Per-Dataset Calibration using Silver Standards and Transfer Learning for Improved Generalization in FLAIR MRI Lateral Ventricle Segmentation

Owen Crystal[1] OCRYSTAL@TORONTOMU.CA
Alan R. Moody[2] ALAN.MOODY@SUNNYBROOK.CA
Pejman J. Maralani[2] PEJMAN.MARALANI@UTORONTO.CA
April Khademi[1] AKHADEMI@TORONTOMU.CA
[1] *Electrical, Computer and Biomedical Engineering, Toronto Metropolitan University, Toronto, ON, Canada*
[2] *Department of Medical Imaging, University of Toronto, Toronto, ON, Canada*

Editors: Under Review for MIDL 2023

Abstract

Subjects with dementia suffer from ventricular enlargement making lateral ventricular (LV) volume an important biomarker for clinical investigation. Fluid attenuated inversion recovery (FLAIR) MRI is an ideal modality for LV segmentation due to the increased the CSF-to-tissue contrast. In this work, we present the first lateral ventricle segmentation deep learning model developed for FLAIR MRI while also using transfer learning to design dataset-calibrated models. Models were trained on silver standard (SS) masks from the target dataset, and fine-tuned on available gold standard (GS) masks. The following three models were tested on held-out test sets from four datasets: 1) trained on target SS masks and fine-tuned on source GS masks (All Layers), 2) trained on source GS masks (GS CAIN Only), 3) trained on target SS masks (SS Only). The All Layers model (mean DSC = 0.886) showed significantly ($p < 0.05$) higher DSC in comparison to the other two models and yielded the most consistent results measured by the coefficient of variation. The All Layers model, in comparison to the GS CAIN Only model, showed slight improvement when testing on the source domain and an increase of approximately 4% DSC when testing on the target domains. The results suggest that providing the model with noisy labels from the target dataset, allows to model to familiarize itself with specific features, thus increasing the performance. This can be used as a per-dataset calibration method, which can accelerate translation opportunities.

Keywords: Deep Learning, Neuroimaging, Transfer Learning, FLAIR MRI, Ventricle Segmentation, Silver Standards, Unsupervised, Noisy Labels, per-site calibration

© 2023 O. Crystal, A.R. Moody, P.J. Maralani & A. Khademi.