Comparison of a 3D convolutional neural network segmentation method to traditional atlas segmentation for CT head and neck contours.

A. Kruzer, H. Wan, M. Bending, C. Halley, D. Darkow, D. Pittock, N. Cole, P. Jacobs, A.S. Nelson

MIM Software Inc., Cleveland, OH

Purpose

The purpose of this study was to compare the efficacy of a 3D convolutional neural network (CNN) segmentation algorithm with a traditional atlas segmentation approach when applied to head and neck contours on CT.

Materials and Methods

A CNN algorithm was trained on a group of 721 subjects from multiple institutions. Ten structures were included: mandible, brainstem, eye (L and R), optic nerve (L and R), optic chiasm, parotid (L and R), and spinal cord. A separate atlas image database was built from 20 subjects with the same structures included. These images were not included in the CNN training set.

Both the CNN and atlas were used to segment the group of 20 atlas subjects from multiple institutions using manual segmentations as ground truth. The atlas segmentation method was configured to use the largest overlapping region of the five most similar images (according to Pearson Correlation Metric), given a majority vote (⅗) of said images. A leave-one-out method was used for the atlas segmentation to prevent any image being used to segment itself.

Mean Dice Similarity coefficient (DSC), mean distance to agreement (MDA), and mean 95% Hausdorff distance (HD95) were calculated for both segmentation methods.

Results

A graphical summary of results is shown through Figures 2, 3, and 4. Figure 2 shows the Dice Similarity Coefficient (DSC) calculated for both the CNN and the atlas segmentations compared with the ground truth. Similarly, Figure 3 shows Mean Distance to Agreement (MDA), and Figure 4 shows Mean 95% Hausdorff Distance (HD95). The HD95 was reported as the max HD at the 95% percentile of the data distribution. Statistical significance was determined using a two-sample t-test. Statistically significant differences are denoted with an asterisk.

The CNN produced statistically significant improvements as follows (CNN, Atlas): Right Eye Dice (0.88, 0.85); Optic Chiasm Dice (0.30, 0.10), MDA (2.29mm, 3.47mm), and HD95 (5.22mm, 8.23mm); Left Parotid Dice (0.79, 0.70); and Right Parotid Dice (0.81, 0.74). The atlas segmentation performed better for the Brainstem MDA (1.56mm, 1.21mm) and for the Mandible MDA (0.70mm, 0.56mm) and HD95 (2.57mm, 1.91mm). No significant change was found for the remaining statistics.

