Deep Learning Models Improve Spatial Resolution of MicroCT Images Across Murine Femur and Spine

Michael David[1], Nicole Szeto[1], Jackson Albright[1], Tillman James[2], Dana Godfrey[1], Cheryl Ackert-Bicknell[1], Douglas Adams[1]
[1]Department of Orthopedics, Colorado Program for Musculoskeletal Research, University of Colorado Anschutz Medical Campus, Aurora, CO
[2]Electrical & Systems Engineering, Washington University in St. Louis, St. Louis, MO
michael.david@cuanschutz.edu

DISCLOSURES: MD (N); NS (N); JA (N); TJ (N); DG (N); CAB (N); DA (N)

INTRODUCTION: Clinically, X-ray Computed tomography (CT) enables 3D imaging of skeletal regions for diagnosis, surgical planning, and research, yet is limited in resolution due to shorter imaging times required to limit motion artifacts, cost, and radiation dose. Recently, our group and others have explored deep learning (DL) as a post-acquisition computational approach[1] to improve resolution in both clinical[2] and preclinical[3,4] imaging by generating synthetically higher resolution images from relatively lower resolution acquisitions. However, limited studies and inconclusive results to date warrant further exploration to find the computational bounds and limitations for translating DL models across skeletal sites. Thus, this study aimed to develop and test DL models built upon a vast dataset of murine femur and spine anatomies to produce a generalizable DL model for increasing microCT image resolution.

METHODS: *Bones and microCT imaging:* Femurs (n = 71) and lumbar spines (n = 68; L3-L5 vertebrae) were harvested from separate 16-week-old Diversity Outbred female and male mice treated with or without parathyroid hormone as part of our other studies (IACUC-approved; n = 139 total with balanced groups). Bones were imaged via microCT at 70 kVp and 200 µA to obtain relatively lower (LR; 1000 projections; 20 µm voxel; either 1 or 3 frame averages) and higher (HR; 2000 projections; 10 µm voxel; 3 frame averages) spatial resolution (µCT50, Scanco Medical), maintaining the same digital registration within specimen holders (Fig. 1). *Model development and generalizability:* 3D image arrays were imported for developing DL models in Python coding language, utilizing cropped co-registered 2D image "slices" in the transverse anatomical plane (LR = 200 pixels × 200 pixels; HR = 400 pixels × 400 pixels; Fig. 1), up-sampling LR to 400×400 via bilinear interpolation, and then grayscale normalizing (min-max scaling). Specimen images (femur/spine) were randomly assigned to either training (n=44/43), validation (n=12/14), or testing (n=15/14) groups for model development, yielding 28,400 femur images (400/femur) and 36,788 spine images (541/spine). Each DL model utilized a U-Net framework[5] and underwent training for 100 epochs using a 0.001 learning rate, Adam optimizer, and cross-entropy loss function. Model accuracy was assessed using peak signal-to-noise ratio (PSNR) and structural similarity index measure (SSIM). To evaluate model generalizability, the DL model developed for femurs was tested on spine and vice versa, evaluating outcomes via PSNR/SSIM metrics. *Bone morphometry and statistics:* 2D slice-wise bone morphometry of bone area fraction (BA/TA) of the LR, HR, and DL-generated (testing group only) images was quantified using an open-source Python package (ITKBoneMorphometry; 1.2 Gaussian blur and 500 mg hydroxyapatite/cm^3 lower threshold cut-off). Confidence interval calculations and t-tests were performed to measure differences in group PSNR, SSIM, and BA/TA for LR vs. HR and DL vs. HR images. Significance was set at $p < 0.05$, and Bland-Altman plots visualized differences in BA/TA between groups.

RESULTS: DL models improved image resolution in the femur and spine based on qualitative and quantitative DL metrics, with little improvement when LR images were acquired at 3 frame averages (Fig. 2A; example data for applying one spine model of LR 3 frame to HR 3 frame; other models were equivocal). Qualitative interpretation of the DL-generated images indicates an appearance of image smoothing or denoising and improvements in resolving the smallest pores evident at HR (Fig. 2A). Examining the DL metrics for each spine specimen revealed diversity in model accuracy, yielding higher results for some and lower for others (Fig. 2B; SSIM shown, but PSNR follows a similar pattern). DL-generated images have a comparable BA/TA to the HR images, which is lacking at LR (Fig. 2C). Qualitatively and quantitatively, the DL models generalized well across bone anatomy and skeletal site, as seen when reversing DL model application by the skeletal site (data not shown, demonstrating equivocal PSNR/SSIM metrics).

