Automated Segmentation Mask Generation of MRI scans for Post-Stroke Lesions

Elizabeth Botbol Ponte, Marcus Paul Lavengco, Oscar Rodriguez Guerrero
[ebotbol, lavengco, oguerrer]@stanford.edu
Department of Electrical Engineering, Stanford University, Stanford, California

Stanford ENGINEERING

Introduction

Objectives: improving upon a model for image segmentation using Convolutional Neural Networks.

Background and Motivation Stroke is the leading cause of adult disability worldwide, with up to two-thirds of individuals experiencing long-term disabilities. As predictive algorithms improve, a long-term goal is for clinicians to use MRI to predict the likelihood of recovery, or more importantly, their likelihood of responding to different and more personalized types of therapies.

Dataset

ATLAS (Anatomical Tracings of Lesions After Stroke) Release 1.1, is an open-source dataset consisting of 304 T1-weighted MRIs with manually segmented diverse lesions and metadata. The goal of ATLAS is to provide the research community with a standardized training and testing dataset for lesion segmentation algorithms on T1-weighted MRIs.

ATLAS 1.1 was released to the public only a few months ago in February 2018, thus there is no published research papers using the dataset.

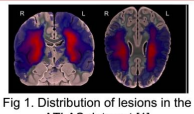

Fig 1. Distribution of lesions in the ATLAS data set [1]

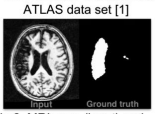

Fig 2. MRI scan (input) and segmentation mask (ground truth) [1]

Implementation: Models, Loss, and Performance Metrics

Dice coefficient: used for comparing the similarity of two samples.

$$DSC = \frac{2\,|P \cap R|}{|P| + |R|} = \frac{2TP}{2TP + FP + FN}$$

Weighted Cross Entropy: $\omega_{Cn} > 1$ decreases the false negative count, hence increasing the recall. A $\omega_{Cn} < 1$ decreases the false positive count and increases the precision.

$$\mathcal{L}_n(\mathbf{W}) = -\gamma w_{c_n} \log \mathbf{y}^{(1)}_{c_n}(\mathbf{x}_n, \mathbf{W}) - (1-\gamma) \log \mathbf{y}^{(2)}_{l_n}(\mathbf{x}_n, \mathbf{W})$$

Tversky Loss: generalized loss function based on Tversky index to address data imbalance with very small number of training images, and achieve better trade-off between precision and recall.

$$T(\alpha, \beta) = \frac{\sum_{i=1}^{N} p_{0i} g_{0i}}{\sum_{i=1}^{N} p_{0i} g_{0i} + \alpha \sum_{i=1}^{N} p_{0i} g_{1i} + \beta \sum_{i=1}^{N} p_{1i} g_{0i}}$$

U-Net

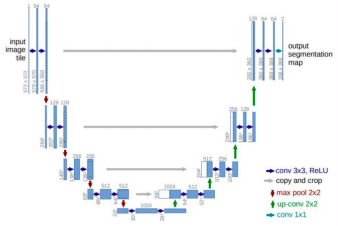

Fig 3. U-Net architecture [3]

AlexNet

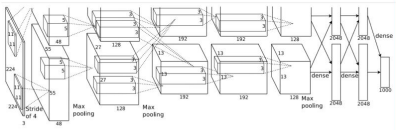

Fig 4. AlexNet architecture [4]

Results

Analysis of baseline performance shows that the model tends to overestimate; we made a few hypotheses as to why. The first was that overestimation was due to the CE weight being too high, excessively overemphasizing true positives. A second hypothesis was that overestimation was due to too small of a batch size, causing parameter updates to occur too often. This is particularly important because of large variance in our dataset. We also experimented with variations in the model architecture, investigating average pooling in place of max pooling, and adding layers in an AlexNet inspired fashion, to deepen the network. The largest improvement was found by adjusting the weight parameter on the weighted CE loss function.

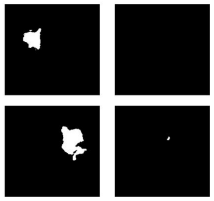

Fig 5. Examples of labels for slice 80 across different patients and sites

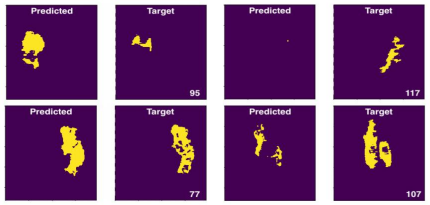

Fig 6. Target and predicted examples with CE weight of 10

Discussion, Conclusions, and Future Work

Our experiments primarily compared individual changes directly to the baseline model. However, the baseline's performance was heavily inhibited by the unoptimized weight parameter in the weighted CE loss function. The weight optimization proved to be a bottleneck in our performance, thus improvement due to independent changes made in other aspects of the model (e.g., hyperparameters, model architecture) was marginal, and provided little insight towards improvement.

Our attempt at solving the lesion segmentation problem employs an end-to-end approach with a limited dataset, inhibited by high unbalance, and low expressivity, mostly due to its small size. Our current end-to-end model shows promising performance, despite the limitations, thus improvements in the dataset may have a profound effect. This may be achieved by data augmentation. Further improvements may be found by segmenting the task into three problems: classification, bounding box prediction, and image segmentation.

References and Acknowledgments

[1] Sook-Lei Liew, et. al., A large, open source dataset of stroke anatomical brain images and manual lesion segmentations. Scientific Data, 2018; 5: 180011 DOI: 10.1038/sdata.2018.11
[2] S. S. M. Salehi, D. Erdogmus, and A. Gholipour, "Tversky loss function for image segmentation using 3D fully convolutional deep networks," International Workshop on Machine Learning in Medical Imaging (2017).
[3] Ronneberger, O., Fischer, P. & Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. International Conference on Medical image computing and computer-assisted intervention 9351, 234–241 (2015).
[4] A. Krizhevsky, I. Sutskever, and G. Hinton. Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems, page 1097--1105 (2012).