

Deep whole brain segmentation of 7T structural MRI

Karthik Ramadass^{*a}, Xin Yu^a, Leon Y. Cai^c, Yucheng Tang^a, Shunxing Bao^b, Cailey Kerley^b, Micah D'Archangel^d, Laura A Barquero^d, Allen T. Newton^{c,e}, Isabel Gauthier^f, Rankin Williams McGugin^f, Benoit M. Dawant^{a,b}, Laurie E. Cutting^d, Yuankai Huo^{a,b}, Bennett A. Landman^{a,b,c,d,e}

^aDepartment of Computer Science, Vanderbilt University, Nashville, TN, USA; ^bDepartment of Electrical and Computer Engineering, Vanderbilt University, Nashville, TN, USA; ^cDepartment of Biomedical Engineering, Vanderbilt University, Nashville, TN, USA; ^dVanderbilt Brain Institute, Vanderbilt University, Nashville, TN, USA; ^eVanderbilt University Institute of Imaging Science, Vanderbilt University, Nashville, TN, USA; ^fDepartment of Psychology, Vanderbilt University, Nashville, TN, USA

ABSTRACT

7T magnetic resonance imaging (MRI) has the potential to drive our understanding of human brain function through new contrast and enhanced resolution. Whole brain segmentation is a key neuroimaging technique that allows for region-by-region analysis of the brain. Segmentation is also an important preliminary step that provides spatial and volumetric information for running other neuroimaging pipelines. Spatially localized atlas network tiles (SLANT) is a popular 3D convolutional neural network (CNN) tool that breaks the whole brain segmentation task into localized sub-tasks. Each sub-task involves a specific spatial location handled by an independent 3D convolutional network to provide high resolution whole brain segmentation results. SLANT has been widely used to generate whole brain segmentations from structural scans acquired on 3T MRI. However, the use of SLANT for whole brain segmentation from structural 7T MRI scans has not been successful due to the inhomogeneous image contrast usually seen across the brain in 7T MRI. For instance, we demonstrate the mean percent difference of SLANT label volumes between a 3T scan-rescan is approximately 1.73%, whereas its 3T-7T scan-rescan counterpart has higher differences around 15.13%. Our approach to address this problem is to register the whole brain segmentation performed on 3T MRI to 7T MRI and use this information to finetune SLANT for structural 7T MRI. With the finetuned SLANT pipeline, we observe a lower mean relative difference in the label volumes of ~8.43% acquired from structural 7T MRI data. Dice similarity coefficient between SLANT segmentation on the 3T MRI scan and the after finetuning SLANT segmentation on the 7T MRI increased from 0.79 to 0.83 with $p < 0.01$. These results suggest finetuning of SLANT is a viable solution for improving whole brain segmentation on high resolution 7T structural imaging.

Keywords: T1-weighted, Deep learning, Whole brain segmentation, 7T MRI, Convolved Neural Network, Transfer Learning

1. INTRODUCTION

Ultra-high field 7T MRI provides increased spatial resolution and enhanced contrast to visualize the structures in the brain [1]. The enhanced details seen in a 7T MRI opens doors into several investigations such as exploring the relationship between cortical laminar structure and brain function [2][3], identifying anatomical and pathological features of multiple sclerosis (MS) [4], and accurate identification of ischemic lesions and periinfarct alterations to support diagnosis of stroke [5]. Use of 7T MRI is further motivated by the potential clinical benefits of using the ultra-high field MRI in presurgical planning [6].