DISCUSSION: This study developed DL models that successfully improve microCT image spatial resolution post-acquisition. The models generalize well across bone anatomy and microarchitecture while reproducing bone morphometry obtained at higher resolution. DL models were built upon a large sample size and the wide-ranging bone microarchitecture associated with the Diversity Outbred mouse population. Additional variance in images was achieved by including 2 skeletal sites, males and females, and treatment with parathyroid hormone, in turn providing robust model training and testing. Although baseline metrics of PSNR and SSIM comparing LR to HR images were relatively high due to general co-registered image congruity, the improvement of ~8% for DL to HR images was associated with matching the presence of small pores and contrast at boundaries. Importantly, this study preserved image registration by maintaining specimens in a fixed spatial position, allowing only the minor inaccuracy associated with the movement of the scanner stage. Overall, these results demonstrated an improvement in using DL to increase image resolution compared to prior reports,[4] possibly due to differences in image registration, microCT instrumentation, variance in specimens, DL model structure (U-Net vs. other), image color scaling, and/or image spatial resolution. Ongoing work is evaluating specific variables that dictate variation in accuracy, toward adapting this DL approach to higher spatial resolution in specimen imaging and clinical research using high resolution peripheral quantitative CT (HR-pQCT).

SIGNIFICANCE/CLINICAL RELEVANCE: This study demonstrates that DL models can improve the post-acquisition resolution of microCT images of the murine skeleton and is a proof-of-concept for translating DL approaches into clinical research for improved diagnosis and monitoring of bone conditions.

ACKNOWLEDGMENTS: NIH R01AR070879 and R01AR073346. **REFERENCES:** [1] Yang+ *IEEE* 2019; [2] Guha+ *SPIE* 2020; [3] David+ *ORS* 2023; [4] Jhuboo+ *EUSIPCO* 2022; [5] Ronneberger+ *arxiv* 2015

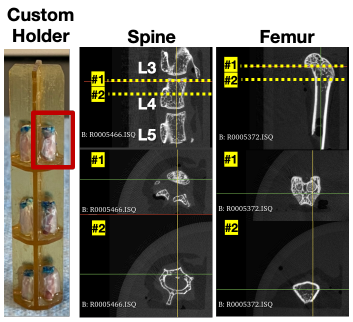

Fig. 1: Image of specimen holder used for micro-CT imaging, and example images of the spine and femur used to develop the DL model (red box=1 bone) visualized using 3D Slicer software.

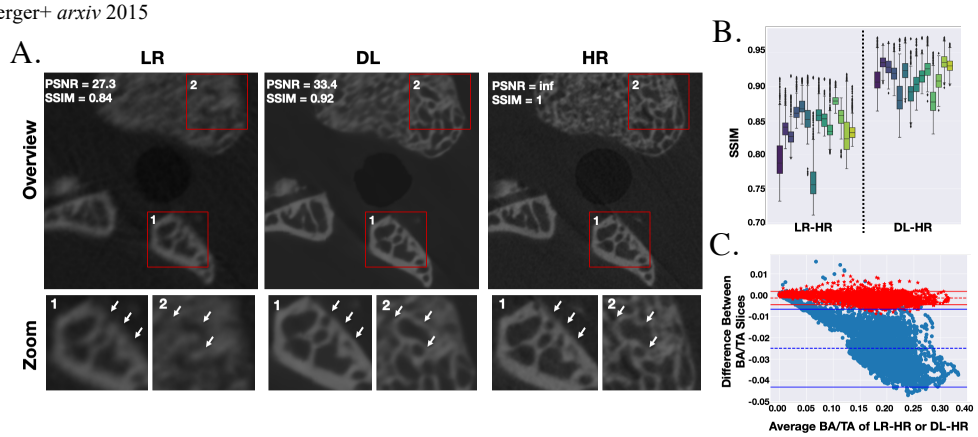

Fig. 2: (A.) MicroCT spine images showing DL-generated images from LR, and PSNR/SSIM vs HR. Red boxes show DL improved resolution of pores within the spongiosa and cortical bone (arrows). (B.) Average SSIM for LR vs HR and DL vs HR for 14 spines (C.) Bland-Altman plots show a decrease in the mean differences and confidence interval (dotted and solid lines, respectively) for BA/TA between DL-HR (red) vs LR-HR (blue).