Improved Auto-Segmentation for CT Male Pelvis: Comparison of Deep Learning to Traditional Atlas Segmentation

C. Halley, H. Wan, Ph.D, A. Kruzer, D. Pittock, D. Darkow, MEM, M. Butler, N. Cole, Ph.D, M. Bending, P. Jacobs, MS, A.S. Nelson, MD

MIM Software Inc., Cleveland, OH, United States

Introduction

When it comes to segmenting anatomical structures on medical images, there are two key features which help the user experience and with the actual diagnosis/treatment planning: speed and accuracy. While traditional atlas contouring methods are relatively fast and accurate, 3D Convolutional Neural Networks (CNN) can significantly improve accuracy for many structures, which may correlate to a reduction in post processing required and time spent contouring. The purpose of this study is to compare the accuracy of a CNN-based automated contouring method to an atlas for male pelvis on CT.

Methods

The atlas segmentation method used the largest overlapping region of the top 5 most similar images (according to Pearson Correlation Metric), given a majority vote (3/5) of said images. The CNN segmentation method was based on RefineNet[1] with additional 3D convolution blocks to leverage contextual information in all directions, an innovative update to the more commonly referenced U-Net architecture[2]. The atlas contained 35 images, and the CNN was trained on 320 separate images, all of which were expert-segmented. Both algorithms were run on the same 35 atlas images and compared to gold-standard (GS) manual segmentations for 6 anatomical structures: prostate, bladder, rectum, both femurs, and seminal vesicles. A leave-one-out analysis was used on the atlas to avoid using images to segment themselves. The Dice score, mean distance to agreement (MDA), and the Hausdorff 95th percentile distance (HD95) were calculated for both methods. Statistically significant improvement was calculated via a two-sample t-test on each structures' statistics.

Results

Figure 1 and **Figure 2** display Atlas (blue), CNN (red), and GS (gold) contours for the right femur and bladder, respectively, of one subject as an example. These images were derived after the initial study to illustrate the trends observed during testing. The values displayed in the graphs are averages across all the segmentations available from the datasets tested. **Figure 3** displays the Dice score. **Figure 4** displays the MDA. **Figure 5** displays the HD95. Structures with an asterisk (*) in the figures showed significant differences between the two methods. Across the majority of structures, the CNN displayed either significant improvement in accuracy or no significant change. However, the HD95 for the rectum, HD95 for the right femur, and MDA for the right femur structures all showed significant improvement through the atlas, instead.

Conclusion

The CNN method was frequently more accurate than the traditional atlas method. However, the atlas performed better on the rectum and right femur. The data used to train the CNN included multiple segmentation styles, generating a single, consistent style for output production. Meanwhile, the atlas was consistently segmented in a single style; different than the CNN. These differences in contouring styles may have led to the atlas performing better than the CNN for the rectum and femurs: structures known for having widely varying contouring styles. In the future, we plan to analyze potential solutions to stylistic differences and analyze time savings, as well as accuracy.

References

[1] G. Lin, A. Milan, C. Shen, and I. Reid, "RefineNet: Multi-path Refinement Networks for High-Resolution Semantic Segmentation," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017.

[2] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," Lecture Notes in Computer Science Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, pp. 234–241, 2015.

