tb.Sp, and Tb.NA measures from nine test subjects are presented in Table 2. A paired t-test was conducted between the r-values from LR and true HR images, as well as predicted and true HR images for all three Tb measures, and the results showed no significant difference ($p > 0.05$) between the two sets of values. However, paired t-test on the CCC for each Tb measure showed that the measurements from predicted HR images had higher CCC with the reference values from true HR images compared to the values derived from the corresponding LR images. Most significant improvement in CCC was observed for Tb.Th ($p < 0.0001$) measure, where mean CCC between LR and true HR images was 0.66, while mean CCC between predicted and true HR images was 0.95. However, relatively smaller improvement in CCC was noticed for Tb.Sp ($p < 0.01$) and Tb.NA ($p < 0.01$) measure as compared to Tb.Th.

Table 3. Results of linear and concordance correlation analysis of different Tb measures from LR, HR, and predicted HR images of nine test subjects combined together.

Tb Measures	**LR – True HR**		**Predicted HR – True HR**	
	Linear Correlation (r)	**CCC**	**Linear Correlation (r)**	**CCC**
Tb.Th (μm)	0.96	0.72	0.96	0.95
Tb.Sp (μm)	0.98	0.97	0.98	0.98
Tb.NA (mm^2/mm^3)	0.91	0.88	0.92	0.91

Measurements from nine test subjects were combined, and r-values, CCC among the microstructural measurements from LR and true HR images, as well as predicted HR and true HR images are reported in 3. For the combined analysis, all three Tb measures computed from LR and predicted HR images produced similar r-values with the reference measures from true HR scans. After combining Tb.Th values from all the test subjects, CCC between Tb.Th values from LR and true HR images was found to be 0.72 while for predicted and true HR CCC was 0.95. Figure 5 shows the linear correlation (a,b) and Bland-Altman (c,d) plot of combined Tb.Th measurements from nine test subjects. As can be seen from (c) and (d), mean difference between Tb.Th measurements from true and predicted HR images are significantly lower than the mean difference in Tb.Th values from LR and true HR images. This observation suggests that the proposed HR reconstructor successfully reduces the bias in Tb.Th measures from LR images. When combined, Tb.Sp and Tb.NA measures from LR images produced CCC of 0.97 and 0.88, respectively, with the values derived from true HR images, whereas the same measures from predicted HR images had CCC of 0.98 and 0.91 with the reference values. These results support our claim that the HR reconstructor successfully recovers the HR microstructural features from a LR CT image through non-linear deblurring and filtering.

4. CONCLUSIONS

In this paper, we have developed and evaluated a deep learning-based method for HR reconstruction of Tb microstructures from LR CT scans using GAN-CIRCLE. To the best of our knowledge, for the first time in biomedical imaging research, true matching LR and HR scans, collected from two different scanners, have been used for training and development of a HR reconstructor. Results of quantitative evaluative experiments on human ankle CT scans from two different scanners have been presented. The performance of the HR reconstructor has been evaluated both in terms of image quality, as well as clinically significant microstructural measures. Results of paired t-test have shown that SSIM values between predicted HR and true HR patches are significantly higher than the SSIM values between the LR and true HR patches. In general, Tb.Th, Tb.Sp, and Tb.NA measures derived from predicted HR images have shown higher agreement with the reference values computed from true HR scans as compared to the same measurements obtained from LR images. Furthermore, Bland-Altman plot of Tb.Th measures have shown that the deep learning network successfully reduces the fixed bias in Tb.Th values from LR images. These observations suggest that the new HR reconstructor will facilitate data harmonization to overcome challenges related to discrepancies in imaging facilities and features at different research performance sites and improve the fidelity of both cross-sectional and longitudinal study data, often collected using different scanners.

ACKNOWLEDGEMENTS

This work was supported in part by the NIH grant R01 HL142042.

REFERENCES

[1] R Zhang, J-B Thibault, CA Bouman, KD Sauer, and J Hsieh, "Model-based iterative reconstruction for dual-energy X-ray CT using a joint quadratic likelihood model," *IEEE Transactions on Medical Imaging*, **33**, 117-134, 2013.

[2] JB Thibault, KD Sauer, CA Bouman, and J Hsieh, "A three-dimensional statistical approach to improved image quality for multislice helical CT," *Medical Physics*, **34**, 4526-4544, 2007.

[3] CA Bouman and K Sauer, "A unified approach to statistical tomography using coordinate descent optimization," *IEEE Transactions on Image Processing*, **5**, 480-492, 1996.