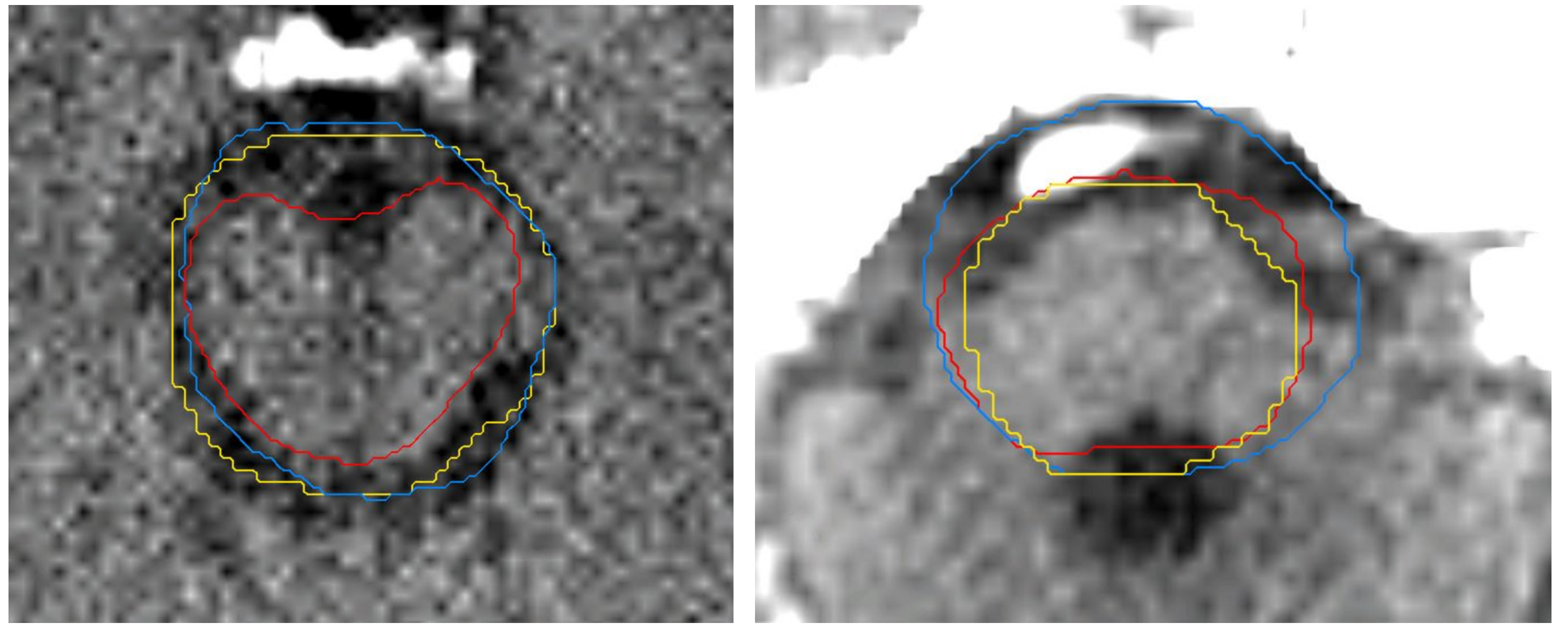

Figure 1. Two examples of results on the brainstem. (A) An example subject where the atlas contour resulted in better Dice and HD compared to the NN contour. (B) An example subject where the NN contour resulted in better Dice and HD compared to the atlas. Reference (yellow), Atlas (blue), NN (red)