Whole brain segmentation is a core neuroimaging technique that is used in common image processing pipelines of human brain MRI [7]. It is widely used for the measurement of various brain regions, to delineate boundaries, and to estimate brain age [8]. The efficacy of many neuroimaging pipelines is contingent upon having accurate whole brain segmentation from structural MRI data [9]. Several tools have been proposed to estimate whole brain segmentation. Deep convolutional neural network (DCNN) segmentation methods have provided efficient and accurate segmentation results. SLANT [10] is

a well-established DCNN segmentation tool that is capable of segmenting human brain MRI into 132 anatomical regions. SLANT has shown superior performance in validation methods since it has been trained with data from over 60 sites that include over 5000 MRI scans. Assessments of SLANT show that it achieves less than 0.05 coefficient of variation (CV) for intra-protocol 3T experiments [11]. However, SLANT has not been able to generate reliable whole brain segmentations on structural 7T MRI owing to the contrast variations in the image caused by RF inhomogeneity in 7T MRI [12]. The severity of the difference between segmentations generated by SLANT for 3T and 7T MRI scans can be clearly seen in Figure 1.

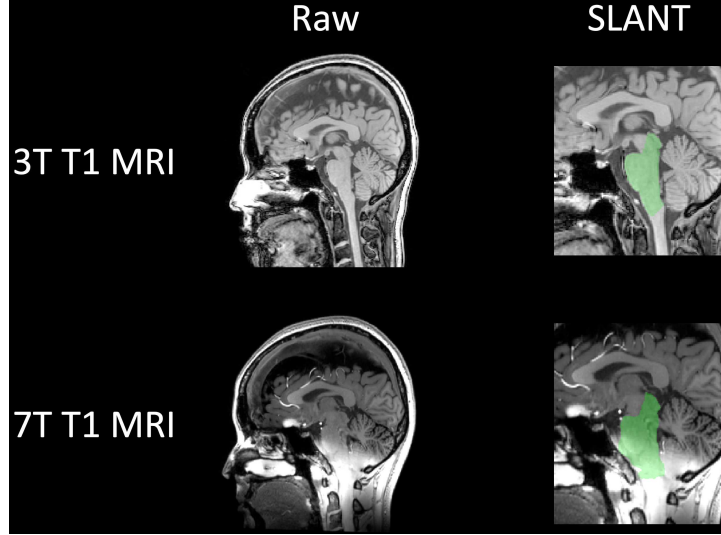


Figure 1. The human brain on the top left was acquired on a 3T MRI scanner. The image on the bottom left is that of the same participant scanned on a 7T MRI scanner. The brain stem segmentation has been overlaid as a black to green color map. The brain stem segmented by SLANT on the 3T scan was of volume 18,384 mm³. The volume of brain stem segmented by SLANT on the 7T scan was of volume 24,307.75 mm³.

As such, we propose to begin addressing this problem by finetuning [13] SLANT by adapting 3T MRI whole brain segmentation for 7T MRI.

2. METHODS

3D T1 scans were acquired for 15 subjects on both a 3T (3T Philips Achieva dStream scanner) and a 7T (7T Philips Achieva scanner) MRI scanner [14]. 14 out of the 15 subjects from the 7T MRI cohort were acquired at 0.7mm isotropic resolution and 1 subject at 0.5mm isotropic resolution. All the 3T T1 scans were acquired at a 1mm isotropic resolution. The 3T T1 scans were segmented using SLANT into 132 anatomical brain regions as defined by the BrainCOLOR protocol [15].

Our study requires the use of labels from the 3T SLANT results as the ground truth for finetuning SLANT on the 7T MRI scans. To achieve this, we use a 2-stage affine registration approach to have the 3T SLANT labels registered to their respective 7T MRI human brains. First, the transformation matrix for affine registering the 3T T1 scans to 7T T1 scans were computed by a 6 DOF (degree-of-freedom) rigid registration model using ANTs [16]. Second, the SLANT brain segmentation results from the 3T T1 were then registered to the 7T MRI human brain with the transformation matrix using multi-label interpolation [17]. The 7T MRI scans were preprocessed with the help of the containerized implementation of SLANT. The preprocessing steps for the SLANT pipeline include affine registration to MNI 305 atlas, N4 bias correction, and intensity normalization. With the transformation matrix used for registering the T1 scan to the MNI 305 atlas, the 3T T1 SLANT labels were registered to the 7T MRI scan in the MNI 305 space.