[4] W Dong, L Zhang, G Shi, and X Wu, "Image deblurring and super-resolution by adaptive sparse domain selection and adaptive regularization," *IEEE Transactions on Image Processing*, **20**, 1838-1857, 2011.

[5] C Jiang, Q Zhang, R Fan, and Z Hu, "Super-resolution ct image reconstruction based on dictionary learning and sparse representation," *Scientific Reports*, **8**, 87-99, 2018.

[6] Z Wang, Y Yang, Z Wang, S Chang, J Yang, and TS Huang, "Learning super-resolution jointly from external and internal examples," *IEEE Transactions on Image Processing*, **24**, 4359-4371, 2015.

[7] J Yang, J Wright, TS Huang, and Y Ma, "Image super-resolution via sparse representation," *IEEE Transactions on Image Processing*, **19**, 2861-2873, 2010.

[8] Y Chen, F Shi, AG Christodoulou, Y Xie, Z Zhou, and D Li, "Efficient and accurate MRI super-resolution using a generative adversarial network and 3D multi-level densely connected network," Proc. of *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 91-99, 2018.

[9] AS Chaudhari, Z Fang, F Kogan, J Wood, KJ Stevens, EK Gibbons, JH Lee, GE Gold, and BA Hargreaves, "Super-resolution musculoskeletal MRI using deep learning," *Magnetic Resonance in Medicine*, **80**, 2139-2154, 2018.

[10] J Park, D Hwang, KY Kim, SK Kang, YK Kim, and JS Lee, "Computed tomography super-resolution using deep convolutional neural network," *Physics in Medicine & Biology*, **63**, 145011, 2018.

[11] H Yu, D Liu, H Shi, H Yu, Z Wang, X Wang, B Cross, M Bramler, and TS Huang, "Computed tomography super-resolution using convolutional neural networks," Proc. of *2017 IEEE International Conference on Image Processing (ICIP)*, 3944-3948, 2017.

[12] C You, Q Yang, L Gjesteby, G Li, S Ju, Z Zhang, Z Zhao, Y Zhang, W Cong, and G Wang, "Structurally-sensitive multi-scale deep neural network for low-dose CT denoising," *IEEE Access*, **6**, 41839-41855, 2018.

[13] C You, L Yang, Y Zhang, and G Wang, *Low-dose CT via deep CNN with skip connection and network-in-network*, SPIE, 2019.

[14] L Yang, RP Ghosh, JM Franklin, C You, and JT Liphardt, "NuSeT: A Deep Learning Tool for Reliably Separating and Analyzing Crowded Cells," *bioRxiv*, 2019.

[15] C Ledig, L Theis, F Huszár, J Caballero, A Cunningham, A Acosta, A Aitken, A Tejani, J Totz, and Z Wang, "Photo-realistic single image super-resolution using a generative adversarial network," Proc. of *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 4681-4690, 2017.

[16] X Wang, K Yu, S Wu, J Gu, Y Liu, C Dong, Y Qiao, and C Change Loy, "Esrgan: Enhanced super-resolution generative adversarial networks," Proc. of *Proceedings of the European Conference on Computer Vision (ECCV)*, 1-16, 2018.

[17] JM Wolterink, T Leiner, MA Viergever, and I Išgum, "Generative adversarial networks for noise reduction in low-dose CT," *IEEE Transactions on Medical Imaging*, **36**, 2536-2545, 2017.

[18] C You, G Li, Y Zhang, X Zhang, H Shan, M Li, S Ju, Z Zhao, Z Zhang, and W Cong, "CT super-resolution GAN constrained by the identical, residual, and cycle learning ensemble (GAN-CIRCLE)," *IEEE Transactions on Medical Imaging*, **39**, 188-203, 2020.

[19] I Goodfellow, J Pouget-Abadie, M Mirza, B Xu, D Warde-Farley, S Ozair, A Courville, and Y Bengio, "Generative adversarial nets," Proc. of *Advances in Neural Information Processing Systems*, 2672-2680, 2014.

[20] Q Lyu, C You, H Shan, and G Wang, "Super-resolution MRI through Deep Learning," *arXiv preprint arXiv:1810.06776*, 2018.

[21] Q Lyu, C You, H Shan, Y Zhang, and G Wang, *Super-resolution MRI and CT through GAN-CIRCLE*, SPIE, 2019.

[22] BL Riggs and LJ Melton, "Bone turnover matters: The raloxifene treatment paradox of dramatic decreases in vertebral fractures without commensurate increases in bone density," *Journal of Bone and Mineral Research*, **17**, 11-14, 2002.