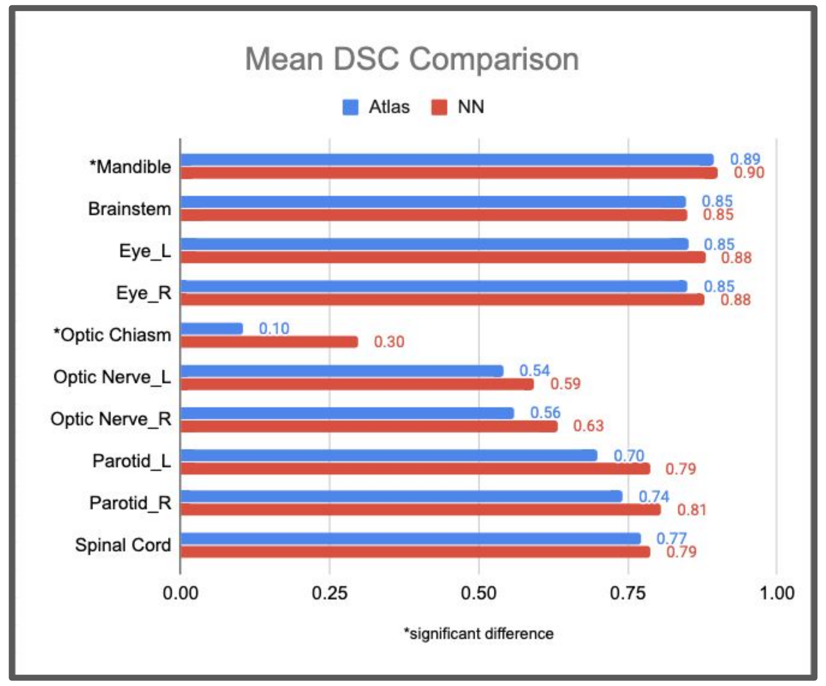

Figure 2. Mean Dice Comparison between CNN and Atlas

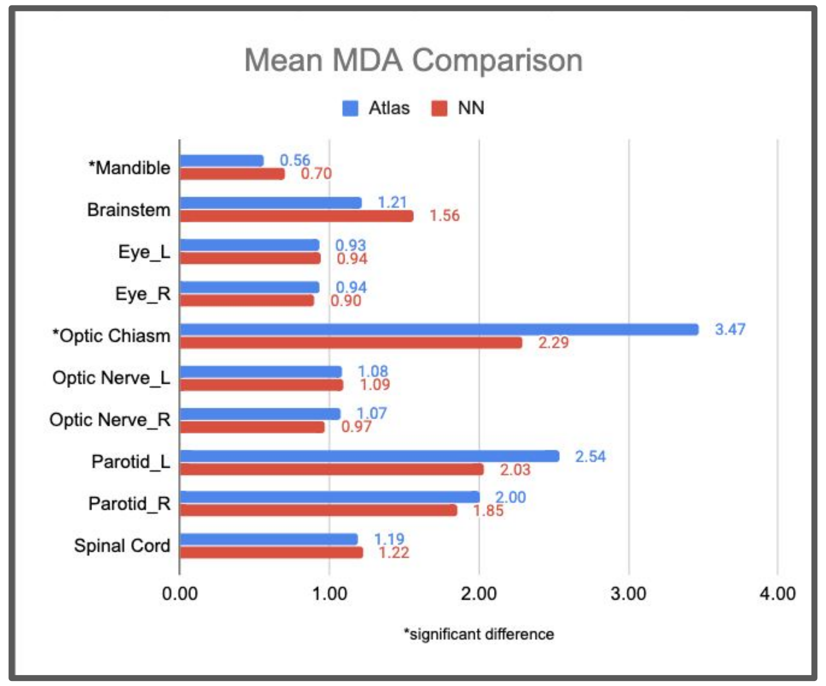

Figure 3. Mean MDA Comparison between CNN and Atlas (mm)

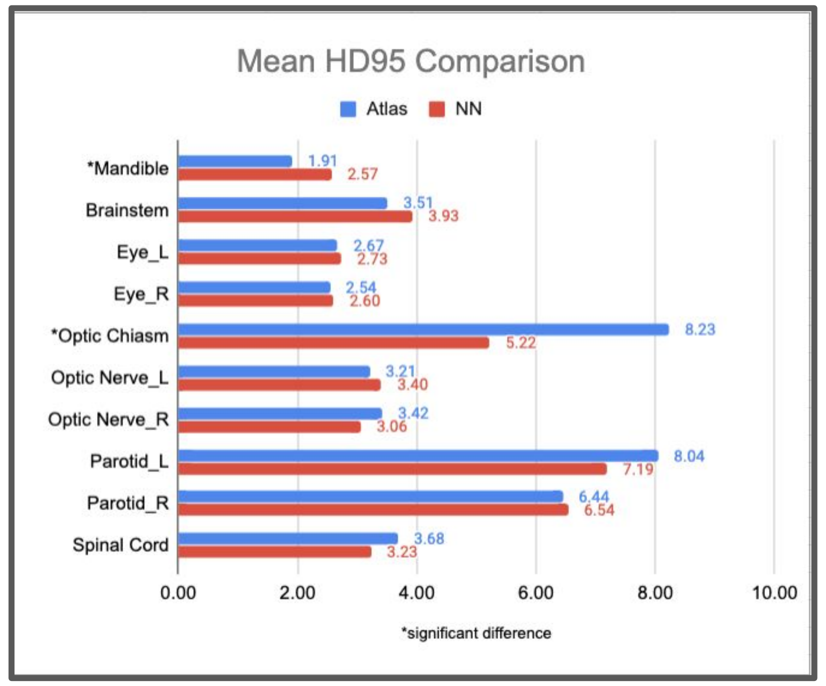

Figure 4. Mean HD95 Comparison between CNN and Atlas (mm)

Innovation and Impact

This study is impactful because it validates the use of neural networks to automate the task of contouring structures on head and neck CT, a tedious task that while aided by traditional atlas segmentation, often still requires manual editing. Further, the convolutional neural network-based auto-segmentation approach (CNN) was based on the RefineNet[1] method with additional 3D convolution blocks added in order to leverage contextual information in all directions, an innovative update to the more commonly referenced U-Net architecture[2].

For eight out of ten structures, the CNN method was found to be the same or better than atlas segmentation. The results seen in the mandible and brainstem appeared to be caused by a stylistic difference between CNN output versus atlas output. 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.