The Application of Convolutional Neural Networks for Subcortical Structure Segmentation

Han Wang[1], Mahmoud Mostapha[1], Siyang Jing[1], Heather Hazlett[2], Michael Graves[2], Junpyo Hong[1], Martin Styner[1,2]

Departments of [1]Computer Science and [2]Psychiatry, The University of North Carolina at Chapel Hill, NC, USA

Abstract

Many image analysis tasks require the identification of structural abnormalities. Hence, the segmentation of medical images is an important step that enables further analysis and diagnosis of patients' conditions. In our study, we adapted a popular multi-resolution, a multi-modal convolutional neural network approach to the segmentation of subcortical brain structures from MRI of 12-24 months infants. In order to improve learning efficiency and reduce data storage, T1- and T2-weighted MRI images were preprocessed and cropped. In the segmentation step, a fully 3D U-Net based architecture was applied to generate multi-label segmentations. We postprocessed the segmentations via morphological operations to ensure simply connected objects. To counter the small amount of well-annotated data, we employed two strategies: a) Pretraining was established via transfer learning from a larger dataset of uncorrected automatic multi-atlas-based segmentation in order to provide a good initialization of parameters compared to training from scratch, b) final training included left-right flipped datasets assuming that appearance of left and right hemispheric structures are near exchangeable. The segmentation accuracy of the proposed method above is comparable to that of an advanced multi-modality, a multi-atlas method with several orders of magnitude speed-up. Further improvements to the deep learning network model as well as the generation of additional training datasets are currently in process to further improve the accuracy of the proposed approach.

Keyword

Convolutional Neural Networks, Neuroimaging, Subcortical Structure Segmentation

1. Introduction

Segmentation of the subcortical area from medical imaging is crucial in many neuroimaging studies, including those of most brain disorders, for example, ADHD (Hoogman et al., 2017), Autism (Cerliani et al., 2013), Schizophrenia (van Erp et al., 2016), and Parkinson (Geevarghese el al., 2015).

However, studies of segmentation of subcortical structures in the developing brain, such as in the brain of 12 to 24 months old children have not been at the same level as other age groups due to the lower signal to noise setting, a larger amount of motion artifacts and the presence of intensity inhomogeneities in the brain white matter due to immature myelination. This work focuses on the automation of segmentation of subcortical structures (hippocampus, amygdala, thalamus, putamen, caudate and globus pallidus) in that age range in order to boost neuroimaging studies in early postnatal brain development.

Deep convolutional networks have recently been widely used in the biomedical imaging field, and they have started to significantly outperform the best prior methods in various visual tasks, especially image classification. Here, we adopted a 3D convolutional neural network(CNN) called U-Net (Ronneberger et al., 2015) with modifications to its original downsampling and upsampling architecture to build better-fitting models for subcortical structure segmentation.

2. Methods

2.1 Data Acquisition

Our training data consists of T1- and T2-weighted MR images of 12-24 months infants (N=32) that are either at high or low risk for autism spectrum disorder (ASD) (Swanson et al., 2017). Roughly 1/5 of those at high risk later developed ASD. The training images are equally distributed from subjects with low risk, high risk with no ASD and high risk with ASD. Scans were performed during natural sleep using 3-T Siemens TIM Trio scanners as part of the NIH funded IBIS project. Figure 1 demonstrates the initial data after intensity inhomogeneity correction, ICBM template registration, and brain masking.

2.2 Data Processing for CNN

We employed a 10-fold analysis and thus split our dataset into training, validation, and inference (testing) in an 8:1:1. Data augmentation using the right-left-hemispheric flipped images was performed. Due to the similarity between the original images and the flipped images, we divided the datasets such that each pair of the original image and its corresponding flipped image can only be included in any one of the training, validation, and inference datasets. If one image would be used for training and its correspondingly flipped image would be used for either validation or testing, optimistically incorrect validation or testing scores could be obtained.