[23] S Khosla, BL Riggs, EJ Atkinson, AL Oberg, LJ McDaniel, M Holets, JM Peterson, and LJ Melton, 3rd, "Effects of sex and age on bone microstructure at the ultradistal radius: a population-based noninvasive in vivo assessment," *Journal of Bone and Mineral Research*, **21**, 124-131, 2006, PubMed Central PMCID: PMC1352156.

[24] LJ Melton, "Epidemiology of fractures," in *Osteoporosis: Etiology, Diagnosis, and Management*, B. L. Riggs and L. J. Melton, Eds., New York, Raven Press, 1988, pp. 133-154.

[25] C Cooper, G Campion, and Lr Melton, "Hip fractures in the elderly: a world-wide projection," *Osteoporosis International*, **2**, 285-289, 1992.

[26] FW Wehrli, PK Saha, BR Gomberg, HK Song, PJ Snyder, M Benito, A Wright, and R Weening, "Role of magnetic resonance for assessing structure and function of trabecular bone," *Topics in Magnetic Resonance Imaging*, **13**, 335-355, 2002.

[27] M Kleerekoper, AR Villanueva, J Stanciu, DS Rao, and AM Parfitt, "The role of three-dimensional trabecular microstructure in the pathogenesis of vertebral compression fractures," *Calcified Tissue International*, **37**, 594-597, 1985.
[28] E Legrand, D Chappard, C Pascaretti, M Duquenne, S Krebs, V Rohmer, MF Basle, and M Audran, "Trabecular bone microarchitecture, bone mineral density, and vertebral fractures in male osteoporosis," *Journal of Bone and Mineral Research*, **15**, 13-19, 2000.
[29] AM Parfitt, CHE Mathews, AR Villanueva, M Kleerekoper, B Frame, and DS Rao, "Relationships between surface, volume, and thickness of iliac trabecular bone in aging and in osteoporosis - implications for the microanatomic and cellular mechanisms of bone loss," *Journal of Clinical Investigation*, **72**, 1396-1409, 1983.
[30] A Vesterby, HJ Gundersen, and F Melsen, "Star volume of marrow space and trabeculae of the first lumbar vertebra: sampling efficiency and biological variation," *Bone*, **10**, 7-13, 1989.
[31] M Hahn, M Vogel, M Pompesius-Kempa, and G Delling, "Trabecular bone pattern factor--a new parameter for simple quantification of bone microarchitecture," *Bone*, **13**, 327-30, 1992.
[32] T Hildebrand and P Rüegsegger, "Quantification of bone microarchitecture with the structure model index," *Computer Methods in Biomechanics and Biomedical Engineering*, **1**, 15-23, 1997.
[33] LA Feldkamp, SA Goldstein, AM Parfitt, G Jesion, and M Kleerekoper, "The direct examination of three-dimensional bone architecture in vitro by computed tomography," *Journal of Bone and Mineral Research*, **4**, 3-11, 1989.
[34] PK Saha and BB Chaudhuri, "Detection of 3-D simple points for topology preserving transformations with application to thinning," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, **16**, 1028-1032, 1994.
[35] PK Saha and BB Chaudhuri, "3D digital topology under binary transformation with applications," *Computer Vision and Image Understanding*, **63**, 418-429, 1996.
[36] PK Saha, BB Chaudhuri, and DD Majumder, "A new shape preserving parallel thinning algorithm for 3D digital images," *Pattern Recognition*, **30**, 1939-1955, 1997.
[37] PK Saha, BR Gomberg, and FW Wehrli, "Three-dimensional digital topological characterization of cancellous bone architecture," *International Journal of Imaging Systems and Technology*, **11**, 81-90, 2000.
[38] BR Gomberg, PK Saha, HK Song, SN Hwang, and FW Wehrli, "Topological analysis of trabecular bone MR images," *IEEE Transactions on Medical Imaging*, **19**, 166-174, 2000.
[39] XS Liu, P Sajda, PK Saha, FW Wehrli, G Bevill, TM Keaveny, and XE Guo, "Complete volumetric decomposition of individual trabecular plates and rods and its morphological correlations with anisotropic elastic moduli in human trabecular bone," *Journal of Bone and Mineral Research*, **23**, 223-235, 2008.
[40] PK Saha, Y Xu, H Duan, A Heiner, and G Liang, "Volumetric topological analysis: a novel approach for trabecular bone classification on the continuum between plates and rods," *IEEE Transaction on Medical Imaging*, **29**, 1821-1838, 2010.
[41] PK Saha, Y Liu, C Chen, D Jin, EM Letuchy, Z Xu, RE Amelon, TL Burns, JC Torner, and SM Levy, "Characterization of trabecular bone plate-rod microarchitecture using multirow detector CT and the tensor scale: Algorithms, validation, and applications to pilot human studies," *Medical Physics*, **42**, 5410-5425, 2015.
[42] G Chang, SK Pakin, ME Schweitzer, PK Saha, and RR Regatte, "Adaptations in trabecular bone microarchitecture in Olympic athletes determined by 7T MRI," *Journal of Magnetic Resonance in Medicine*, **27**, 1089-1095, 2008, PubMed Central PMCID: PMCPMC3850284.
[43] M Stauber and R Muller, "Volumetric spatial decomposition of trabecular bone into rods and plates--a new method for local bone morphometry," *Bone*, **38**, 475-84, 2006.
[44] G Chang, L Wang, G Liang, JS Babb, PK Saha, and RR Regatte, "Reproducibility of subregional trabecular bone micro-architectural measures derived from 7-Tesla magnetic resonance images," *MAGMA*, **24**, 121-5, 2011, PubMed Central PMCID: PMCPMC3710719.
[45] G Chang, L Wang, G Liang, JS Babb, GC Wiggins, PK Saha, and RR Regatte, "Quantitative assessment of trabecular bone micro-architecture of the wrist via 7 Tesla MRI: preliminary results," *MAGMA*, **24**, 191-9, 2011, PubMed Central PMCID: PMCPMC3723135.
[46] S Dudley-Javoroski, PK Saha, G Liang, C Li, Z Gao, and RK Shields, "High dose compressive loads attenuate bone mineral loss in humans with spinal cord injury," *Osteoporosis International*, **23**, 2335-2346, 2012, PubMed Central PMCID: PMC3374128.
[47] S Dudley-Javoroski, R Amelon, YX Liu, PK Saha, and RK Shields, "High bone density masks architectural deficiencies in an individual with spinal cord injury," *Journal of Spinal Cord Medicine*, **37**, 349-354, 2014, PubMed Central PMCID: PMC4064585.
[48] GA Ladinsky, B Vasilic, AM Popescu, M Wald, BS Zemel, PJ Snyder, L Loh, HK Song, PK Saha, AC Wright, and FW Wehrli, "Trabecular structure quantified with the MRI-based virtual bone biopsy in postmenopausal women contributes to vertebral deformity burden independent of areal vertebral BMD," *J Bone Miner Res*, **23**, 64-74, 2008, PubMed Central PMCID: PMC2663589.
[49] M Benito, B Gomberg, FW Wehrli, RH Weening, B Zemel, AC Wright, HK Song, A Cucchiara, and PJ Snyder, "Deterioration of trabecular architecture in hypogonadal men," *The Journal of Clinical Endocrinology and Metabolism*, **88**, 1497-502, 2003.
[50] PK Saha, Y Liu, TL Burn, JC Torner, and SM Levy, "Effects of physical activity on trabecular bone micro-architecture: a comparative study in young men and women using multi-detector CT and volumetric topological analysis," Proc. of *IEEE*

