SUPPLEMENTAL MATERIAL

The fraction of false positives + fraction of false negatives loss is

$$\mathcal{L}(n) = \lambda(n)\mathrm{FP}_{\mathrm{frac}} + \mathrm{FN}_{\mathrm{frac}}$$

Where

$$\mathrm{FP}_{\mathrm{frac}} = \frac{1}{N}\sum_{i\in\mathcal{I}}(1-t_i)p_i \qquad \mathrm{FN}_{\mathrm{frac}} = \frac{1}{N}\sum_{i\in\mathcal{I}} t_i(1-p_i)$$

Where N is the total number of voxels, t is the ground truth binary label, and p is the predicted softmax probability, and i is the voxel index. This loss varies between 0 and 1.

The loss depends on n, the iteration number during training . $\lambda(n)$ is a sigmoid function defined by :

$$\lambda(n) = \lambda_{\min} + \frac{1-\lambda_{\min}}{1+e^{-\vartheta(n)}} \quad \text{where} \quad \vartheta(n) = \frac{n - {n_{\max}}/{2}}{{n_{\max}}/{10}}$$

$\lambda_{\min}$ and $n_{\max}$ had to be tuned by hand and were set to 0.01 and 20,000. Thus, λ is approximately equal to 1 after 20,000 iterations.

Generalized Dice for 2 labels (foreground and background) is defined as :

$$\mathrm{GDL} = 1 - 2\frac{\sum_{l=1}^{2} w_l \sum_i t_{li}p_{li}}{\sum_{l=1}^{2} w_l \sum_i t_{li} + p_{li}}$$

Where l is the class index, i is the voxel index, and the weighting factors are:

$$w_l = \frac{1}{(\sum_i t_{li})^2}$$

(for more info see https://arxiv.org/pdf/1707.03237.pdf)

The final loss was :

$$\mathcal{L}(n) = 10\mathrm{GDL} + \lambda(n)\mathrm{FP}_{\mathrm{frac}} + \mathrm{FN}_{\mathrm{frac}}$$

The generalized Dice loss was weighted 10 the FP+FN loss. This final choice of weighting was fairly arbitrary, but was based on the idea that we desired very accurate segmentations, and the intuition that generalized Dice loss may be better for that. The choice of these two loss functions was arrived at after a few iterations of trial and error with several different loss functions considered, including voxel-wise cross-entropy and traditional Dice loss.