Supplementary

A. Boxplots of the results of the five-fold cross validation on two individual scanners.

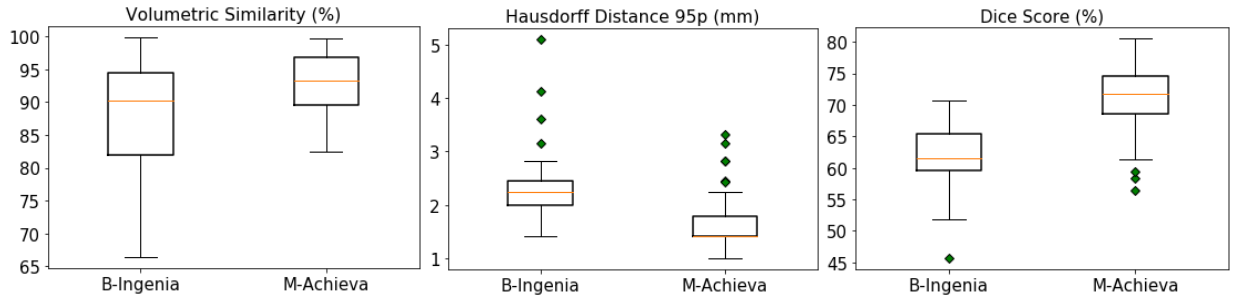

***Figure S1**: Results of individual scanners validation on the 2nd scanner (Bonn-Ingenia) and the 3rd scanner (Munich-Achieva). Each box plot summarizes the segmentation performance on images from two scanners using one specific metric.*

B. Training process with training loss and validation loss.

In the k-fold cross-validation experiment, we split the training set into a nested training set and a validation set by randomly picking 80% and the remaining 20% scans from each scanner respectively. Figure S2 shows the curves of training loss and validation results over 350 epochs. It could be observed that the validation result (i.e. volumetric similarity and Dice score) did not show an ascending trend at around 150 epochs. Hence, we choose a number 150 epochs to avoid over-fitting and to keep a low computational cost.

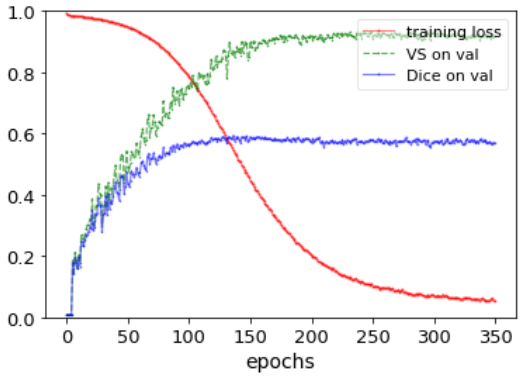

***Figure S2**: The curves of training loss and validation results over 350 epochs.*

C. The correlation between Dice scores and claustrum volumes.

We plot the distribution of ground-truth volumes of claustrum and Dice scores achieved by illustrating that Dice scores correlate with the volumes of regions of interest.

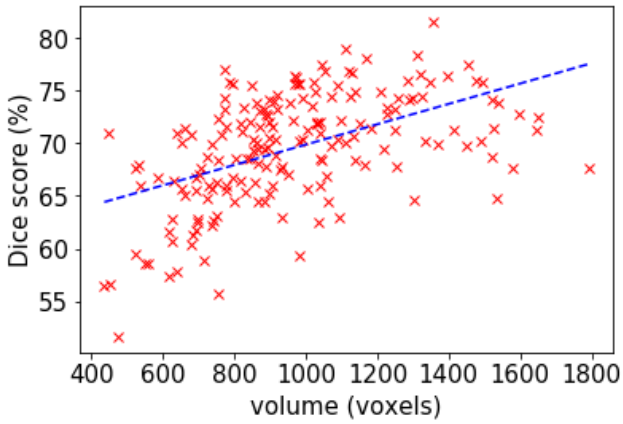

***Figure S3**: The distribution of volumes of claustrum and Dice score. Dice scores are highly correlated with the volumes of claustrum and may be low when the volume is small.*