# Generalizing Trabecular Bone Super-resolution to Out-of-distribution Data

Trevor J. Chan, Chiadika Eleh, Rasleen Grewal, Christiana Cottrell, Nada Kamona, Caroline Paik, David Barreto, Ganen Chinniah, Chamith S. Rajapakse
University of Pennsylvania, Philadelphia, PA
tjchan@seas.upenn.edu
**Disclosures:** Trevor J. Chan (N), Chiadika Eleh (N), Rasleen Grewal (N), Christiana Cottrel (N), Nada Kamona (N), Caroline Paik (N), David Barreto(N), Ganen Chinniah (N), Chamith S. Rajapakse (N)

**INTRODUCTION:** The ability to resolve detailed trabecular microstructure through non-invasive imaging carries numerous benefits. Quantitative metrics derived from trabecular microstructure in conjunction with DXA, which measures only bone mineral density, provide a more comprehensive view of bone health. Clinically, this could allow for more sensitive tests for diagnosing osteoporosis and assessing patient responses to treatments.

Despite the benefits, there are numerous obstacles limiting the incorporation of high-resolution trabecular bone imaging in clinical practice, particularly in the central skeleton, including the hip and spine. These include hardware limitations–CT scanners capable of the sub 0.2 mm resolution necessary to accurately capture trabecular bone are expensive and rare–and safety limitations–scanning at higher resolutions necessitates a higher radiation dose to the patient. Recently, groups have developed deep learning computational methods able to sidestep these constraints by imputing high-resolution spatial information from lower resolution routine clinical scans. Adapting these super-resolution models to bone imaging, they allow clinicians to resolve trabecular bone structure from routinely acquired CT images, without the need for increased radiation or specialized hardware.

An open question remains as to the flexibility and generalizability of these methods beyond the narrow application areas these models were initially developed for. As with most deep learning models, the performance and consistency of super-resolution models is highly dependent on the quality, size, and diversity of the dataset used to train the model–in this case, high-resolution scans of trabecular bone. However, due to the aforementioned obstacles, obtaining such a dataset is, at a minimum, time consuming and expensive and, in the worst case, potentially dangerous (i.e., high radiation dose). A feasible alternative is to obtain a training set of high-resolution images from cadaveric scans, where radiation dose is not a concern. However, it is unknown how well a model trained on cadaveric data can generalize to patient data. Therefore, we examined the performance of a super-resolution model trained only on high-resolution cadaveric CT scans of the femur on a range of out-of-distribution images.

**METHODS:** We trained a diffusion deep learning super-resolution model on a training set of 20 human cadaveric femur HR-pQCT scans, comprising 70,000 axial images at 30 micron resolution. Validation and testing sets comprised an additional 2 and 4 cadaveric femurs respectively. We additionally tested super-resolution performance on a range of images external to the training set. These include cadaveric scans of the femur on a Siemens clinical CT scanner and a Siemens photon counting CT scanner, patient scans on a Siemens clinical CT scanner, and patient CT scans of the skull. The use of data in this study was approved by institutional review board. In this preliminary study, we assessed performance subjectively using visual comparison.

**RESULTS SECTION:** Model performance recovering trabecular bone microstructure from retrospectively downsampled HR-pQCT images of cadaveric femurs (considered in-distribution) was predictably very high. Across four out-of-distribution images, performance ranged from plausible to implausible, with the plausibility of model outputs largely correlating with image quality, and not the anatomical similarity, of the initial images. The four conditions were: scans of a cadaver femur in a standard clinical CT scanner, scans of a cadaver femur in a photon counting CT scanner, patient routine abdominal pelvic scans (here cropped to the femur) in a standard clinical CT scanner, and retrospectively downsampled CT scans of a patient's skull (cropped to the temporal bone). Representative images are shown in figure 1. Of these, reconstruction in the skull CT and photon counting CT images seemed to perform the best, while the lower resolution clinical CT scans, both for cadaver and for patient images, exhibited unnatural artifacts, such as the creation of trabeculae in soft tissue regions (Figure 1, row 3, column 2) and the bleeding of trabeculae into the acetabulum (Figure 1, row 3, column 4).

**DISCUSSION:** Conventional wisdom surrounding deep learning models dictates that performance on images outside of the training distribution would be severely impacted. Surprisingly, our model performed well when qualitatively evaluated on these out-of-distribution images. Future work will investigate the reasons behind the good performance on out-of-distribution images, and why the model fails to produce plausible outputs in some cases.

There are major limitations to this preliminary study, chiefly the lack of validation. Without knowing additional information about these specimens/patients, we can only base conclusions of model performance and generalizability on the plausibility of outputs, not the accuracy. Therefore, future work will also focus on verifying the accuracy of model outputs on external data by incorporating a combination of high-resolution ground truth scans and clinical diagnoses, as well as a levy of quantitative trabecular microstructural metrics.

**SIGNIFICANCE/CLINICAL RELEVANCE:** This work presents preliminary evidence that deep learning super-resolution models trained to resolve trabecular microstructure in clinical resolution CT scans are capable of a higher degree of generalization than previously thought. Such insights are a step towards comprehensive bone health evaluations using low cost and low radiation imaging modalities.

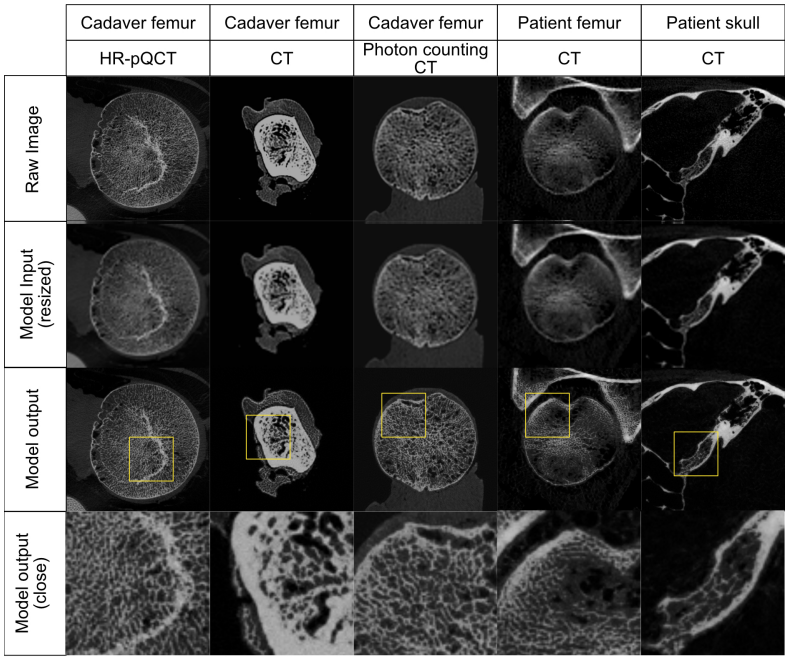


**Figure 1**. Representative images show one example of super-resolution performance on an in-distribution image (Left column, cadaveric femur scanned on a HR-pQCT system) and four examples of super-resolution performance on out-of-distribution images (Columns 2-5). For each condition, we show an initial raw image (row 1), a resized image for input into the model (row 2), and the model prediction for the high-resolution image (row 3). We can qualitatively assessed the plausibility of the trabecular bone microstructure in the model outputs, but we cannot infer their accuracy without a ground truth comparison.