International Conference on Intelligent Computation and Bio-Medical Instrumentation, 283-286, Wuhan, China, December 14-17 2011.

[51] PK Saha, Y Liu, P C. A, TL Burns, JC Torner, and SM Levy, "Quantitative bone micro-architecture in young adults using multi-detector CT imaging and volumetric topological analysis – a feasibility study," Proc. of *Annual Meeting of the American Society for Bone and Mineral Research*, San Diego, CA, 2011.

[52] PK Saha, CA Calarge, C Li, Y Liu, JM Fishbaugher, BC Tyler, NM Baker, TL Burns, KF Janz, JC Torner, and SM Levy, "Trabecular bone micro-architecture during SSRI treatment using multi-detector CT imaging and topological analysis on a continuum between plates and rods," Proc. of *Annual Meeting of the American Society for Bone and Mineral Research*, Minneapolis, MN, 2012.

[53] PK Saha, RE Amelon, Y Liu, C Li, D Jin, C Chen, JM Fishbaugher, EM Letuchy, CA Calarge, KF Janz, DB Hornick, J Eichenberger-Gilmore, TL Burns, JC Torner, and SM Levy, "In vivo study of trabecular and cortical bone in young adults with varying trajectories of bone development using multi-row detector CT imaging," Proc. of *Annual Meeting of the American Society for Bone and Mineral Research*, Baltimore, MD, 2013.

