Semantic Segmentation for MRI using Deep Learning

Nikitas Sourdakos -i6212455

2 February 2021

1 Abstract

Whole brain segmentation from structural magnetic resonance imaging (MRI) is a necessary process for most morphological analyses. However, it is computationally heavy by traditional methods, and extremely time consuming if done by human experts. Thus here we propose an automated process for producing brain segmentation, based on Convolutional Neural Networks, that reduces the duration to the order of seconds, instead of hours, for a single subject. Our approach was to use the output of existing segmentation software as ground truth for training such a model, and thus a great amount of data can be acquired, without the many workhours of experts needed. Our proposed models work on segmenting the brains cortical as well as subcortical structures, achieving accuracies over 96 percent, with 30 different classes.

2 Introduction

Magnetic Resonance Imaging (MRI) provides detailed in vivo morphological information about the human brain, which is crucial to perform a variety of different studies, such as degenerative diseases or development. In order to know the volume or shape of a certain structure, the brain needs to be segmented, which is very time consuming when performed manually. Computational tools have been developed to do that automatically, such as Freesurfer (Fischl, 2012), however, even those methods are not without error, and are very computationally intensive, making them unsuitable for studies with a large number of subjects. In this work, we utilize the power of deep fully convolutional neural networks, to learn the segmentation process, using the output of Freesurfer as ground truth for our training. The trained network can then be run on a single GPU to automatically segment a brain in a few seconds, making it a viable alternative for processing large batches of brain data.

Our work here is twofold. First, we test our training paradigm with a classic architecture for semantic segmentation such as the Unet (Ronneberger, Fischer, & Brox, 2015), that works as a baseline, and then we compare some more recent variants, like Unet++ (Zhou, Siddiquee, Tajbakhsh, & Liang, 2018), as well as adding attention mechanisms via convolutional, non-local layers, as proposed by (Wang, Girshick, Gupta, & He, 2018). Finally, we propose our own architecture, named the SonicNet, that improves the results of the previous architectures, when compared with equal number of parameters, especially when working in a limited GPU memory regime.

3 Related Work

There have been numerous attempts in the last few years to tackle the brain segmentation problem, including the subcortical structures. In (Dolz, Desrosiers, & Ayed, 2018), they tackled the problem of segmenting between 8 subcortical structures, and the rest of the brain, with a simple fully convolutional network. In (Bontempi, Benini, Signoroni, Svanera, & Muckli, 2020) they use a 3D Unet, with 3 different scales, to segment brains into 7 different classes. A combination of a 3-D convolutional neural network with a Conditional Random Field for the final segmentation was used in DeepNat, (Wachinger, Reuter, & Klein, 2018), and on another study, named QuickNat, (Roy et al., 2019), they built an ensemble of 3 different 2-D Unets, for processing the coronal, sagittal and axial slices respectively, and combining them to give a final result for each voxel.

A modification of the Unet architecture, was used by the authors of (Zhou et al., 2018),to tackle the problem of segmenting organs other than the brain (liver and lungs) where nested, smaller U-nets were used to create a more densely connected architecture, and give additional outputs used for deep supervision (Lee, Xie, Gallagher, Zhang, & Tu, 2015). Their architecture was actually one that we tried out in this work, however, when controlling for the same number of parameters, it showed no improvement over the basic Unet. Moreover, trying deep supervision seemed to be counterproductive for this particular problem.

4 Methods

4.1 The Dataset

For this work, the OASIS 1 dataset was used (Marcus et al., 2007), both to train and test the efficiency of our models. This dataset consists of a cross-sectional collection of 416 subjects aged 18 to 96. For each subject, 3 or 4 individual T1-weighted MRI scans obtained in single scan sessions are included. The subjects are all right-handed and include both men and women. 100 of the included subjects over the age of 60 have been clinically diagnosed with very mild to moderate Alzheimer's disease. This dataset was chosen both for its availability, as it not only offers freely all of the scans, but also a secondary dataset of freesurfer outputs, which were used as ground truth, as we will discuss in the methods section, and also for the diversity of its subjects.

