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


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 Metrics), given a majority vote of 5 of 10. 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 (H95S) 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) comparison for both the CNN and the atlas segmentation compared with the ground truth. Similarly, Figure 3 shows Mean Distance to Agreement (MDA), and Figure 4 shows Mean 95% Hausdorff Distance (H95S). The H95S 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 vs. Atlas): Right Eye Dice (0.88, 0.85); Optic Chiasm Dice (0.30, 0.31); MDA (2.29mm, 5.47mm); and H95S (5.22mm, 8.23mm); Left Parotid Dice (0.79, 0.76); 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 (0.10mm, 0.50mm) and H95S (2.57mm, 1.91mm). No significant change was found for the remaining statistics.

Innovation and Impact

This study is important because it validates the use of neural networks to automate the task of contouring structures on head and neck CT, a tedious task that was only done by traditional atlas segmentation, often still requires manual editing. Further, the convolutional neural network-based auto-segmentation approach (CNN) was based on the RefineNet1 method with additional 3D convolutional blocks added in order to leverage contextual information in all directions, an innovative update to the more commonly referenced U-Net architecture2.

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
