Segmentation of Subcortical Brain Structures Using Convolutional Neural Network

Bryan Li[a], Judy Wang[a], Jason Zhang[a], Will Ye[a]
The Johns Hopkins University, Department of Computer Science [a]

***Abstract* – This project attempts to perform 3D subcortical brain segmentation of MRI scans through the use of a convolutional neural network (CNN). Accurate brain segmentation is crucial for the study of certain disorders due to brain trauma, such as anterograde amnesia, sectoranopia, and hemihypesthesia. We utilize the CNN to segment different brain structures, then create a mask based on each segment to fix a bounding box around each brain segment. We trained and executed our CNN on an Internet Brain Segmentation Repository dataset and obtained results that show successful segmentation of subcortical brain sections.**

I. Introduction

Currently, Convolutional Neural Networks (CNN) have become recognized as a new efficient method in classification. CNNs have proven to be more effective at image classification than humans, which has established the use of deep learning in brain imaging. Thus, CNN has rapidly become the choice in improving brain diagnosis, treatment and characterization, which benefits from accurate segmentation. [1] For example, the ability to segment and visualize the thalamus would allow us to determine whether trauma or disfigurement of certain areas of the thalamus would indicate the possibility of developing certain disorders or syndromes. One example of spatial specific injury that leads to disorder is trauma to the Anterior thalamic nucleus, which would lead to the development of anterograde amnesia. [5] While current methods require us to segment and analyze by hand, our goal is to optimize this process for faster and more accurate diagnosis.

II. Methods

A. Theory

CNNs consists of three parts, which include input neurons, output, and convolution layers. Because our input data are 3 dimensional scans, our input neurons will be our image voxels. The convolutional layers' task is to convolve our input scan with a 3x3x3 kernel which will produce a feature map projection of feature detections. Each consecutive layer detects and accumulates more complex features as the network gets deeper, then a full convolutional layer aggregates all the feature maps to label the object in question. [2] The network that we adopt is similar to the one proposed by Dolz, J. et al. [3], which utilizes three sets of three convolutional layers followed by three fully connected convolutional layers. For each consecutive layer, we implemented the Parametric Rectified Linear Unit (PReLU) as a nonlinearity defined as such:

$$f(x_i) = max(0, x_i) + a_i * min(0, x_i) \qquad (1)$$

Where f(xi) is our output, xi is our input and ai is our coefficient that improves the adaptiveness of our PReLU for accuracy. We also implement a softmax function for voxel classification as such:

$$\sigma(\mathbf{z})_\mathbf{j} = \frac{\mathbf{e}^{\mathbf{x_j}}}{\sum_{\mathbf{k=1}}^{\mathbf{K}} \mathbf{e}^{\mathbf{x_j}}} \qquad (2)$$

in which K represents the possible number of classes. Our algorithm utilizes seven classes: left and right thalamus, left and right putamen, left and right caudate, and other. We will train our network with training data and will utilize cross entropy as the network's loss function and back propagation.

B. Datasets

We used brain MRI images from the open access Internet Brain Segmentation Repository (IBSR) dataset to train and develop our network. The dataset consists of 18 T1-weighted volumetric brain scans acquired from healthy patients of ages 38.0 +/- 22.4 (14 males and 4 females), which have been positionally normalized into the Talairach orientation and pre-processed for bias field correction [2]. Manual segmentations were provided for 43 individual structures of principle grey and white matter, including the six subcortical structures we focused on for the scope of this project, which we used as ground-truth for the training and testing phases. A total of 10 scan sets were used for training and validation, while the remaining 8 MRIs made up our testing set. We normalized all image resolution to be 0.9375 x 1.5 x 0.9375 mm, with a consistent matrix size of 256 x 128 x 256. The dataset was provided by the Center for Morphometric Analysis at Massachusetts General Hospital.

C. Pre-Processing

Using the FMRIB Software Library (FSL) v6.0, we performed file type conversion from IMG to NIFTI format on the original IBSR data. Skull-stripping was used to remove the skull and other non-brain tissue from the images before segmentation can be performed. For the CNN we made, we had to crop the image into a fixed size of (160, 200, 40) and tried to extract as much brain data as possible during our compression and ignore as much background noise as possible. We also attempt a zero mean intensity normalization for images to improve prediction performance.

D. Convolutional Neural Network

The architecture of the CNN we utilized is based on the work of Dolz, J. et al. [3], which is shown in Figure 1.

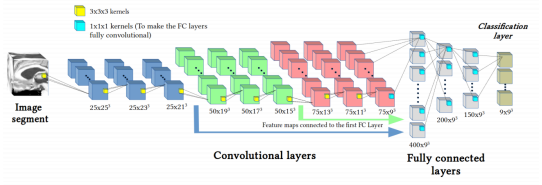


Fig. 1 Convolutional Neural Net proposed by J. Dolz.