The dataset was split into two different ways for training and testing. The first split was for 100 subjects to become the test set, and the 316 that remained were put in a pool, from which 100 were chosen randomly, one for each epoch that our models were trained. The second setup was focused more on testing, as the test set was comprised of 200 subjects, and the training and validation sets where made from the rest. The brain size was varied between all these subjects. Thus, as a pre-processing step, besides cropping all the unnecessary "void" around the head, as well as centering the brains, the data were trimmed to fit the size of the smallest of them. We found out that this way we gained the most, since a bigger percentage of the volume was occupied by the more difficult classes, the subcortical structures. Losing a small percentage of the outer parts of the brain and the skull proved to have zero effect on the outcome, since the white-gray matter disambiguation is one of the easiest parts for our networks to handle.

4.2 Models

The models that were used fall into two categories, the Unet, (and some variants), and the SonicNet. The Unet used was created with 4 layers of maxpooling, so the analysis was on 5 different scales. Two convolutional layers, each followed by a batch normalization layer were used before and after each Max-Pooling layer. Attention was added in the three "deepest" layers of the Unet, as the all-to-all maps became too much for the memory to handle, as the volume increased in size. We used two different data input sizes on both models, cubes of 64x64x64 and 48x48x48. The choice of the second input size, although proving a little inferior in performance, was made to accommodate devices with below 12 and above 6 gigabytes of VRAM.

4.3 SonicNet

The SonicNet consists of 6 Unets of depth 3, called U-blocks in the following diagrams. After the cycle of each U-block, the final layer is concatenated with the first layer, and they are fed to the next U-block. What is more, each of the layers has skip feedforward connections to every other layer in the same level of encoding in the rest of the U-blocks, creating many different paths where information can flow. To keep the parameters to a comparable size to the Unet (around 28 million, going from the uppermost layers having 64 filters, to the deepest ones having 384 filters), each U-block was a lot "thinner", having 64 filters per layer in the first 4 U-blocks, and 128 filters in the final two.

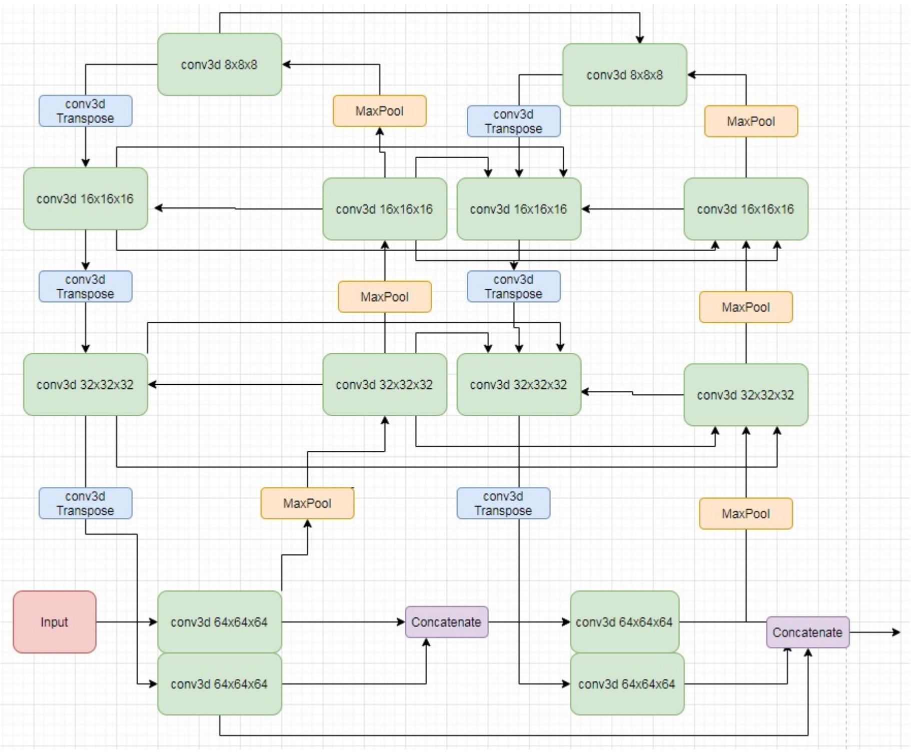

Figure 1: Detailed diagram of the first 2 Ublocks of SonicNet, with a starting block of 64x64x64. One small detail that was omitted for space economy is that those convolutional layers, actually are convolutional blocks, that include two different convolutional layers of the same resolution, and two layers of batch normalization