The 7T MRI T1 scans were tiled into 27 ($3 \times 3 \times 3$) overlapping subspaces, each with dimensions 96×128×88 voxels (as in SLANT). The ground truth labels were also tiled in a similar fashion for the training and validation cohort. We used a $k = 5$ fold cross validated modeling approach to finetune SLANT [18]. Each fold used 12 subjects for training and 3 for validation. Thus, we evaluated performance of each fold on 3 unique subjects that were not included in the training cohort in that fold. Each piece from SLANT's 27 models is a 3D U-Net that has been modified to accommodate the 132 different labels. The models were finetuned with a batch size of 1, an input resolution of 96×128×88 voxels, 1 input channel, 133

output channels, and the Adam optimizer with a learning rate = 0.0001. We finetuned SLANT for 30 epochs for each of the 5 folds.

The training was run on a NVIDIA Quadro RTX 5000 (16GB RAM) and a NVIDIA GeForce RTX 2080 Ti (11GB RAM) for all 5 folds. The training for each fold with 30 epochs and 27 independent models took ~8 training hours. With parallel training, the overall task was completed in ~24 hours. The inference run on the NVIDIA Quadro RTX 5000 finished in ~2 minutes per 7T T1 scan. The results of inference from the 5 fold model were postprocessed using the postprocessing pipeline from the containerized SLANT.

The volumes of individual labels and Dice similarity coefficients were used for comparing results before and after finetuning. The Dice similarity coefficients were computed with using the segmentation results on 3T MRI as the ground truth. The statistical significance of the mean Dice similarity coefficients before and after finetuning were evaluated using Wilcoxon signed rank test [19].

3. RESULTS

With a paired dataset of 15 participants scanned on both 3T and 7T MRI machines, our goal was to adapt the SLANT whole brain segmentation from 3T MRI onto 7T MRI. The SLANT segmentation results from 3T MRI were taken as the ground truth and we took the transfer learning approach to finetune SLANT for use on 7T MRI. With the finetuned SLANT model, we observed that the 3T-7T MRI inter-scanner mean relative difference dropped to ~8.43% from the earlier ~15.13% mean relative difference. The quantitative results from Figure 2 show improvements in the Dice similarity coefficients across 14 of the 15 subjects. The subject, ST7007, was acquired with 0.5 mm isotropic resolution where the contrast inhomogeneity was severe. The mean of the overall Dice similarity coefficient of SLANT results for all subjects before finetuning was at 0.79. After finetuning, the overall mean Dice similarity coefficient improved to 0.83. There is an increase in the median Dice scores of the 132 labels segmented by SLANT from 0.82 to 0.85. The Wilcoxon signed rank test returned a p-value less than 0.01. Although this increase in Dice similarity coefficient does not give us the sufficient accuracy required from whole brain segmentation, it goes on to show that we can expect improvement in segmentation by finetuning SLANT with more data.

