Brain structure segmentation utilizing 3D convolutional neural networks.

Panagiotis Korfiatis PhD, Christopher G. Schwarz PhD, Jeffrey L. Gunter PhD, Matthew L. Senjem MS, Bradley J. Erickson MD, PhD

Mayo Clinic
Disclosures

Nothing to disclose
Investigate deep neural architectures for deep brain structure segmentation from MRI scans
Deep Brain Structures

- Thalamus (L/R)
- Caudate (L/R)
- Putamen (L/R)
- Pallidum (L/R)
Deep Brain Structure Role

Deep brain regions:
- cognitive
- affective
- social functions in humans

Quantification:
- diagnosis and monitoring of diseases in observational studies and clinical trials:
  - Diagnosis of schizophrenia
  - Autism
  - Parkinsonism
  - Multiple sclerosis
State of the art

- Early approaches:
  - Atlas based approaches
    - FIRST and PICL reporting accuracies in terms of dice coefficient higher than 0.820

- Deep learning:
  - Milletari et al. Hough CNN, Dice ~ 0.85
  - Wachinger et al. CNN patch based, Dice ~ 0.92
  - Dolz et al. dice coefficient range between 0.86 and 0.92 (skull stripped data)
Our approach

Deep learning: UNET variant
Architecture

- Modified UNET architecture
  - 3D
  - Multiple level output
    - Reduces coarseness of segmentation
  - Residual Connections and context modules
    - two 3x3x3 convolutional layers and a dropout layer (stride 2)
  - Localization modules
    - 3x3x3 convolution followed by 1x1x1
  - Strided convolutions
    - No maxpooling
Publicly available dataset (Alzheimer’s Disease Neuroimaging Initiative)

- 14350 scans
  - T1 weighted scans
  - Multiple scans per patient
  - Multiple manufactures

Gold Standard

- A pipeline based on SPM12, the MCALT brain template, and ANTs*. 
- Reviewed by image analysts

Preprocessing

- **Intensity standardization**
  - subtracting the mean value (excluding zero values) and dividing by the standard deviation across all brain voxels
    - Background pixels were excluded.
    - [-4, 4] clipping
    - Rescale [0, 1]

- **96x96x96**
  - Minimum bounding box was cropped around the head.
Training

- **Dataset**
  - 12000 training/validation
  - 2350 testing

- **UNET**
  - Adaptive dice loss coefficient
  - 300 epochs
    - Early stopping was implemented (patients =
  - Batch size: 8
  - Optimizer Adam (lr=0.01)
  - Instance Normalization
  - leaky ReLU
  - Softmax
  - 8 classes (left/right)

- **Implementation**
  - Keras (tensorflow backend)
  - Horovod
  - DGX-1
## Results (1/3)

<table>
<thead>
<tr>
<th>Structure</th>
<th>Dice</th>
<th>Average Hausdorff distance (pixels)</th>
<th>False positive error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thalamus (L)</td>
<td>0.934 (0.041)</td>
<td>0.093 (0.069)</td>
<td>0.072 (0.039)</td>
</tr>
<tr>
<td>Thalamus (R)</td>
<td>0.933 (0.019)</td>
<td>0.100 (0.032)</td>
<td>0.069 (0.018)</td>
</tr>
<tr>
<td>Caudate (L)</td>
<td>0.937 (0.058)</td>
<td>0.094 (0.071)</td>
<td>0.063 (0.039)</td>
</tr>
<tr>
<td>Caudete (R)</td>
<td>0.938 (0.042)</td>
<td>0.080 (0.046)</td>
<td>0.063 (0.032)</td>
</tr>
<tr>
<td>Putamen (L)</td>
<td>0.916 (0.054)</td>
<td>0.083 (0.076)</td>
<td>0.087 (0.047)</td>
</tr>
<tr>
<td>Putamen (R)</td>
<td>0.923 (0.048)</td>
<td>0.011 (0.098)</td>
<td>0.080 (0.047)</td>
</tr>
<tr>
<td>Pallidum (L)</td>
<td>0.954 (0.034)</td>
<td>0.049 (0.117)</td>
<td>0.045 (0.019)</td>
</tr>
<tr>
<td>Pallidum (R)</td>
<td>0.954 (0.042)</td>
<td>0.064 (0.055)</td>
<td>0.044 (0.042)</td>
</tr>
</tbody>
</table>

**Table 1:** Results in term of Dice coefficient, average Hausdorff distance and false positive error rate on a test set consisting of 2350 exams. The results are reported in term of mean and standard deviation. No statistical significant difference was observed between left and right results.
Results (2/3)

<table>
<thead>
<tr>
<th>Case 1</th>
<th>Ground Truth</th>
<th>Output</th>
</tr>
</thead>
<tbody>
<tr>
<td>(representative slice)</td>
<td><img src="image1" alt="Image" /></td>
<td><img src="image2" alt="Image" /></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Case 2</th>
<th>Ground Truth</th>
<th>Output</th>
</tr>
</thead>
<tbody>
<tr>
<td>(representative slice)</td>
<td><img src="image3" alt="Image" /></td>
<td><img src="image4" alt="Image" /></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Case 3</th>
<th>Ground Truth</th>
<th>Output</th>
</tr>
</thead>
<tbody>
<tr>
<td>(representative slice)</td>
<td><img src="image5" alt="Image" /></td>
<td><img src="image6" alt="Image" /></td>
</tr>
</tbody>
</table>
Results (3/3)

<table>
<thead>
<tr>
<th></th>
<th>Thalamus (L)</th>
<th>Thalamus (R)</th>
<th>Caudate (L)</th>
<th>Caudate (R)</th>
<th>Putamen (L)</th>
<th>Putamen (R)</th>
<th>Pallidum (L)</th>
<th>Pallidum (R)</th>
</tr>
</thead>
<tbody>
<tr>
<td>100 cases</td>
<td>0.682</td>
<td>0.640</td>
<td>0.664</td>
<td>0.674</td>
<td>0.631</td>
<td>0.624</td>
<td>0.568</td>
<td>0.593</td>
</tr>
<tr>
<td>500 cases</td>
<td>0.828</td>
<td>0.830</td>
<td>0.849</td>
<td>0.850</td>
<td>0.796</td>
<td>0.801</td>
<td>0.866</td>
<td>0.865</td>
</tr>
<tr>
<td>1000 cases</td>
<td>0.881</td>
<td>0.881</td>
<td>0.863</td>
<td>0.8871</td>
<td>0.859</td>
<td>0.871</td>
<td>0.861</td>
<td>0.918</td>
</tr>
<tr>
<td>5000 cases</td>
<td>0.902</td>
<td>0.923</td>
<td>0.897</td>
<td>0.890</td>
<td>0.893</td>
<td>0.887</td>
<td>0.901</td>
<td>0.923</td>
</tr>
</tbody>
</table>

UNET size changes depending on number of cases available.
Future Steps

Average Dice for 127 cortical regions ~0.92±0.04
Discussion

- High accuracy
- Multi center data
- Inference Time
- No need for skull stripping
Questions ?