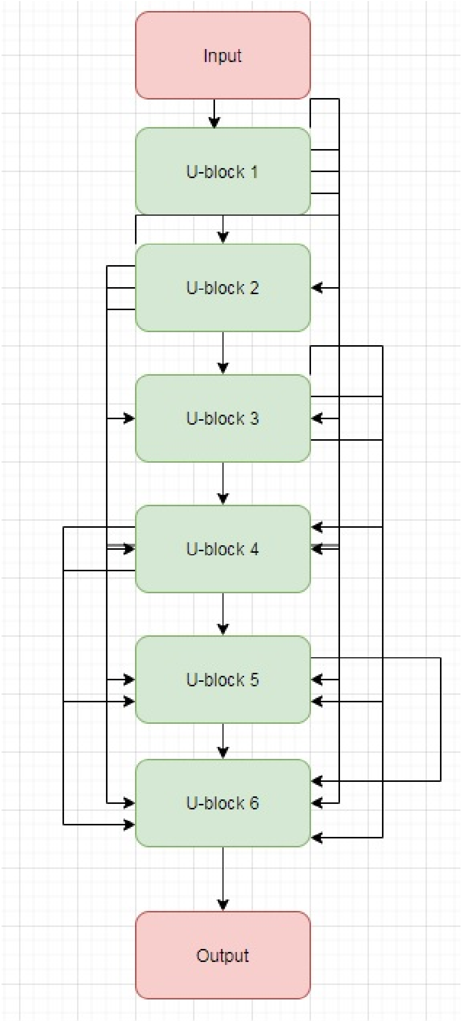

Figure 2: All 6 U-blocks of the SonicNet

5 Results

We measure out networks power with a simple accuracy metric, between the output and the ground truth of freesurfer labels. As we see in the table below, the decrease in accuracy for our more minimal setup is substantial, a 0.8 percent, from its more memory intensive counterpart. SonicNet seems to improve by 1.0 percent versus the Unet in the 48-sized cube regime, while it offers only a 0.23 improvement in the 64-sized cube regime. That is the improvement on average.

On a sample of 100 test subjects, SonicNet, in the 48-cube regime, shows a significant improvement, ($p < 0.05$) with a margin of 0.38 percent in accuracy.

Model	Accuracy
U-NET-48	94.95
SonicNET-48	**96.00**
U-NET-64	96.58
SonicNET-64	**96.81**

We also show in the table below, the class specific accuracies for the SonicNet-48. The classes that did not exist in the data are shown with **, and those that did not exist in the predictions only are shown with *.

Class	Accuracy(Unet)	Accuracy(SonicNet)
0-Unknown	97.81	97.95
1-Cerebral-Exterior	**	**
2-Cerebral-White-Matter	96.16	97.29
3-Cerebral-Cortex	85.56	91.01
4-Lateral-Ventricle	97.06	95.38
5-Inferior Lateral-Ventricle	67.41	**74.41**
6-Cerebellum-Exterior	**	**
7- Cerebellum-White-Matter	88.84	97.31
8-Cerebellum-Cortex	96.76	94.57
9-Thalamus	95.09	93.86
10-Caudate	94.00	91.31
11-Putamen	93.90	91.43
12-Pallidum	68.26	**81.65**
13-3rd-Ventricle	76.39	**96.98**
14-4th-Ventricle	*	**96.09**
15-Brain-Stem	93.56	92.31
16-Hippocampus	81.46	87.14
17-Amygdala	83.72	84.92
18-Cerebrospinal Fluid	91.65	85.83
19-Accumbens-area	73.28	78.61
20-Substancia-Nigra	**	**
21-VentralDC	69.85	**80.68**
22-Choroid-plexus	**	**
23-5th-Ventricle	**	**
24-hypointensities	60.98	68.27
25-Optic-Chiasm	**	**
26-Corpus-Callosum	**	**
27-Dura Matter	**	**
28-Other	**	**
29-Cingulate Cortex	**	**

Sonicnet does better in most of the classes. Considerable improvement is shown for the Pallidum, the 3rd and 4th ventricles, as well as the VentralDC, where we see an increase of more than 10 percent in comparison with the basic Unet. Of particular interest is the 4th ventricle, which Unet seems to not notice entirely. The weakest classes are the inferior lateral ventricle, the Accumbens area, the Ventral DC, and certain hypointensity areas.

