**Supporting Figure S1:** Effect of initialization on final accuracy and convergence curves. validation accuracy (Dice) and training loss for a network with random initialization vs. initialization from a network trained on a 3T dataset is shown. The segmentation results indicate a statistically significant improvement over two nuclei (VA and Hb) when initialized from the 3T network.

**Supporting Table S1:** Image acquisition parameters for WMn- and CSFn-MPRAGE for 7T and 3T data used in this study

|  |  |  |  |
| --- | --- | --- | --- |
|  | **WMn-MPRAGE** | | **CSFn-MPRAGE** |
| **Scanner** | 3T (GE) | 7T (GE) | 7T (GE) |
| **TR/TS (ms)** | 10/4500 | 10/6000 | 7.2/3000 |
| **TI (ms)** | 500 | 680 | 1200 |
| **Flip (deg)** | 9 | 4 | 6 |
| **Matrix** | 180 × 220 × 180 | 180 × 220 × 180 | 180 × 220 × 180 |
| **Acq. resolution (mm)** | 1x1x1 | 1x1x1 | 1x1x1 |
| **Rec. resolution (mm)** | 0.7x0.5x0.7 | 0.7x0.5x0.7 | 0.7x0.5x0.7 |
| **Parallel imaging** | None | 1.5 × 1.5 | 3x1 |
| **Coil** | 8-channel GE | 32-channel Nova | 32-channel Nova |

**Supporting Table S2:** Effect of loss function on Dice and VSI. Average Dice and VSI shown for 11 subjects.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Dice** | | **VSI** | |
| **Nuclei** | Dice Loss | BCE Loss | Dice Loss | BCE Loss |
| **Thalamus** | 0.92† | 0.91 | 0.98† | 0.96 |
| **Pul** | 0.86 | 0.85 | 0.97 | 0.96 |
| **VLP** | 0.78 | 0.79 | 0.95 | 0.95 |
| **MD-Pf** | 0.85 | 0.85 | 0.95 | 0.94 |
| **VPl** | 0.64† | 0.61 | 0.92 | 0.89 |
| **VA** | 0.70 | 0.69 | 0.92 | 0.93 |
| **AV** | 0.76 | 0.74 | 0.87 | 0.85 |
| **VLa** | 0.69 | 0.67 | 0.92 | 0.89 |
| **CM** | 0.70† | 0.65 | 0.93† | 0.86 |
| **LGN** | 0.70† | 0.66 | 0.90 | 0.85 |
| **MGN** | 0.70† | 0.64 | 0.89 | 0.81 |
| **MTT** | 0.68 | 0.63 | 0.88 | 0.81 |
| **Hb** | 0.77 | 0.75 | 0.89 | 0.88 |

† p < 0.05 BCE vs Dice loss function