Using neural networks for post-stroke lesion detection in the ATLAS dataset

Diana D. Chin
ddchin@stanford.edu

William R. T. Roderick
wrtr@stanford.edu

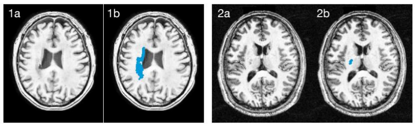
Karen M. Wang
kmwang14@stanford.edu

Overview

Post-stroke lesion detection is a process that currently takes skilled tracers up to an hour per scan [1]. Automating lesion detection will not only help radiologists catch problematic lesions in many clinical domains but will also enable us to rapidly expand neuroimaging datasets, which can then be used to improve our understanding of how MRI brain scans relate to recovery prognoses and suitable treatments. In this work, we test deep learning methodologies used in prior medical image segmentation studies [2-4] on USC's new (2018) Anatomical Tracings of Lesions After Stroke (ATLAS) dataset. Given a series of MRI slices of the brain as input, our model predicts segmentation masks identifying the locations of post-stroke lesions as output. We demonstrate how a U-Net architecture, applying dilation, or using Gaussian blurring are relatively ineffective for improving the dice coefficient of our predictions while the greatest performance can be derived by cascading an encoding/decoding neural network architecture.

The ATLAS Dataset

Data. The publicly available ATLAS dataset [1] includes 229 T1-weighted MRI scans (from n=220 patients) with segmented lesions. Each scan includes a series of (grayscale) MRI slices and one or more series of lesion masks. The original images are converted to Numpy arrays of pixel intensity values, which are then normalized to values between 0 and 1 to be input into the neural network.

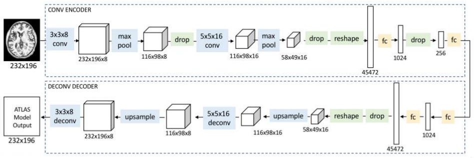


Examples of two pairs of MRI slices (1a,2a) with the corresponding lesion mask overlaid in blue (1b,2b) from the ATLAS dataset.

Features. Each brain scan is a 232 x 196 image, so the raw input data contains 45,472 features per image. We chose not to introduce any additional features in this project in order to assess the ability of our network to detect lesions based only on the original MRI scans.

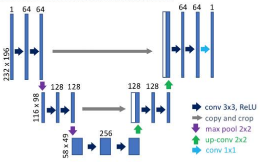
Models

The Atlas Model



The original baseline neural network architecture.

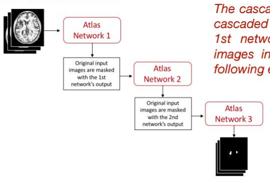
U-Net



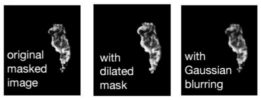
The U-Net architecture, adapted from [2].

We also reduced the network size in a "medium" version, by reducing the number of filters by a factor of 4, and a "small" version, which additionally replaced the 2x2 max pools with one 4x4 max pool and the 2x2 up-conv steps with one 4x4 up-conv.

Cascaded Atlas


Atlas Network 1
Original input images are masked with the 1st network's output
Atlas Network 2
Original input images are masked with the 2nd network's output
Atlas Network 3

The cascaded network architecture, shown here for a twice cascaded system. We also tried dilating the output of the 1st network before masking, and blurring the masked images input into the second network, as shown in the following examples:



Results

Our training and dev sets had 12984 and 1000 samples, respectively, and were evaluated by their DICE* scores (training scores were calculated on 100 sample subsets). Numbers in parentheses show the change from the baseline model after training for the same number of epochs (per network for the cascaded models).