[54] FW Wehrli, GA Ladinsky, C Jones, M Benito, J Magland, B Vasilic, AM Popescu, B Zemel, AJ Cucchiara, AC Wright, HK Song, PK Saha, H Peachey, and PJ Snyder, "In vivo magnetic resonance detects rapid remodeling changes in the topology of the trabecular bone network after menopause and the protective effect of estradiol," *Journal of Bone and Mineral Research*, **23**, 730-40, 2008, PubMed Central PMCID: PMC2674544.

[55] XS Liu, AH Huang, XH Zhang, P Sajda, B Ji, and XE Guo, "Dynamic simulation of three dimensional architectural and mechanical alterations in human trabecular bone during menopause," *Bone*, **43**, 292-301, 2008, PubMed Central PMCID: PMC2526101.

[56] XS Liu, A Cohen, E Shane, E Stein, H Rogers, SL Kokolus, PT Yin, DJ McMahon, JM Lappe, RR Recker, and XE Guo, "Individual trabeculae segmentation (ITS)-based morphological analysis of high-resolution peripheral quantitative computed tomography images detects abnormal trabecular plate and rod microarchitecture in premenopausal women with idiopathic osteoporosis," *Journal of bone and mineral research* **25**, 1496-505, 2010, PubMed Central PMCID: PMC3131618.

[57] XS Liu, EM Stein, B Zhou, CA Zhang, TL Nickolas, A Cohen, V Thomas, DJ McMahon, F Cosman, J Nieves, E Shane, and XE Guo, "Individual trabecula segmentation (ITS)-based morphological analyses and microfinite element analysis of HR-pQCT images discriminate postmenopausal fragility fractures independent of DXA measurements," *J Bone Miner Res*, **27**, 263-72, 2012, PubMed Central PMCID: PMC3290758.

[58] S Majumdar, D Newitt, A Mathur, D Osman, A Gies, E Chiu, J Lotz, J Kinney, and H Genant, "Magnetic resonance imaging of trabecular bone structure in the distal radius: relationship with X-ray tomographic microscopy and biomechanics," *Osteoporosis International*, **6**, 376-385, 1996.

[59] TM Link, S Majumdar, P Augat, JC Lin, D Newitt, Y Lu, NE Lane, and HK Genant, "In vivo high resolution MRI of the calcaneus: differences in trabecular structure in osteoporosis patients," *Journal of Bone and Mineral Research*, **13**, 1175-1182, 1998.

[60] S Boutroy, ML Bouxsein, F Munoz, and PD Delmas, "In vivo assessment of trabecular bone microarchitecture by high-resolution peripheral quantitative computed tomography," *The Journal of Clinical Endocrinology and Metabolism*, **90**, 6508-6515, 2005.

[61] M Burrows, D Liu, and H McKay, "High-resolution peripheral QCT imaging of bone micro-structure in adolescents," *Osteoporosis International*, **21**, 515-520, 2010.

[62] R Krug, AJ Burghardt, S Majumdar, and TM Link, "High-resolution imaging techniques for the assessment of osteoporosis," *Radiologic Clinics of North America*, **48**, 601-621, 2010, PubMed Central PMCID: PMC2901255.

[63] C Chen, X Zhang, J Guo, D Jin, EM Letuchy, TL Burns, SM Levy, EA Hoffman, and PK Saha, "Quantitative imaging of peripheral trabecular bone microarchitecture using MDCT," *Medical Physics*, **45**, 236-249, 2018.

[64] PA Yushkevich, J Piven, HC Hazlett, RG Smith, S Ho, JC Gee, and G Gerig, "User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability," *Neuroimage*, **31**, 1116-1128, 2006.

[65] K He, X Zhang, S Ren, and J Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," Proc. of *Proceedings of the IEEE International Conference on Computer Vision*, 1026-1034, 2015.

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

[67] ML Bouxsein, SK Boyd, BA Christiansen, RE Guldberg, KJ Jepsen, and R Muller, "Guidelines for assessment of bone microstructure in rodents using micro-computed tomography," *Journal of Bone and Mineral Research*, **25**, 1468-1486, 2010.

[68] Y Liu, D Jin, C Li, KF Janz, TL Burns, JC Torner, SM Levy, and PK Saha, "A robust algorithm for thickness computation at low resolution and its application to in vivo trabecular bone CT imaging," *IEEE Transactions on Biomedical Engineering*, **61**, 2057-2069, 2014.

[69] KS P, D Jin, Y Liu, EC G, and C Chen, "Fuzzy Object Skeletonization: Theory, Algorithms, and Applications," *IEEE Transactions on Visualization and Computer Graphics*, **24**, 2298-2314, 2018.

[70] Z Wang, AC Bovik, HR Sheikh, and EP Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, **13**, 600-612, 2004.