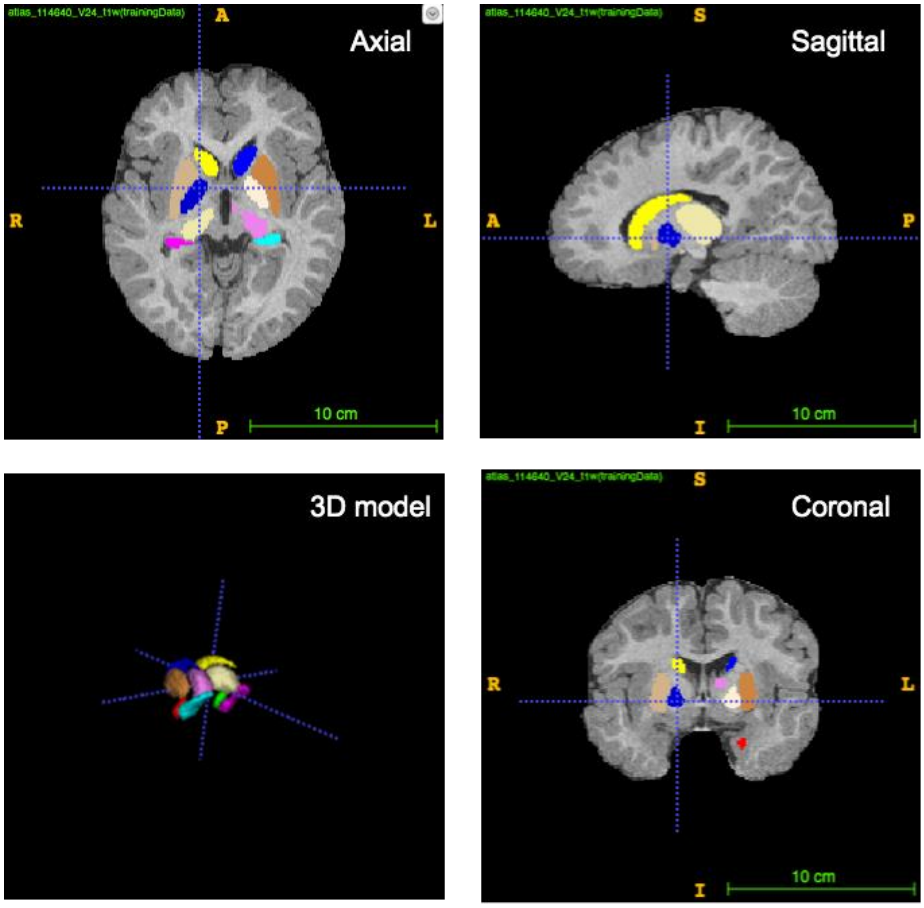

Figure 1. MRI scan images of 12-24 months old infants.

For deep 3D CNNs, image size is of particular issues as the size of most standard GPU graphics cards limited to 8GB. As storage space is limited on the GPU, a larger image goes along a smaller training batch size. To maximize batch size and provide efficient training, we preprocessed our data accordingly. The steps are further illustrated in figure 2. It is noteworthy that we adopted a "fully convolutional network" method as such methods have shown a significant advantage in applications with small training datasets.

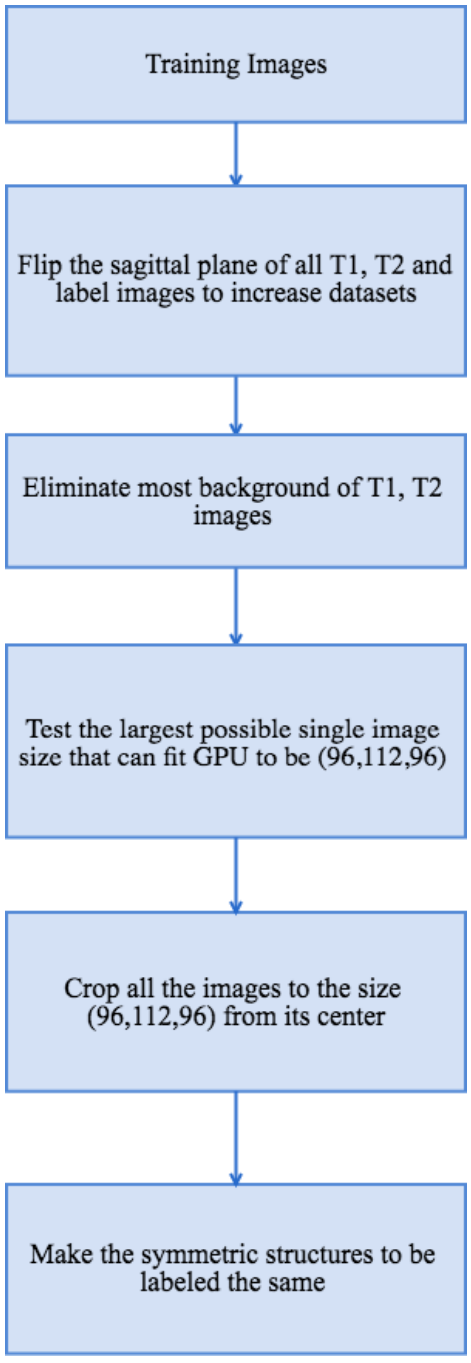

Figure 2. Flowchart of data preprocessing before inputting into the network.