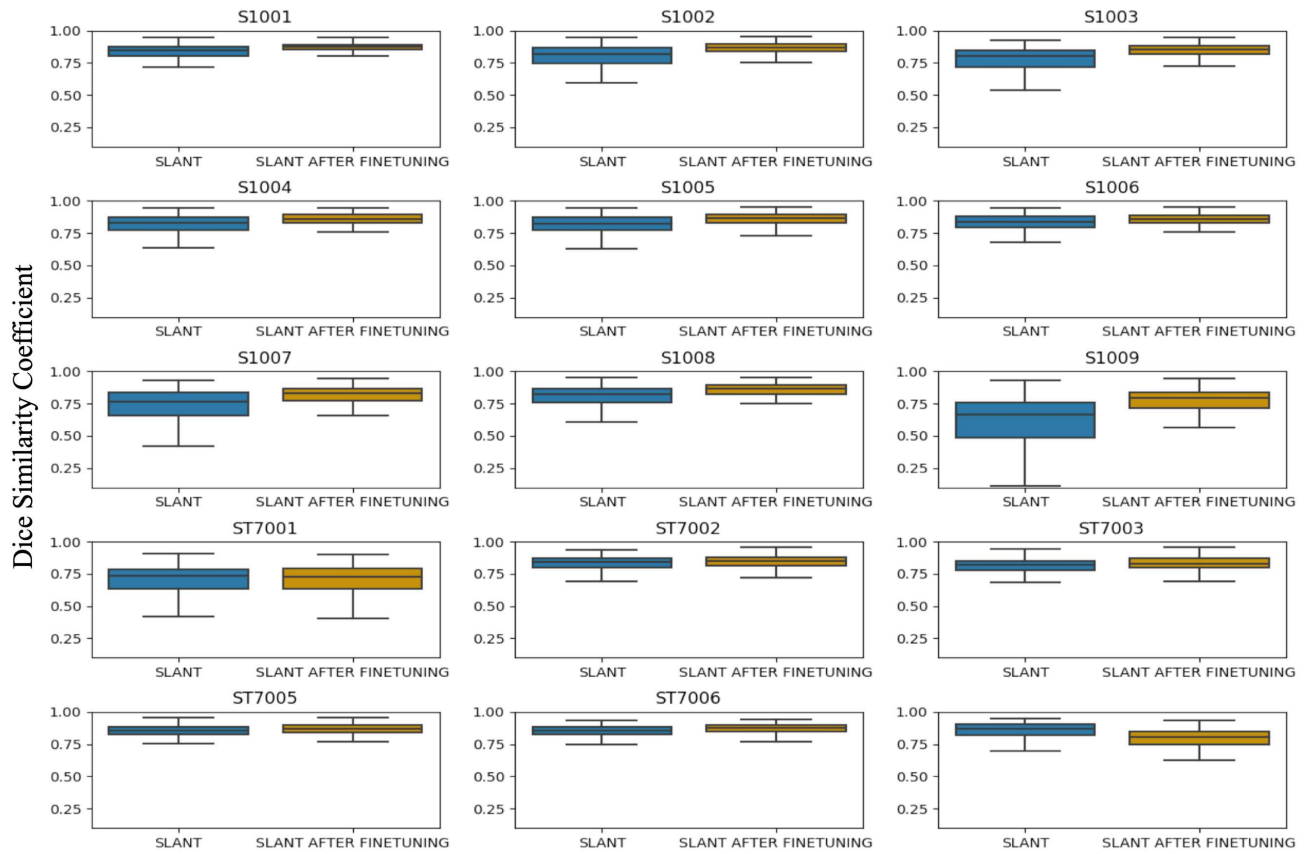


Figure 2. Dice similarity coefficients between 3T T1 and 7T T1 segmentations using raw SLANT and the after finetuning results from cross validation of the 15 subjects were showed as boxplots. Each box shows performance in a fold where the subject was not included in the training set. The overall mean of the Dice similarity coefficients across all subjects show an improvement after finetuning.

The Dice similarity coefficient of each ROI (region of interest) was acquired for every subject before and after finetuning SLANT. The mean of the Dice similarity coefficients for each ROI across subjects has been plotted as a grouped bar plot in Figure 3. 127 out of the 132 labels have an improved mean Dice similarity coefficient after finetuning. For the 5 labels where finetuning did not result in an improvement, the mean Dice similarity coefficients decreased from 0.79 to 0.78.