So SonicNet achieves better or comparably close results for all of the classes. However, as the no free lunch theorem suggests, there is a tradeoff. And that comes with an increased test time, and training time, as it can be seen in the table below.

Model	Train Time	Test Time
U-NET-64	14 minutes	22 seconds
SonicNET-64	33 minutes	29 seconds

In the following figure we can see the results of the predictions of our network, as visualized with the BrainVoyager software (Goebel, 2012).

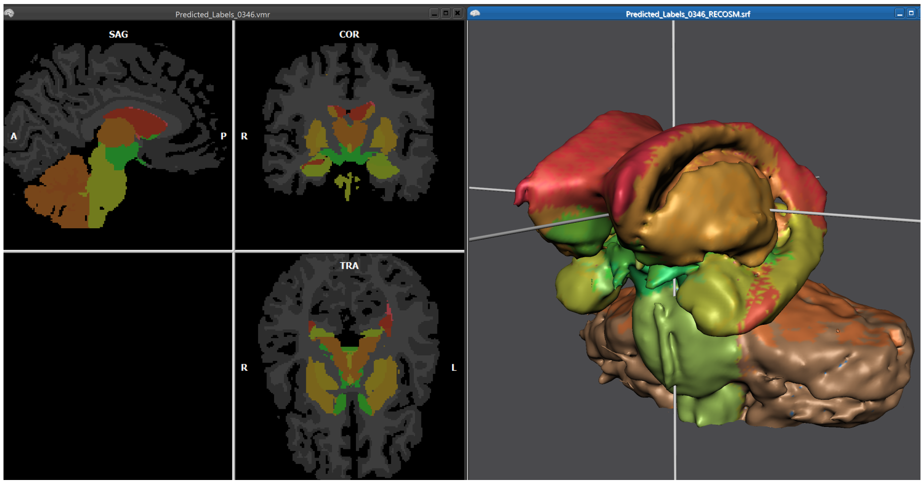

Figure 3: This is a visualization of the predicted labels, in 2D and 3D, for subject 0346 of the OASIS1 dataset. On the left we have the 2D slices, and on the right is a 3D reconstruction.

6 Future Work

Although this work showed some improvements in the problem of subcortical segmentation, a lot of work still can be done. First on the list, would be to test the current methods with more datasets, to see their generalization properties.

That of course requires a the processing through Freesurfer of another dataset, which is a time consuming process. Another issue is the reduced accuracy on some of the classes that are not heavily represented. A possible improvement there might be to tweak the loss function so that it penalizes mistakes in those areas more heavily.

An open issue is the boundary surface reconstruction. Right now, a max function is posed on the output probabilities of the network, so that we have a definitive class in each voxel.

Alternatively, we might keep the original output, so that it might perhaps better inform the decisions of the reconstructive process. If neither of the approaches satisfies, a new direction might be to artificially increase the resolution of our data, using super-resolution techniques. Recent advances, such as (Ledig et al., 2017), have made huge improvements over traditional methods like the somewhat blurry bicubic interpolation. Perhaps running the classification on the augmented data, might provide with better results for the final reconstruction process, although there is a tradeoff there, as increasing the resolution, with the GPU memory limits as they are, means a smaller field of view, as a percentage of the whole brain. Finally, with the acquisition of some hand annotated data in the future, a seconds training could take place, as a fine-tuning step, to make some final improvements.

7 Conclusions

While both the classic Unet, and our new model, the SonicNet were able to disambiguate accurately, for the most part, the subcortical structures, the SonicNet was able to have a small advantage in accuracy, in both of the data setups, with cubes of size 64 and especially in the cubes of size 48. The improved performance can be attributed to two factors, the increased number of skip connections, and the increased overall depth. These improvements however, come at a cost, which is larger training time. SonicNet takes about 2.35 times the hours it takes to train the basic Unet, always for a similar number of parameters.

Although accuracies of over 96 percent are pretty high, the ability of our networks to improve even further is perhaps hampered by the fact that our input data might not be one hundred percent consistent, since they are the output of the Freesurfer software, and not manually hand annotated, and while Freesurfer has been shown to perform well on most of the voxels, in the parts that belong to decision boundaries between two structures a lot of ambiguity remains. Even experts may not label some borders in the same way, since the MRI contrast does not provide precise tissue boundaries for some subcortical structures. Data overlap was another factor for training where values between 0.5 and 0.8 were tested. Overall, it seems that the higher overlap produced better results, though that might be due to essentially more data being created per epoch. Overall, further experimentation will be needed, to test wider, and slimmer architectures, as well as testing the performance of our network on different datasets.