Subcortical structures only cover a smaller part of the medial sections of the brain. As such we first cropped all datasets to a 10 percent enlarged version of the minimal size covering all subcortical structures in a larger database of automatic atlas-based segmentation. The corresponding cropped images are of 96 × 112 × 96 size (1x1x1 mm resolution). A sample T1-weighted MRI image of such size is shown in figure 3. We further employed a full padding mode for convolutional layers, as well as added dropout layers at each convolutional layer. Our adapted 3D U-Net network structure is shown in figure 4 below.

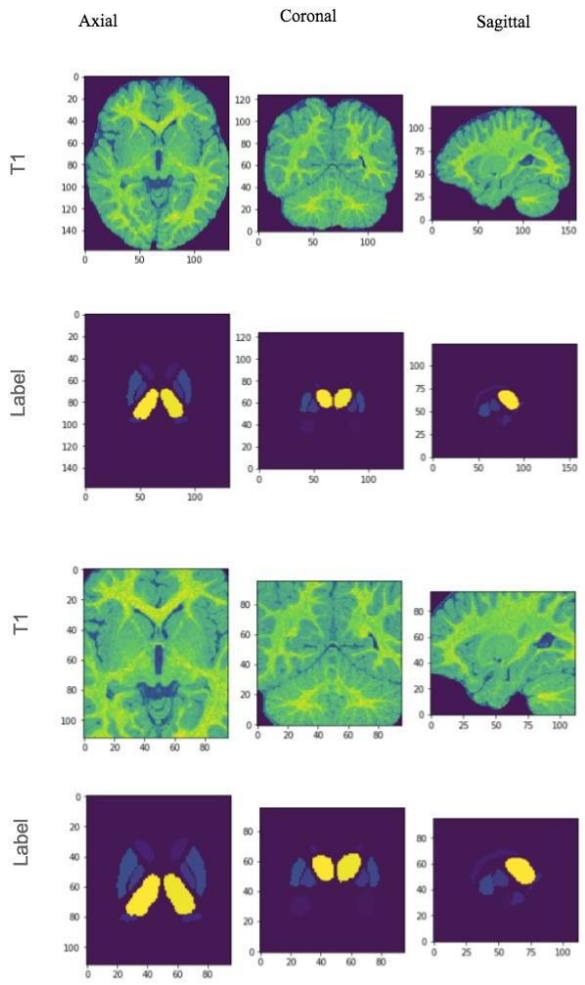

Figure 3. Comparison between background-removed and cropped T1-weighted MRI image.

It is difficult to train a deep neural network from scratch, especially with the limited amount of annotated data that is present in our setting. Previous studies demonstrated that pretraining on a large number of images with less accurate annotation could boost model accuracy. To best utilize the advantage of transfer learning, we chose to use images with the same subcortical structures segmentations from subjects within the same age range as our regular training dataset. Overall, we employed (N=860) multi-modality, multi-atlas based segmentations (Wang et al, 2014) in the pretraining step

Class imbalance is an issue with training segmentation of subcortical structures because they usually only occupy a small part of the brain. There is also a disparity between each label's volume. Experiments with training without considering class imbalance can lead to loss of the small labels due to their insignificant volumes. Adding weights can balance the size disparity between labels and background, as well as the disparity between labels per se. Therefore, we generated

weight maps for each training label image inversely weighting voxels of a given subcortical structure (weight = 1/Volume).

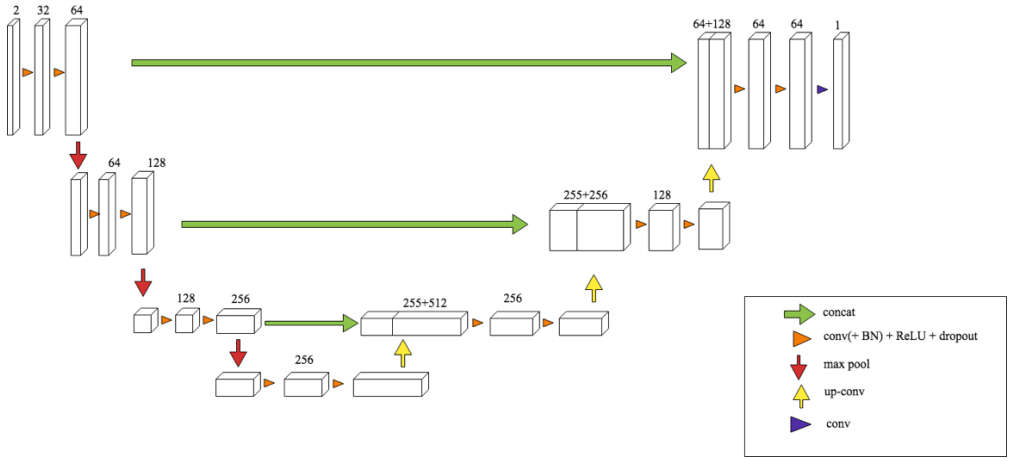

Figure 4. The modified 3D U-Net architecture we used in this study. Notice that the first input has only two channels (T1 and T2 images) instead of the original 3 RGB channels. Besides conv, BN, ReLU, we added Dropout.