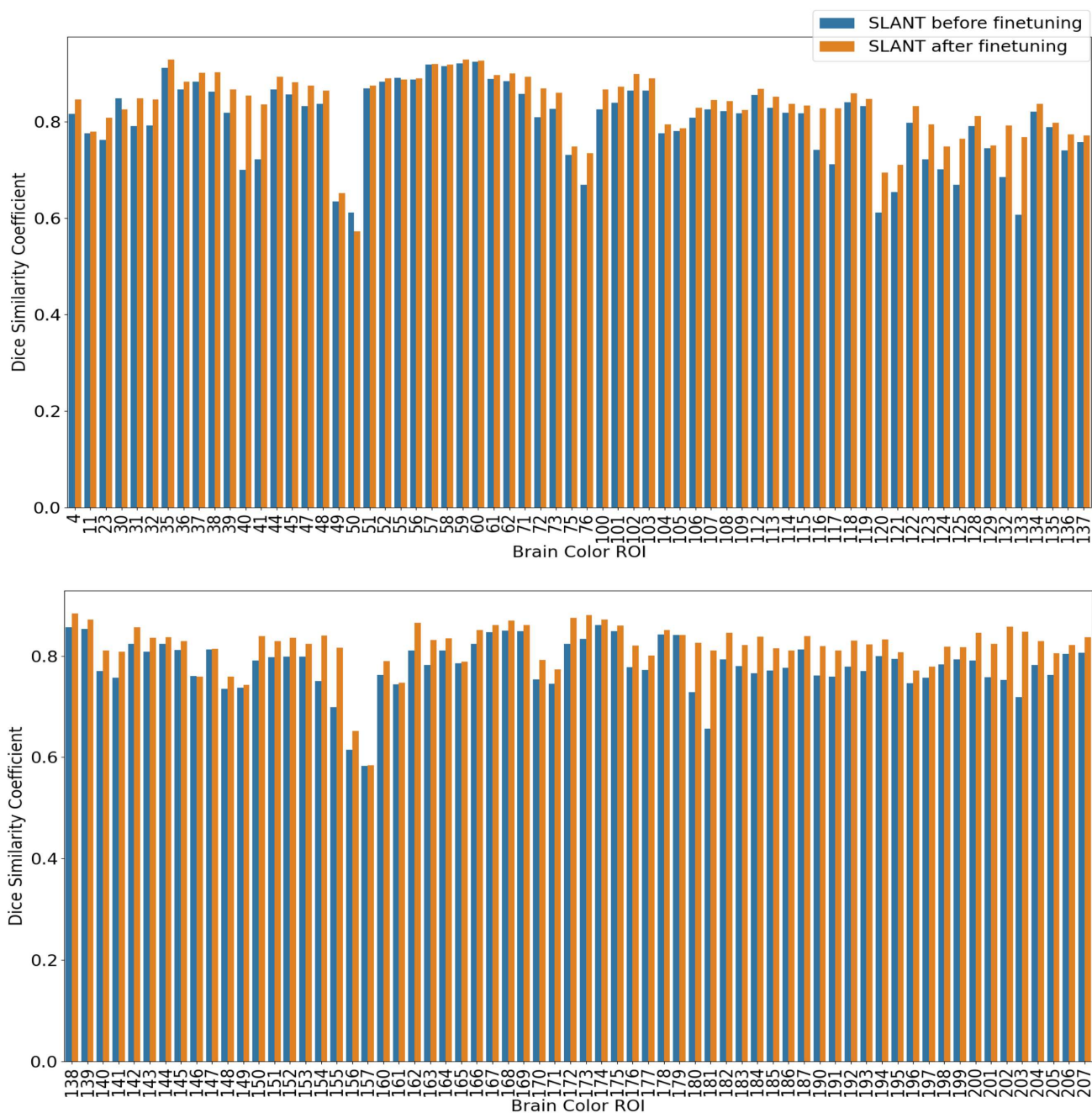


Figure 3. Mean Dice similarity coefficients of each ROI from the segmentations before and after finetuning were showed as grouped bar plots. The labels along x-axis are the label numbers corresponding to the regions in the BrainCOLOR protocol.