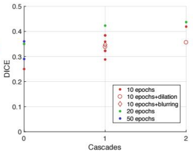
Model	Training DICE	Dev DICE
U-Net, small (3 epochs)	0.020 (-74%)	0.017 (-79%)
U-Net, medium (3 epochs)	0.030 (-62%)	0.024 (-70%)
U-Net, large (3 epochs)	0.071 (-8.9%)	0.028 (-65%)
Single Cascaded Atlas (20 epochs/network)	0.55 (+20%)	0.42 (+21%)
Single Cascaded w/ dilation (10 epochs/network)	0.39 (+56%)	0.34 (+35%)
Single Cascaded w/ blurring (10 epochs/network)	0.47 (+88%)	0.35 (+39%)
Double Cascaded w/ dilation (10 epochs/network)	0.44 (+76%)	0.36 (+42%)
Double Cascaded Atlas (20 epochs/network)	0.58 (+26%)	0.44 (+25%)

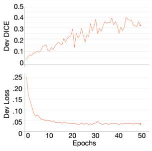
$$^*\text{DICE} = \frac{2\,TP}{2^*TP + FP + FN}$$

TP = true positive, FP = false positive, FN = false negative pixel count

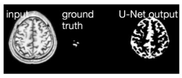
Results & Discussion

Cascading the Atlas model successfully improved segmentation performance, but with diminishing returns for each additional cascade. Dilating was likely less helpful because the relative recall/ precision weightings in the loss function of the first network favored false positives, so the predicted mask was generally already larger than the target. Blurring may have been ineffective because it leads to the loss of potentially useful information.





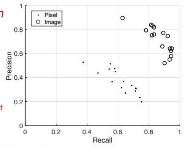
As expected, running more epochs also initially increased the DICE score, but only for ~45 epochs, and with diminishing returns after about 20.



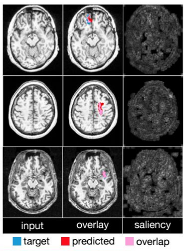
The U-Net performed worse than expected as it appeared to only be thresholding, but we suspect that it would improve after much more training.

We also noted a tradeoff between recall and precision produced by adjusting the weighting in our cross entropy loss function.

Recall_{pixel} = TP/(TP+FN)
Precision_{pixel} = TP/(TP+FP)
Recall_{image} =O/T
Precision_{image} = O/M
O = #images where predicted mask overlaps with target
T = #images with a target
M = #images with a predicted mask



A saliency map illustrates the gradient of the predicted mask probabilities with respect to the input image pixels, where a high gradient, bright pixel indicate the importance of that input pixel in determining the output probabilities. The most salient pixels seem to cluster around gray and high contrast areas.



Future Work

The difference between the training and dev set performances indicate that the model has high variance, so a promising next step would be to use more training data to resolve this discrepancy. Additional regularization and further tuning the hyperparameters would also likely help. To improve the training and dev performance further, we would recommend training a larger model, exploring volumetric segmentation models from sparsely labeled images, or incorporating lesion metadata (ex. primary stroke location and hemisphere or vascular territory).

References

[1] S.-L. Liew, J. M. Anglin, N. W. Banks, M. Sondag, K. L. Ito, H. Kim, J. Chan, J. Ito, C. Jung, N. Khoshab, S. Lefebvre, W. Nakamura, D. Saldana, A. Schmiesing, C. Tran, D. Vo, T. Ard, P. Heydari, B. Kim, L. Aziz-Zadeh, S. C. Cramer, J. Liu, S. Soekadar, J.-E. Nordvik, L. T. Westlye, J. Wang, C. Winstein, C. Yu, L. Ai, B. Koo, R. C. Craddock, M. Miham, M. Lakich, A. Pienta, and A. Stroud, "A large, open source dataset of stroke anatomical brain images and manual lesion segmentations," *bioRxiv*, 2017.

[2] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," *CoRR*, vol. abs/1505.04597, 2015. [Online]. Available: http://arxiv.org/abs/1505.04597

[3] G. Wang, W. Li, S. Ourselin, and T. Vercauteren, "Automatic brain tumor segmentation using cascaded anisotropic convolutional neural networks," *CoRR*, vol. abs/1709.00382, 2017. [Online]. Available: http://arxiv.org/abs/1709.00382

[4] L. Spies, A. Tewes, P. Suppa, R. Opfer, R. Buchert, G. Winkler, and A. Raji, "Fully automatic detection of deep white matter t1 hypointense lesions in multiple sclerosis," *Physics in Medicine and Biology*, vol. 58, pp. 8323–8337, 2013.

Acknowledgments

We would like to thank our mentor, David Eng, for providing baseline code to help us get started and for his guidance and support throughout the project. We are also grateful to the CS 230 teaching team for their hard work in making this course a success.