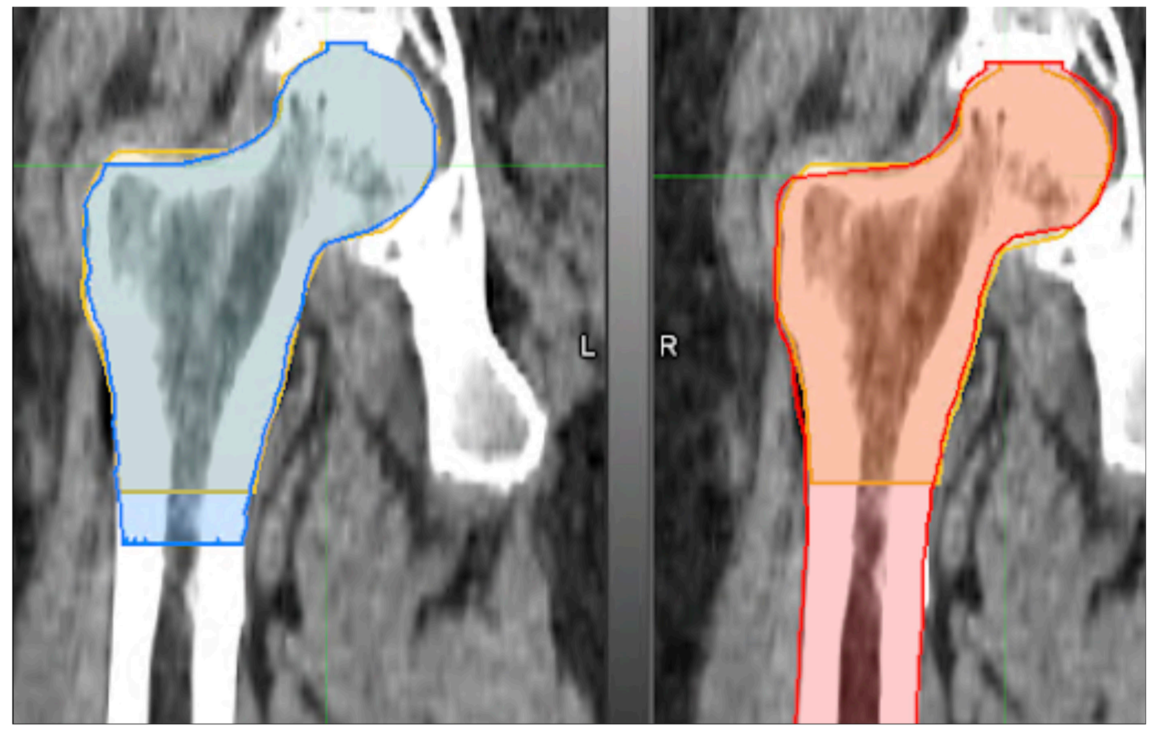

Figure 1: Screen capture comparing Atlas to GS Contours (Left) and CNN to GS Contours (Right) for the right femur