The qualitative results from segmentations are shown in Figure 4. In some regions such as the brain stem(label 35) and left calcarine cortex(label 109), we observed that the boundaries of segmentation have slightly improved. We observed some labels where the segmentation results from using SLANT and the finetuned model on 7T MRI showed little difference in the Dice scores. The mean Dice similarity coefficient of regions such as the left and right putamen(labels 57 and 58 from Figure 2) showed very small improvement(0.918 to 0.919). From Figure 3, we saw that 5 of the label volumes had mean Dice similarity coefficient with SLANT after finetuning perform slightly worse than before. The labels left accumbens area (label 30), left inferior lateral ventricle (label 50), right pallidum (label 55), right medial orbital gyrus (label 146) and left post orbital gyrus (label 176) showed worse mean Dice similarity coefficient after finetuning.

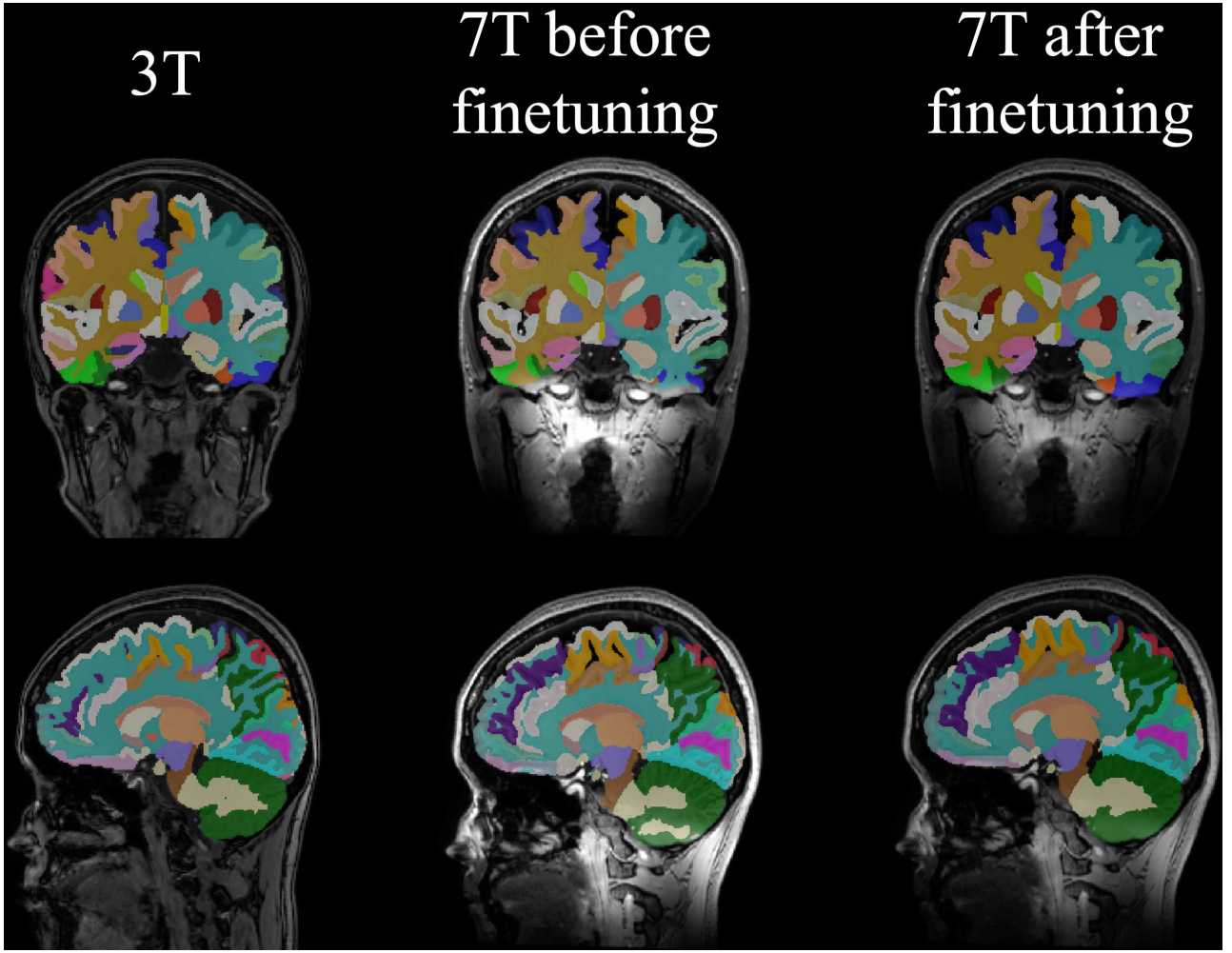


Figure 4. The sagittal and coronal slices with overlays of segmentations from SLANT on 3T MRI, SLANT on 7T MRI and finetuned SLANT on 7T MRI are shown. Some of the labels such as the left cerebellum white matter show clear improvement in the segmentation boundaries in the sagittal slices.

4. DISCUSSION

The accuracy of whole brain segmentation on a structural 7T MRI strongly influences the reproducibility and efficacy of the neuroimaging pipelines which can positively exploit the many benefits of the ultra-high field 7T MRI. In our previous work, we have used surfaces extracted from a 3T MRI to inform an automatic mapping of cortical layers on a 7T MRI scan [20]. One of the major limiting factors in that work was the requirement for paired data from a 3T and 7T MRI. Surface extraction on the ultra-high resolution scan was unattainable since the whole brain segmentation results were unreliable. The improvement in accuracy by finetuning SLANT is a promising step since accurate extractions of brain regions from structural 7T MRI can elevate several such research tools. Although our current results from the finetuned SLANT model do not reproduce the level of reliability with structural 3T scans, the trend of improving similarity to segmentation on 3T MRI is a promising start that can be further improved with more data.

SLANT was originally trained with 5111 multi-site images and finetuned with 35 images. Therefore, our study was designed such that we made the most out of the data by using 15 paired scans for training a 5 fold cross validated model. We did not test our results on external validation datasets. With more data acquisitions, we would be able to better analyze the degree of improvement that the finetuned model brings across all 132 labels. We only used the sensitivity parameters used by SLANT and have not tested other parameters such as different learning rates or epochs in our study.

ACKNOWLEDGEMENTS

Our work was supported by the National Institutes of Health (NIH) under award numbers U54 HD083211, P50HD103537, R37 HD095519 R01 NS095291 and NSF 1840896. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. This work has not been submitted for publication or presentation elsewhere.

