DualNetFC: Brain Lesion Segmentation via Dual Pathway Network

Weston Ungemach (wungemach@stanford.edu) and Beite Zhu (jupiterz@stanford.edu)

CS230 Deep Learning

Problem Statement

1. Brain lesions in stroke sufferers inform medical treatment during surgical consultation and when assessing speech and motor function impairment.
2. The industry standard for lesion segmentation is still expert hand-labeling, which is time-intensive and inefficient.
3. Here, we implement DualNetFC: a dual-pathway network on the ATLAS dataset to produce lesion segmentation masks.

The ATLAS Dataset

1. We analyze the ATLAS dataset, newly compiled and released by USC, containing 229 T1-weighted MRI scans labeled with a 3D mask segmenting (possibly) multiple lesions.
2. We process this as a dataset of 43281 horizontal cross-sections.
3. MRI scans in ATLAS are more detailed than the multimodal clinical scans in the ISLES dataset and the experts that compiled this dataset do not expect transfer learning from that dataset to be successful here.

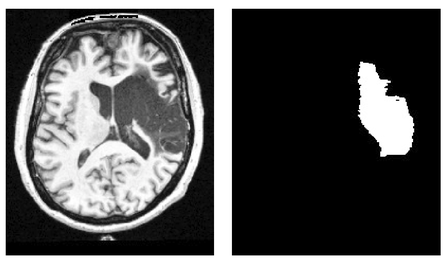

Figure 1: A sample lesion-mask pair from the ATLAS dataset

Objective

The loss function here is a weighted pixel-wise cross-entropy:

$$\mathcal{L} = -\left[\sum_{p \in \text{pixels}} 5\hat{p}\log(p) + (1-\hat{p})\log(1-p)\right]$$

The relative weights of the terms encourages the network to make positive predictions, which effectively balances the positive/negative class sizes.

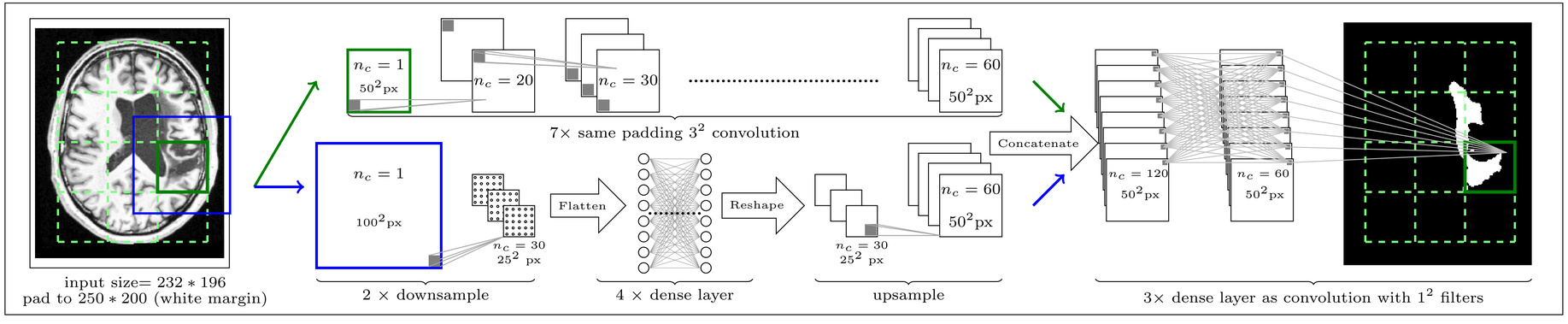

Figure 2: The network takes in a 232×196 image and passes slices of size 50×50 (green) and 100×100 (blue) through the upper and lower paths, respectively. The upper path is fully convolutional, while the lower path has fully connected layers. These pathways are merges with a series of 1×1 convolutions, then the various croppings are synthesized into a predicted output mask.

Network Details

We employ a dual-pathway approach. We tile the image with with 50×50 crops (in green), each of which is centered in a larger 100×100 crop (in blue), as pictured above. These crops are successively passed through the network and synthesized at the end to a target mask.

1. The intuition for the two pathways is that the top path learns the local shape of the lesion while the lower path provides context for the location of that patch in the slice.
2. The network has approximately 19.5 million parameters.
3. The upper pathway is 10 layers deep while the lower pathway is 11 layers.
4. Convergence takes approximately 50 epochs, with a plateau at 40 epochs solved by learning rate decay.
5. The runtime on a single GPU is 6 minutes/epoch.

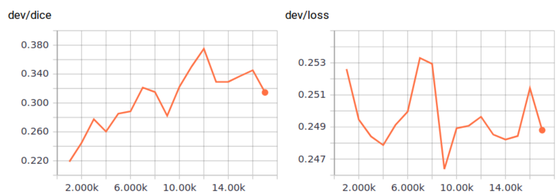

Figure 3: Plots of the Sørensen-Dice Coefficient and the loss on the test set. Note that the scale on the right is 0.002, while that on the left is 0.04.

Results

Accuracy in image segmentation is usually measured with respect to the Sørensen-Dice Coefficent, which for two sets X, Y is given by

$$\text{Dice}(X,Y) = \frac{\text{Area of } X \cap Y}{\text{Area of } X \cup Y} \in [0,1]$$

Our network achieves a maximum coefficient of 0.375 on the test set. This score takes about 50 epochs, after which the network begins overfitting the training set.

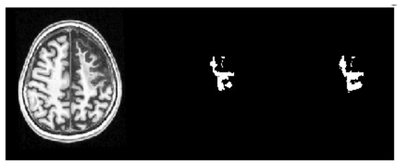

Figure 4: A sample output. From left to right, the input image, the target output, and the predicted output.

Network	Maximum SDC	Number of Parameters
Encoder/Decoder	0.40	93,705,297
DualNetFC	0.38	19,469,835
DualNetVeryFC	0.26	45,233,281
DualNetBigBlue	0.20	39,064,203
DualNetOverlap	0.31	19,592,925

Table 1: Encoder/decoder is a simple conv/fc/deconv, DualNetVeryFC has fully connected layers in the upper pathway, DualNetBigBlue passes the entire input as the blue crop, and DualNetOverlap overlaps the green crops.

Discussion/Further Directions

1. A standard encoder/decoder network achieves a similar Sørensen-Dice Coefficient in the same amount of time, but uses almost five times the number of parameters.
2. Our model has a bias issue: the Sørensen-Dice Coefficeint on the train set only reaches 0.60. Simple network modifications proved difficult to train and did not improve results.
3. Our network does use both pathways, as removing the upper pathway achieves a lower Sørensen-Dice coefficient.
4. A natural next-step is to process each input as a 3D object. Our local-to-global approach will hopefully continue to be effective there, as shown in [1].

Acknowledgements

We would like to thank our mentor, David Eng, for his guidance throughout this project.

References

[1] Konstantinos Kamnitsas, Christian Ledig, and Virginia F. J. Newcombe et al. Efficient multi-scale 3d CNN with fully connected CRF for accurate brain lesion segmentation. *CoRR*, abs/1603.05959, 2016.

[2] Marcel Prastawa, Elizabeth Bullitt, Sean Ho, and Guido Gerig. A brain tumor segmentation framework based on outlier detection*1. 8:275–83, 10 2004.

[3] B. H. Menze, A. Jakab, and S. Bauer et al. The multimodal brain tumor image segmentation benchmark (brats). *IEEE Transactions on Medical Imaging*, 34(10):1993–2024, Oct 2015.