Deep-Learning-Based Super-Resolution and Segmentation for Clinical and Research Musculoskeletal MRI

- Akshay Chaudhari, Zhongnan Fang, Feliks Kogan, Eric Gibbons, Jeff Wood, Kathryn Stevens, Jin Hyung Lee, Garry Gold, Brian Hargreaves.

Target Audience: MR physicists, musculoskeletal radiologists and disease researchers, deep learning researchers and practitioners.

Purpose: Near-isotropic high-resolution MRI of the knee is beneficial for both clinical and research applications. Clinically, near-isotropic resolutions can reduce partial volume effects and enable interrogating tissues in arbitrary oblique planes. Similarly, such resolution is vital for accurately measuring morphological changes of thin tissues like cartilage in researching diseases such as osteoarthritis[1]. However, previous methods exploring high resolution scans typically compromised in-plane resolution for thin slices due to signal-to-noise ratio (SNR) limitations[2]. Moreover, segmenting tissues from high-resolution MRI scans is a laborious manual procedure that is not feasible for large clinical studies. To overcome such limitations, we explore deep-learning used for both super-resolution and segmentation methods in musculoskeletal MRI. Using 3D convolutional neural networks (CNNs), we demonstrate a **threefold through-plane resolution enhancement** and use the resultant images to **fully-automate cartilage segmentation**.

Theory: We implemented two 3D CNNs, one each for super-resolution (SR-CNN) and segmentation (Seg-CNN). SR-CNN learned a residual difference function between low-resolution and high-resolution input patches from knee MR images, inspired by previous work[3]. Such a residual function could be applied to arbitrary low-resolution images during testing time to generate super-resolution images. Seg-CNN, a 3D extension of the U-Net architecture, utilized dynamic learning rates and network regularization to learn volumetric segmentations of femoral cartilage using manual segmentation labels[4]. A state-of-the-art musculoskeletal MRI segmentation network (2D Segnet) was also implemented for use as a benchmark[5].

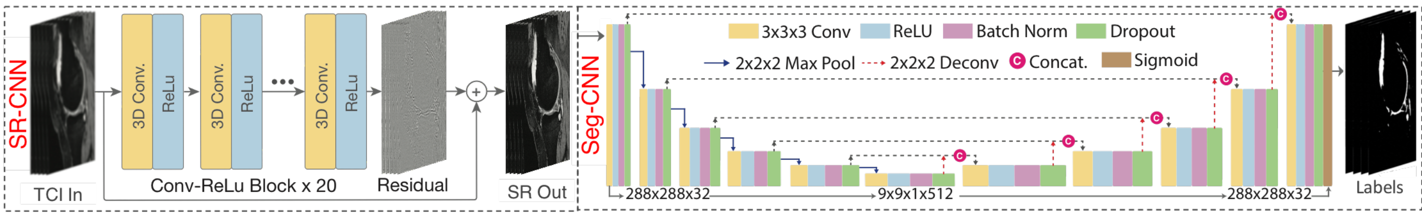

Figure 1: *3D SR-CNN super-resolves (SR) 3x downsampled and interpolated images (TCI) for performing femoral cartilage segmentation in 3D Seg-CNN.*

Data: Both networks (Fig. 1) were trained using 3D double-echo steady-state (DESS) sequences obtained from the Osteoarthritis Initiative (OAI) (relevant scan parameters: FOV=14cm, Matrix=384x307 (zero-filled to 384x384), TE/TR=5/16ms, 160 slices with a thickness of 0.7mm)[1]. 176 patients in total were split in 124 patients for training, 35 for validation, and 17 for testing. All datasets had manual femoral cartilage segmentations available.

Methods: Low-resolution DESS images with 3x thicker slices were synthesized from the high-resolution OAI data using a finite impulse response filter and were tricubicly interpolated (TCI) at the ground-truth slice locations. The TCI images were the inputs to the SR-CNN, which would eventually be transformed into the SR images using learned residuals. For SR-CNN training, each imaging volume was divided into voxel patches of size 32x32x32 with a 16x16x16 stride (~0.5 million total patches). SR-CNN consisted of 20 layers with 64 feature maps each and was trained with L2-loss using a static learning rate of 0.0001 over 20 epochs. Image quality of the 3x downsampled output SR and the input TCI was compared to the ground-truth using structural similarity (SSIM) and peak-signal-to-noise ratio (pSNR). Two musculoskeletal radiologists assessed the diagnostic image quality (DIQ) (1=poor, 5=excellent) of the ground-truth, TCI, and SR-CNN images concurrently in all three scan planes and also ranked the quality of all three image sets (1=worst, 3=best). All images were blinded and presented in random orders.