NN
Atlas
GS

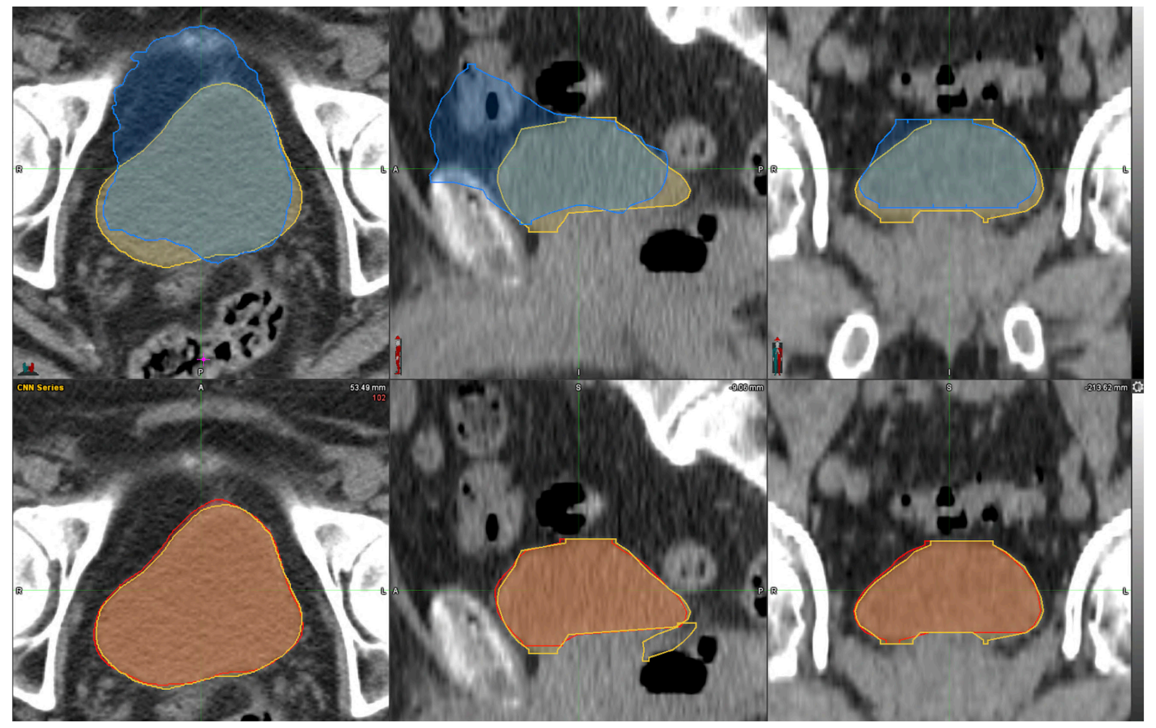

Figure 2: Screen capture comparing Atlas to GS Contours (Top) and CNN to GS Contours (Bottom) for the bladder