REFERENCES

- [1] M. Cosottini and L. Roccatagliata, "Neuroimaging at 7 T: are we ready for clinical transition?," *European Radiology Experimental*, vol. 5, no. 1, p. 37, 2021, doi: 10.1186/s41747-021-00234-0.
- [2] S. O. Dumoulin, A. Fracasso, W. van der Zwaag, J. C. W. Siero, and N. Petridou, "Ultra-high field MRI: Advancing systems neuroscience towards mesoscopic human brain function," *Neuroimage*, vol. 168, pp. 345–357, 2018, doi: <https://doi.org/10.1016/j.neuroimage.2017.01.028>.
- [3] R. W. McGugin, A. T. Newton, B. Tamber-Rosenau, A. Tomarken, and I. Gauthier, "Thickness of deep layers in the fusiform face area predicts face recognition," *bioRxiv*, 2019, doi: 10.1101/788216.
- [4] M. Inglese, L. Fleysher, N. Oesingmann, and M. Petracca, "Clinical applications of ultra-high field magnetic resonance imaging in multiple sclerosis," *Expert Review of Neurotherapeutics*, vol. 18, no. 3, pp. 221–230, Mar. 2018, doi: 10.1080/14737175.2018.1433033.
- [5] V. I. Madai *et al.*, "Ultrahigh-Field MRI in Human Ischemic Stroke – a 7 Tesla Study," *PLOS ONE*, vol. 7, no. 5, pp. e37631–, May 2012, [Online]. Available: <https://doi.org/10.1371/journal.pone.0037631>
- [6] S. Trattnig *et al.*, "Key clinical benefits of neuroimaging at 7T," *Neuroimage*, vol. 168, pp. 477–489, 2018, doi: <https://doi.org/10.1016/j.neuroimage.2016.11.031>.
- [7] Z. Liu *et al.*, "Deep learning based brain tumor segmentation: a survey," *Complex & Intelligent Systems*, 2022, doi: 10.1007/s40747-022-00815-5.
- [8] H. Jiang *et al.*, "Predicting Brain Age of Healthy Adults Based on Structural MRI Parcellation Using Convolutional Neural Networks," *Frontiers in Neurology*, vol. 10, 2020, doi: 10.3389/fneur.2019.01346.
- [9] I. Despotović, B. Goossens, and W. Philips, "MRI Segmentation of the Human Brain: Challenges, Methods, and Applications," *Computational and Mathematical Methods in Medicine*, vol. 2015, p. 450341, 2015, doi: 10.1155/2015/450341.
- [10] Y. Huo *et al.*, *Spatially Localized Atlas Network Tiles Enables 3D Whole Brain Segmentation from Limited Data*. 2018.
- [11] Y. Xiong, Y. Huo, J. Wang, L. T. Davis, M. McHugo, and B. A. Landman, "Reproducibility evaluation of SLANT whole brain segmentation across clinical magnetic resonance imaging protocols," in *Proc.SPIE*, Mar. 2019, vol. 10949. doi: 10.1117/12.2512561.
- [12] J. Stockmann and L. Wald, "In vivo B0 field shimming methods for MRI at 7T," *Neuroimage*, vol. 168, Jun. 2017, doi: 10.1016/j.neuroimage.2017.06.013.
- [13] F. Zhuang *et al.*, "A Comprehensive Survey on Transfer Learning," *Proceedings of the IEEE*, vol. PP, pp. 1–34, Jul. 2020, doi: 10.1109/JPROC.2020.3004555.

- [14] Y. Liu, P.-F. D’Haese, A. T. Newton, and B. M. Dawant, “Generation of human thalamus atlases from 7 T data and application to intrathalamic nuclei segmentation in clinical 3 T T1-weighted images,” *Magnetic Resonance Imaging*, vol. 65, pp. 114–128, 2020, doi: <https://doi.org/10.1016/j.mri.2019.09.004>.
- [15] A. Klein *et al.*, “Open labels: online feedback for a public resource of manually labeled brain images.” [Online]. Available: <http://www.braincolor.org/openlabels/roygbiv>
- [16] B. B. Avants, N. J. Tustison, G. Song, P. A. Cook, A. Klein, and J. C. Gee, “A reproducible evaluation of ANTs similarity metric performance in brain image registration,” *Neuroimage*, vol. 54, no. 3, pp. 2033–2044, 2011, doi: <https://doi.org/10.1016/j.neuroimage.2010.09.025>.
- [17] A. C. Bovik, “Basic Gray Level Image Processing,” in *The Essential Guide to Image Processing*, Elsevier Inc., 2009, pp. 43–68. doi: 10.1016/B978-0-12-374457-9.00003-2.
- [18] D. Opitz and R. Maclin, “Popular Ensemble Methods: An Empirical Study,” vol. 11, Dec. 1999, doi: 10.1613/jair.614.
- [19] F. Wilcoxon, “Individual Comparisons by Ranking Methods,” *Biometrics Bulletin*, vol. 1, no. 6, pp. 80–83, 1945, doi: 10.2307/3001968.
- [20] K. Ramadass *et al.*, “Ultra-high-resolution mapping of cortical layers 3T-guided 7T MRI,” in *Proc.SPIE*, Apr. 2022, vol. 12032. doi: 10.1117/12.2611857.