For Seg-CNN, each volume was divided into patches of 288x288x32. Seg-CNN utilized dynamic learning rates (initial rate = 0.001, 60% drop every 5 epochs, for 50 epochs), a Dice coefficient score (DSC) loss, and 25% dropout. Segmentation was performed on the ground-truth, TCI, and SR-CNN images. Segmentation was also performed on the ground-truth images using 2D-Segnet as a benchmark. DSC and volumetric overlap error (VOE%) were used to evaluate segmentation accuracy on all image sets.

3D SR-CNN Performance			
	Ground-Truth	SR-CNN	TCI
SSIM	N/A	79.1 ± 1.3	73.8 ± 1.3*
pSNR	N/A	29.1 ± 1.0	26.8 ± 1.2*
DIQ	4.4±0.5	3.8±0.8†	3.1±1.0†*
Rank	2.8±0.2	2.1±0.5†	1.1±0.4†*
3D Seg-CNN Accuracy			
	Ground-Truth	SR-CNN	TCI
DSC	90.2 ± 1.8	89.6 ± 2.0	86.3 ± 3.2†
VOE	17.8 ± 3.0	18.8 ± 3.3	24.0 ± 4.9†

Table 1: *SR-CNN and Seg-CNN metrics.*
**: Significantly different than SR-CNN (p<0.001).*
†: Significantly different than ground-truth.

Wilcoxon signed-rank tests (WSRT) evaluated variations in SSIM and pSNR for SR-CNN and TCI (compared to the ground-truth), and also differences in overall diagnostic image quality. WSRT also compared whether the SR-CNN segmentation DSC and VOE scores were comparable to the ground-truth and TCI scores. A comparison of 3D Seg-CNN and 2D Segnet DSC and VOE scores on the ground-truth data was also performed using WSRT.

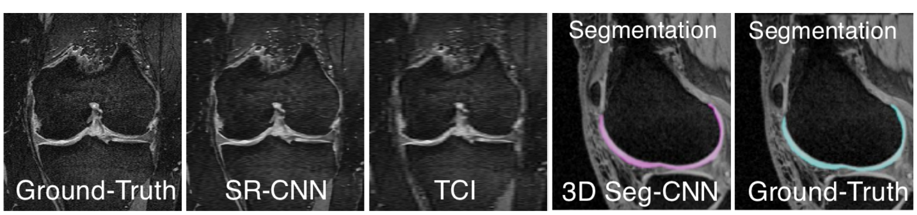

Figure 2: *Example coronal ground-truth SR-CNN, and TCI images and sagittal Seg-CNN segmentations.*

Results: All quantitative results are summarized in Table 1 and example images shown in Fig. 2. SR-CNN had significantly higher SSIM and pSNR than TCI images (p<0.001). In the reader study, ground-truth images were ranked the best and had the highest diagnostic image quality (DIQ), however SR-CNN had significantly higher DIQ and ranking (p<0.01) than TCI. For segmentation, SR-CNN had near identical DSC and VOE scores as the ground-truth, while TCI had significantly worse accuracies (p<0.01). The 3D Seg-CNN network significantly outperformed (p<0.0005) the state-of-the-art 2D Segnet (DSC=83.0±2.4% and VOE=29.0±3.5%). The inference duration for SR-CNN was 10s per volume, while Seg-CNN required 1s per volume.

Discussion and Conclusion: 3D SR-CNN considerably outperformed the commonly-utilized TCI method for enhancing through-plane image resolution. SR-CNN had better quantitative image quality metrics than TCI as well as better diagnostic image quality ratings from radiologists. While SR-CNN did not match the ground-truth image quality, implementing the long DESS sequence to acquire 0.7mm slices (11 min) is not pragmatic in clinical or research applications. The SR-CNN images had near identical femoral cartilage segmentation accuracy as the ground-truth images in Seg-CNN, demonstrating minimal blurring of thin tissues such as cartilage. 3D Seg-CNN outperformed the current state-of-the-art 2D Segnet method, likely due to 3D convolutions and the use of dropout to increase network generalizability. Overall, we have demonstrated the promise of combining a near-instantaneous super-resolution and segmentation method that may enhance the quality, speed, and value of both clinical and research workflows.

References: 1: Peterfy, OAC, 2008. 2: Kijowski, Radiology, 2009. 3: Kim, CVPR, 2016. 4: Ronnenberger, MICCAI, 2015. 5: Liu, MRM, 2017.