NN
Atlas
GS

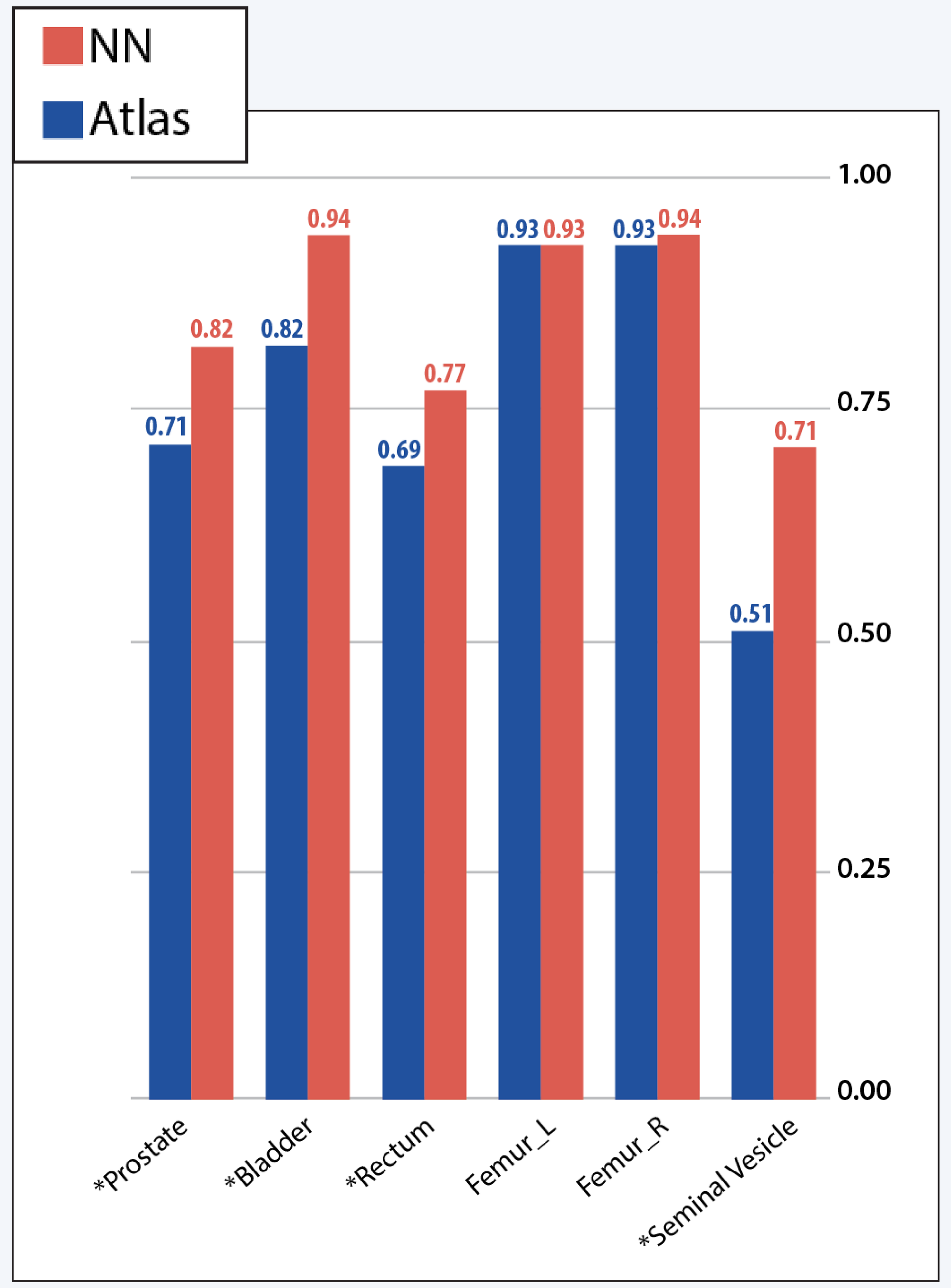

Figure 3: Mean Dice Comparison between CNN and Atlas

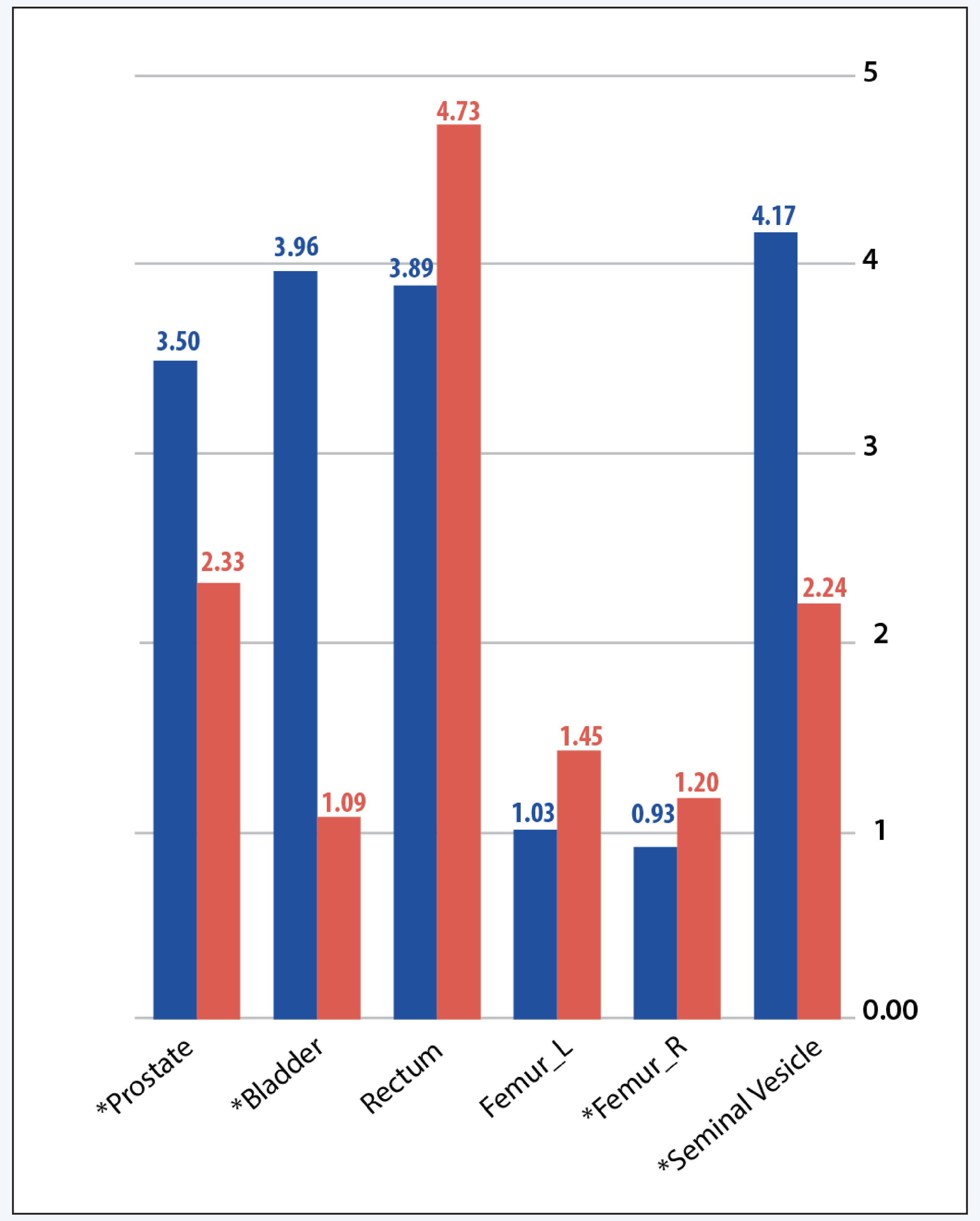

