Generating Knee CT Scans from Bi-Planar Radiographs Using a Modular Diffusion Model

Bardia Khosravi, Pouria Rouzrokh, Elizabeth Sun Kaji, Austin Grove, Joaquin Sanchez-Sotelo[1], A. Noelle Larson, Bradley James Erickson[1], Michael J Taunton[1], Cody Wyles[1]
[1]Mayo Clinic

INTRODUCTION: Purpose: To develop a modular diffusion model capable of generating multi-resolution CT scans from bi-planar knee radiographs, reducing radiation dose and enhancing preoperative planning in orthopedic surgery.
METHODS: A dataset of 16,730 CT scans and 54,340 AP and lateral knee radiographs from patients aged 28-90 (54% female) across three tertiary centers was used. The 400-million-parameter model is based on the Matryoshka diffusion architecture, which is a powerful generative model that iteratively denoises Gaussian noise to create realistic images. The model encodes radiographs using a ConvNext CNN, and the features are cross-attended during the diffusion process to generate CTs at 32^3, 64^3, and 128^3 resolutions. The model is conditioned on pixel spacing and
slice thickness. The 128^3 volumes are supersampled to 512^3 using a patch-based diffusion super-resolution model. The model was incrementally trained on batch sizes of 224, 80, and 16. Generated bone accuracy was evaluated using the bone overlap between real and generated CTs on 10 samples after registration.
RESULTS: Model training on 16 A100 GPUs took 10 days. The model generated 128^3-resolution CTs in 2 minute and 27 seconds, with supersampling to 512^3-resolution taking an additional 6 minutes (under 9 minutes total). The generated bone had an 84% overlap with the original CT images, **indicating a high degree of correspondence between the generated and real CT scans**. The Fréchet Inception Distance (FID), a metric quantifying the similarity between real and generated images, was 32.8 on 1000 generated images, suggesting that the generated images have a very close distribution to real CT scans.
DISCUSSION AND CONCLUSION: The modular diffusion model successfully generates high-resolution CT scans from bi-planar knee radiographs, offering a radiation dose reduction solution. The integration of this technology with robotic surgery, virtual reality, and augmented reality will enhance surgical planning and execution.