2.3 Training parameters selection:

We employed the categorical cross-entropy as the cost function and applied optimization via the Adam method with a fixed learning rate of 0.001. We decided to use the whole image instead of patches, so the spatial window size is (96, 112, 96). We enabled normalization, with percentile norm type. The training was performed in batches of 1 image during each iteration. It took approximately 9000 iterations to reach an asymptote as shown in figure 5 below. The training model of the subcortical segmentation tasks was saved every 3000 iterations and was stopped after 9000 iterations, as they were found to be sufficient for convergence. The same parameters were used for pretraining with images segmented by the multi-atlas method. Training regular datasets required approximately 6 hours to complete. When batch size equals 1, training required 7.93GB total memory, with 7.81GB free memory. The training was conducted for 6 structures at the same time. Inferencing 81 imagers required 32 seconds.

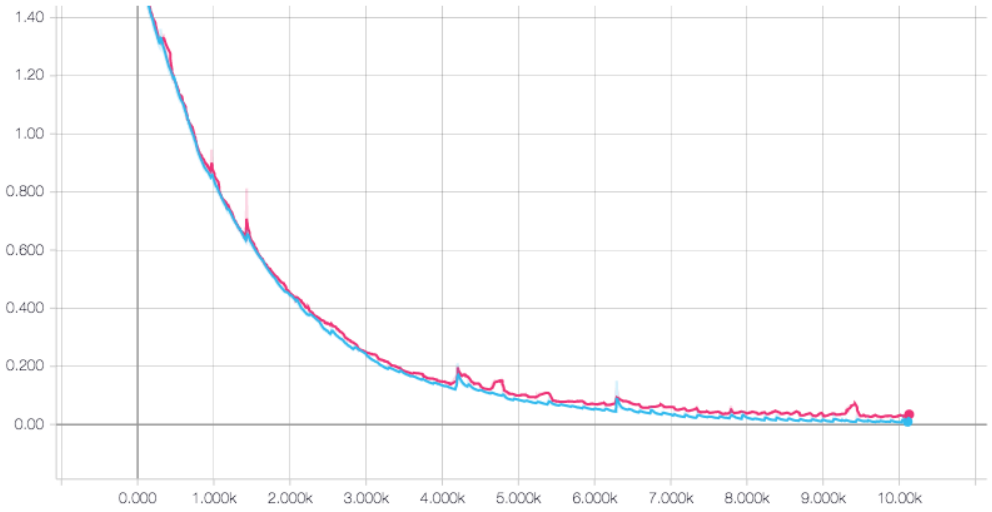

Figure 5. training and validation curve start to approach asymptote at 8k iterations

2.4 Postprocessing

We found several appearances of disconnected parts in nonuniform areas in images generated from inferencing testing cases. Those parts all are much smaller than those in the subcortical region. We postprocessed the segmentation by selecting the largest component for each label separately. Figure 6 shows the change after postprocessing

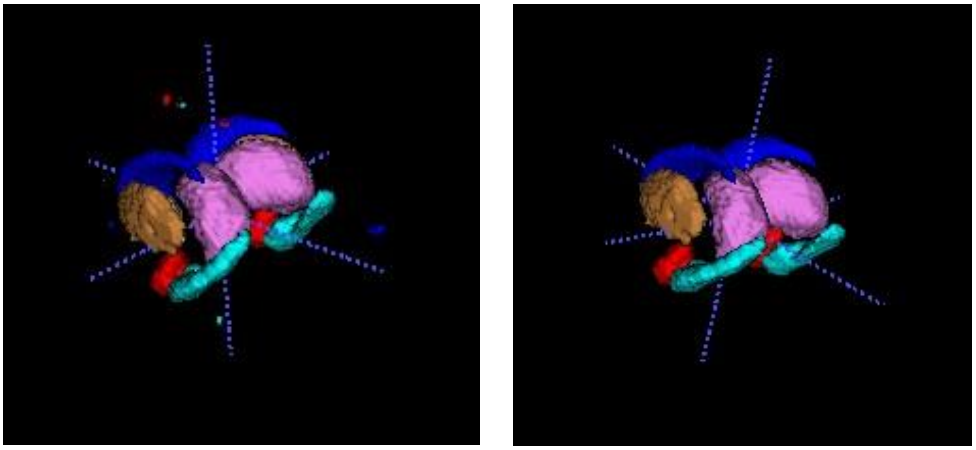

Figure 6. The left image is the original inference on testing cases, the right image is the left image after post-processing.