The input corresponds to our input voxels, which is fed into the convolution neural network to generate a classification of each voxel. The first three layers utilize 3D convolutional layers that utilize a kernel with the size of 3x3x3, filter size of 25, and a stride value of 1. These detect and map out general details and certain edges. The second set is similar but with a filter size of 50. The

previous feature map is used to produce a new feature map output that detects more complex features. We perform this again with a third set where the new filter size is 75. Shown in Figure 1, feature maps are generated by the third, sixth, and ninth layers, which will be fed into the fully convolutional layers with filter sizes of 400, 200 and 150 respectively. [3] This uses a pyramid architecture, enabling the network to make complex nonlinear combinations of each subsequent feature map. Each convolutional layer is fitted with our PReLU activation function (1) to ensure better adaptability than the conventional ReLU function, as PReLU allows the network to change the 'a' parameter, giving us greater accuracy. In this case, PReLU outperforms ReLU in terms of accurate classification with only slightly more computational complexity. Once we obtain the output of the fully convolutional layers, we pass the output through a softmax activation function (2), enabling us to compute voxel classification likelihood. We chose ten images to feed into the neural network with a pre-labeled segmentation mask. The CNN compares each prediction with the labeled mask to compute a loss function. We chose cross entropy loss as our loss function. We then use back propagation to update kernel value weights with ADAM, our chosen optimizer for traditional CNNs. Once our CNN is fully trained, we use the same weights to make predictions with new input images.

E. Alternative CNN

We visually compared the segmentation performance of our CNN with an existing F-CNN, developed by Shakeri, M. et al. [4] and adapted for our input dataset. The architecture, summarized in Table I, is built upon the MatConvNet toolbox in MATLAB, and consists of five pairs of convolutional/max pooling layers. Inspired by the DeepLab network [6], the input is subsampled for the first two layers, while holes are introduced for the remaining layers to increase the receptive field of filters.

TABLE I
OVERVIEW OF LAYERS AND PARAMETERS

Block	conv kernel	# filters	hole stride	pool kernel	pool stride	dropout
1	7×7	64	1	3×3	2	no
2	5×5	128	1	3×3	2	no
3	3×3	256	2	3×3	1	yes
4	3×3	512	2	3×3	1	yes
5	3×3	512	2	3×3	1	yes
6	4×4	1024	4	no pooling		yes
7	1×1	39	1	no pooling		no

III. RESULTS AND DISCUSSION

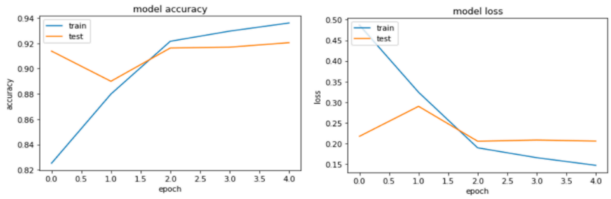
Both CNN's was trained on the first 10 IBSR patient data and tested on the remaining 8. We obtained clear and fluid segmentations shown in Figures 3 and 4. Our CNN used an ADAM optimization of 0.0001 for the highest accuracy as shown in table 2.

Figs. 2-4 (Left to Right) Processed IBSR scan, skull-stripped; Subcortical segmentations produced by our CNN; Subcortical segmentations of F-CNN, overlayed with manual segmentations at slice 182 of patient 01.

TABLE II
TRAINING PARAMETERS

ADAM Optimizer Parameter	Resulting Accuracy
0.1	40%
0.001	70%
0.0001	93%
0.00001	8%

We utilized 5 epochs because less epochs displayed less feature data from the input image while more epochs had negligible improvement. We were careful not to overfit the data, which results from training on the same data set repeatedly.



Figs. 4-5 Graphs of model accuracy and loss versus epoch number.

Additionally, we varied our CNN by changing the kernel size. Larger kernel size meant more required memory for weights and more computation time for back-propagation. Smaller kernel size meant a drop in accuracy. Thus, we chose a 3x3x3 kernel, which straddle the two extremes. Our CNN performed poorly when we experimented with training on all 43 individual structures. Our filter size provided limited feature detection, which can be an area for improvement. We may also improve network performance by increasing the number of convolutional layers to deepen the network, which would boost performance by learning more complex, non-linear behaviors. However, this approach will likely pose extensive training difficulty.

IV. CONCLUSION

In this paper, we had decided to experiment with a convolutional neural network to accurately segment subcortical structures in brain MRI image scans. Our tests indicate that our proposed method can be an effective and efficient approach at brain segmentation for diagnosis, however, we must understand that there are limitations: currently our model cannot detect more structures as it had trouble learning more than our seven proposed classes. Also, due to system and memory limitations, we couldn't get more accurate results. However, with better infrastructure, and a deeper neural network, segmentation can become more accurate and better optimized.

V. REFERENCES

[1] H. Choi and K. H. Jin, "Fast and robust segmentation of the striatum using deep convolutional neural networks," Journal of Neuroscience Methods., vol. 274, pp. 146-153, Dec. 2016.

[2] T. Rohlfing, "Image similarity and tissue overlaps as surrogates for image registration accuracy: Widely used but unreliable," IEEE Transactions on Medical Imaging, vol. 31, no. 2, pp. 153-163, 2012.

[3] J. Dolz, C. Desrosiers, and I. B. Ayed, "3D fully convolutional networks for subcortical segmentation in MRI: A large-scale study," NeuroImage, vol. 170, pp. 456-470, 2018.

[4] M. Shakeri, S. Tsogkas, E. Ferrante, S. Lippe, S. Kadoury, et al., "Sub-cortical brain structure segmentation using F-CNN's," ISBI 2016: International Symposium on Biomedical Imaging, Prague, Czech Rep., 2016.

[5] A. G. Roy et al, "QuickNAT: A Fully Convolutional Network for Quick and Accurate Segmentation of Neuroanatomy," NeuroImage, Nov, 2018. (In press).

[6] L. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A.L. Yuille, "Semantic image segmentation with deep convolutional nets and fully connected crfs," arXiv preprint arXiv:1412.7062, 2014.