Figure 4: Mean MDA Comparison between CNN and Atlas in mm

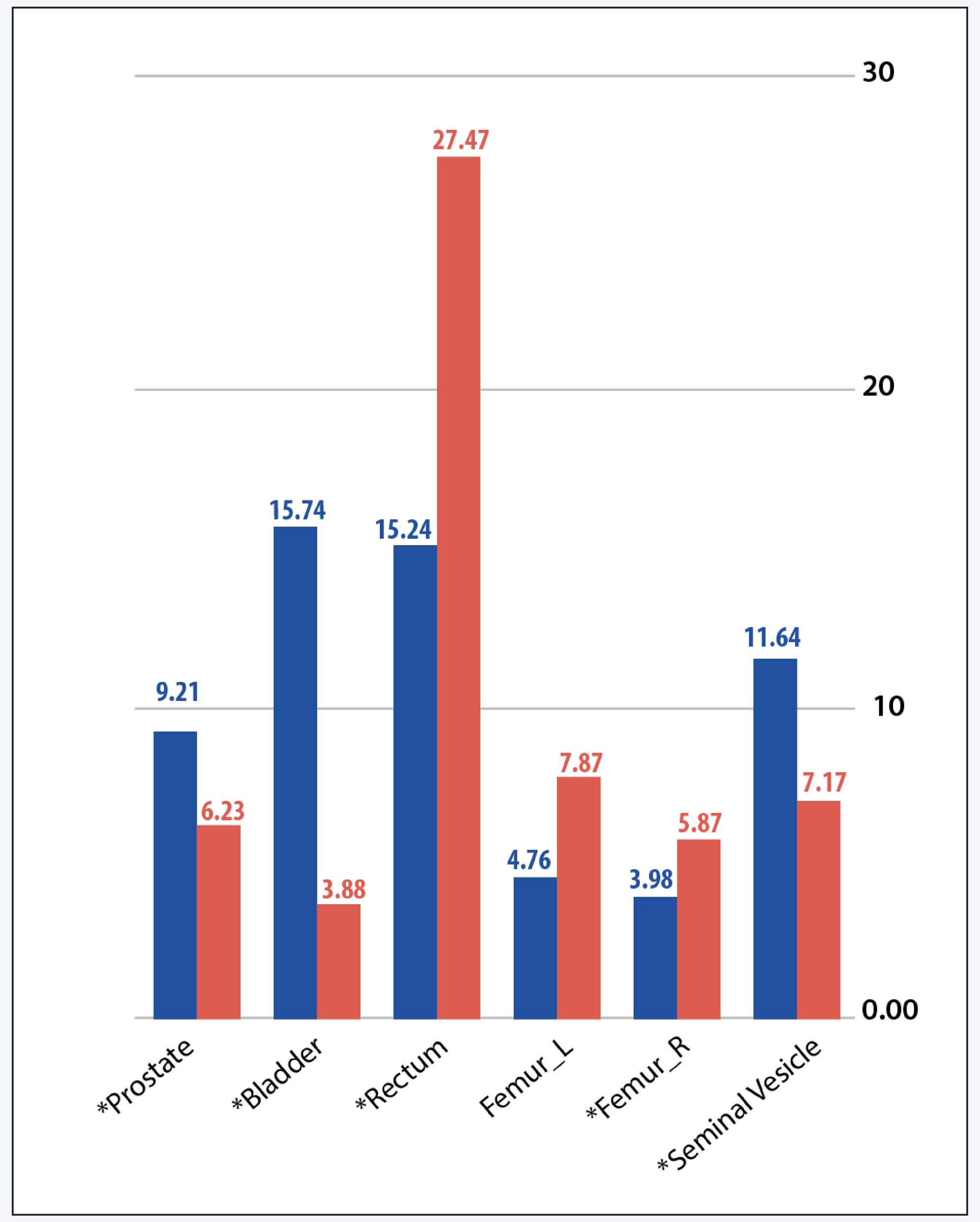

Figure 5: Mean HD95 Comparison between CNN and Atlas in mm