References

Bontempi, D., Benini, S., Signoroni, A., Svanera, M., & Muckli, L. (2020). Cerebrum: a fast and fully-volumetric convolutional encoder-decoder for weakly-supervised segmentation of brain structures from out-of-the-scanner mri. *Medical image analysis*, *62*, 101688.

Dolz, J., Desrosiers, C., & Ayed, I. B. (2018). 3d fully convolutional networks for subcortical segmentation in mri: A large-scale study. *NeuroImage*, *170*, 456–470.

Fischl, B. (2012). Freesurfer. *Neuroimage*, *62*(2), 774–781.

Goebel, R. (2012). Brainvoyager—past, present, future. *Neuroimage*, *62*(2), 748–756.

Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., ... others (2017). Photo-realistic single image super-resolution using a generative adversarial network. In *Proceedings of the ieee conference on computer vision and pattern recognition* (pp. 4681–4690).

Lee, C.-Y., Xie, S., Gallagher, P., Zhang, Z., & Tu, Z. (2015). Deeply-supervised nets. In *Artificial intelligence and statistics* (pp. 562–570).

Marcus, D. S., Wang, T. H., Parker, J., Csernansky, J. G., Morris, J. C., & Buckner, R. L. (2007). Open access series of imaging studies (oasis): cross-sectional mri data in young, middle aged, nondemented, and demented older adults. *Journal of cognitive neuroscience*, *19*(9), 1498–1507.

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *International conference on medical image computing and computer-assisted intervention* (pp. 234–241).

Roy, A. G., Conjeti, S., Navab, N., Wachinger, C., Initiative, A. D. N., et al. (2019). Quicknat: A fully convolutional network for quick and accurate segmentation of neuroanatomy. *NeuroImage*, *186*, 713–727.

Wachinger, C., Reuter, M., & Klein, T. (2018). Deepnat: Deep convolutional neural network for segmenting neuroanatomy. *NeuroImage*, *170*, 434–445.

Wang, X., Girshick, R., Gupta, A., & He, K. (2018). Non-local neural networks. In *Proceedings of the ieee conference on computer vision and pattern recognition* (pp. 7794–7803).

Zhou, Z., Siddiquee, M. M. R., Tajbakhsh, N., & Liang, J. (2018). Unet++: A nested u-net architecture for medical image segmentation. In *Deep learning in medical image analysis and multimodal learning for clinical decision support* (pp. 3–11). Springer.

8 Appendix

This is a test on a new dataset, from (http://brain-development.org/ixi-dataset/). It consists of 581 unprocessed scans of adults. Here our results are for only the first 10. It was scanned with the latest version of Freesurfer, which includes tha label 29 for the Cingulate Cortex. As a result the accuracy of our network for the cerebral cortex drops a lot, since it missclassifies the cingulate cortex as class 3, Cerebral Cortex, in accordance with the scans from the OASIS dataset, that did not include that label. Other than that, the weakest classes are the Inferior-Lateral-Ventricle, the Putamen and the Pallidum.

Class	Accuracy(SonicNet)
0-Unknown	95.87
1-Cerebral-Exterior	**
2-Cerebral-White-Matter	94.83
3-Cerebral-Cortex	78.75
4-Lateral-Ventricle	94.06
5-Inferior Lateral-Ventricle	33.76
6-Cerebellum-Exterior	**
7- Cerebellum-White-Matter	87.90
8-Cerebellum-Cortex	87.35
9-Thalamus	80.85
10-Caudate	67.60
11-Putamen	36.65
12-Pallidum	42.87
13-3rd-Ventricle	95.19
14-4th-Ventricle	96.76
15-Brain-Stem	91.7
16-Hippocampus	83
17-Amygdala	68.63
18-Cerebrospinal Fluid	81.96
19-Accumbens-area	50.54
20-Substancia-Nigra	**
21-VentralDC	82.28
22-Choroid-plexus	**
23-5th-Ventricle	**
24-hypointensities	29.22
25-Optic-Chiasm	**
26-Corpus-Callosum	**
27-Dura Matter	**
28-Other	**
29-Cingulate Cortex	**