3. Results

In the following tables, we present our current best results. We use the Dice score to measure the accuracy of our model. All presented images and Dice scores are postprocessed with the removal of disconnected parts from the original network outputs. Figure 7 below shows the comparison between ground truth, pretraining-only model, and completely-trained model. Figure 8 lists the Dice score of each subcortical structure of three models above. Our adapted multi-modality 3D U-Net implemented on subcortical structures segmentation can achieve Dice score around 90 percent on most structures. This is slightly better, yet comparable accuracy to the multi-atlas approach employed in the pretraining. Computational efficiency is much improved with our adapted 3D U- Net performing segmentation in under 8 hours (on a standard GPU), whereas the multi-atlas method runs for over 36 hours (on a standard desktop workstation).

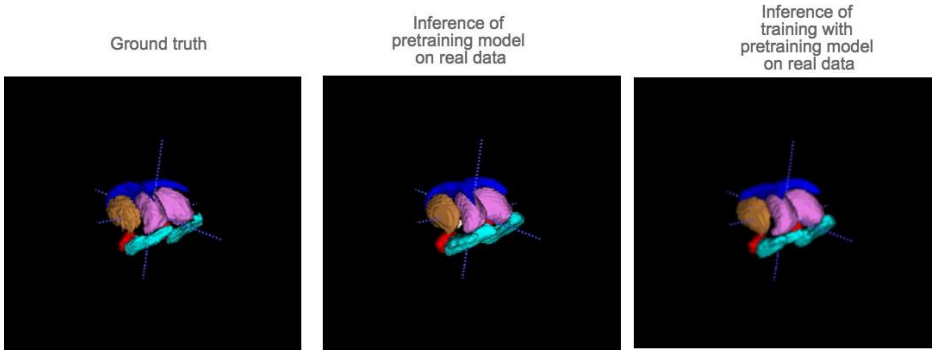

Figure 7. Comparison between ground truth, inference of pretraining model on real data and inference of training with the pretraining model on real data.

Label number	Structure names	Pretraining inference dice score	Pretraining inference on training data dice score	Pretraining + regular training inference on training data dice score
1	Amygdala	82.913	74.769	81.143
3	Caudate	92.548	83.640	87.952
5	Hippocampus	88.116	76.918	80.995
7	GP	87.294	66.414	82.639
9	Putamen	90.968	87.145	88.667
40	Thalamus	93.592	92.595	92.493

Figure 8. Dice score chart of 6 structures.

Conclusions

Our research shows that with relative straightforward extensions, a 3D-UNet approach performs at comparative accuracy and several orders of magnitude improved computational efficiency to multi-atlas segmentation. We expect that this result extends to many other biomedical volumetric segmentation tasks. Further improvements to the deep learning network model as well as the generation of additional training datasets are currently in process to further improve the accuracy of the proposed approach.

Acknowledgment

We thank NIH grants U54HDO79124 and R01EB021391 for support.

References

1. Shaw, P., et al (2017). 99. A Large Scale Study of Cortical and Cerebellar Morphology in ADHD across the Life span: An ENIGMA-ADHD Collaboration. *Biological Psychiatry*, 81(10), pp.S41-S42.
2. Geevarghese, R., et al (2015). Subcortical structure volumes and correlation to clinical variables in Parkinson's disease. *J. Neuroimaging* 25 (2), 275–280.
3. Ronneberger, O., et al (2015). U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention* (pp. 234-241).
4. Cerliani, L., et al (2015). Increased Functional Connectivity Between Subcortical and Cortical Resting-State Networks in Autism Spectrum Disorder. *JAMA Psychiatry*, 72(8), p.767.
5. van Erp, T., et al., (2016). Subcortical brain volume abnormalities in 2028 individuals with schizophrenia and 2540 healthy controls via the enigma consortium. *Molecular Psychiatry*, *21*(4), 547–553.
6. Swanson, M., et al. (2017). Subcortical Brain and Behavior Phenotypes Differentiate Infants With Autism Versus Language Delay. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, 2(8), pp.664-672.
7. Wang, J., et al (2014). Multi-atlas segmentation of subcortical brain structures via the AutoSeg software pipeline. *Frontiers in Neuroinformatics*, 8, 7. PMCID: PMC3915103