Paper Reading No.11

Boundary loss for highly unbalanced segmentation

Sheng Lian

August 2019

1 Brief Paper Intro

- ***Paper ref:*** MIDL 2019, oral paper. http://arxiv.org/abs/1812.07032

  The authors presentation on MIDL 2019 can be viewed in https://www.bilibili.com/video/av58414058?pop_share=1, in 209:40.

- ***Authors:*** See Fig 1.

Hoel Kervadec[*1] hoel.kervadec.1@estmtl.net
Jihene Bouchtiba[*1] jihene.bouchtiba.1@ens.etsmtl.ca
Christian Desrosiers[1] christian.desrosiers@etsmtl.ca
Eric Granger[1] eric.granger@etsmtl.ca
Jose Dolz[1] jose.dolz@etsmtl.ca
Ismail Ben Ayed[1] ismail.benayed@etsmtl.ca
[1] *ÉTS Montreal*

Figure 1: authors' brief intro.

- ***Paper summary:*** Dice loss and cross-entropy loss are widely used in the task of medical image segmentation. Both of them are based on integrals (summations) over the segmentation regions. They are sensitive to high class-imbalance. This paper propose a boundary loss, which takes the form of a distance metric on the space of contours (or shapes), not regions. It can deal well with high class-imbalance task.

- ***Reading motivation:*** Nearly all the loss we've seen in the task of medical image segmentation is region-based, for example, dice, cross-entropy, focal loss, etc. This paper reminds me of the KITS19 challenge. Our model didn't perform very well in the challenge. One of the main reasons is that there are serious class imbalances in the images of the KITS dataset. The tumor area usually occupies a small percentage of the entire CT image. This caused our dice accuracy rate to be only about 75%. Let's see if this paper can solve this problem.

2 Methods

The region-based loss are widely used because they are convenient for training deep segmentation networks. While the boundary loss takes the form of a distance metric on the space of contours (or shapes), not regions. It can provide information that is complimentary to regional losses. However, almost no one chooses to use boundary-based losses in a CNN model. Because it is challenging to represent the boundary points corresponding to the regional softmax outputs of a CNN. So, how to make the boundary loss differentiable and implementable to CNN is the main goal of this paper.

First, let's see how region based loss work, for example, cross-entropy loss. When the foreground and background are very imbalanced, and there are few foreground pixels, see Eq.1, the trained model will bias to the background. When splitting the integral region of the cross-entropy into foreground and background, it can be found that the background loss is the main part when back propagating.

$$\begin{aligned} \mathcal{L}_R(\theta) &= \int_\Omega -g(p)\log S_\theta(p)dp + \int_\Omega -(1-g(p))\log\left(1-s_\theta(p)\right)dp \\ &= \int_G -\log S_\theta(p)dp + \int_{\Omega\setminus G} -\log\left(1-s_\theta(p)\right)dp \end{aligned} \tag{1}$$

As for the gradient of the regional loss (CE/DICE), it treats the background and foreground pixels equally, ignoring the spatial information, such as the mis-segmentation far from the groudtruth, which should give a larger penalty weight.

So, for solving this problem, this paper's aim is to minimize the distance between the predicted segmentation boundary and the ground truth boundary, which is denoted as $\mathrm{Dist}\left(\partial G, \partial S_\theta\right)$, with ∂G denoting a representation (the set of points) of the boundary of ground-truth region G.

See Eq. 2, the Dist() function is formed as the L_2 distance of the point on groundtruth's boundary and the corresponding points on predicted boundary.

$$\mathrm{Dist}(\partial G, \partial S) = \int_{\partial G} \left\| y_{\partial S}(p) - p \right\|^2 dp \tag{2}$$

So, what's the definition of 'corresponding'？这里是通过找到grountruth边缘法线方向与预测边缘相交的关联点。

But, this differential loss framework cannot be used directly as a loss. However, let's do a brainstorm, we can express 2 using an integral approach. We can formulate boundary change as a regional integral. See Fig 2.

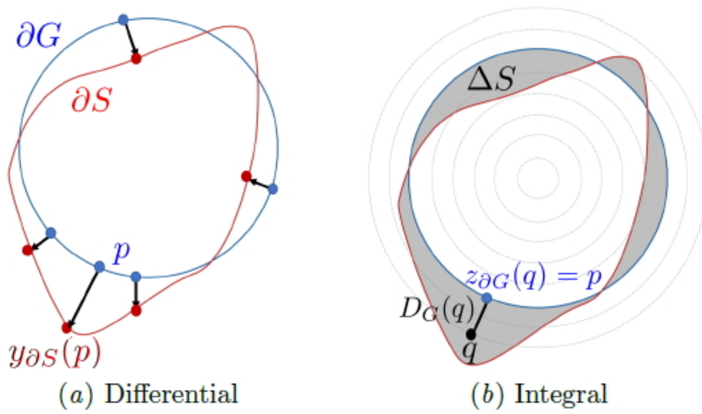

Figure 2: The relationship between differential and integral approaches for evaluating boundary change (variation).

And this goes as

$$\mathrm{Dist}(\partial G, \partial S) = 2 \int_{\Delta S} D_G(q) dq \tag{3}$$

Here, D_G is the distance map with respect to boundary. We have a variable change:

$$\int_{p}^{y_{\partial S}(p)} 2D_G(q)dq = \int_{0}^{\|y_{\partial S}(p)-p\|} 2D_G dD_G = \|y_{\partial S}(p) - p\|^2 \tag{4}$$

$$\frac{1}{2}\mathrm{Dist}(\partial G, \partial S) = \int_S \phi_G(q)dq - \int_G \phi_G(q)dq = \int_\Omega \phi_G(q)s(q)dq - \int_\Omega \phi_G(q)g(q)dq \tag{5}$$

Through this, the non-symmetric L_2 distance between contours in Eq.2 can be expressed as a sum of regional integrals based on a level set representation of boundary ∂G.

When replacing the $s(q)$ with the softmax probability outputs of the network, denoting $s_\theta(q)$, we finally get the boundary loss:

	ISLES		WMH	
Loss	DSC	HD (mm)	DSC	HD (mm)
$\mathcal{L}_B$	0.321 (0.000)	NA	0.569 (0.000)	NA
$\mathcal{L}_{GD}$	0.575 (0.028)	4.009 (0.016)	0.727 (0.006)	1.045 (0.014)
$\mathcal{L}_{GD} + \mathcal{L}_B$	**0.656 (0.023)**	**3.562 (0.009)**	**0.748 (0.005)**	**0.987 (0.010)**

Figure 3: DSC and HD values achieved on the validation subset.

$$\mathcal{L}_B(\theta) = \int_{\Omega} \phi_G(q) s_\theta(q) dq \tag{6}$$

Here, the second part of Eq 5 is omitted because it is independent of network parameters.

In the experiments, authors use the boundary loss in conjunction with the regional generalized Dice loss:

$$\alpha \mathcal{L}_{GD}(\theta) + (1 - \alpha)\mathcal{L}_B(\theta) \tag{7}$$

3 Experiment

The model is implemented on ISLES and WMH datasets. Authors choose U-Net as the training model. The α in Eq. 6 is initially set to 1, and decreased by 0.01 after each epoch, following a simple scheduling strategy, until it reached the value of 0.01. In this way, authors give more importance to the regional loss term at the beginning while gradually increasing the impact of the boundary loss term.

The quantitative evaluation results on ISLES and WHM can be referred in Fig 3. We can find that only using boundary loss is unreasonable.

From Fig 4, we can also observe that the boundary loss term helps stabilizing the training process, yielding a much smoother curve as the network training converges.

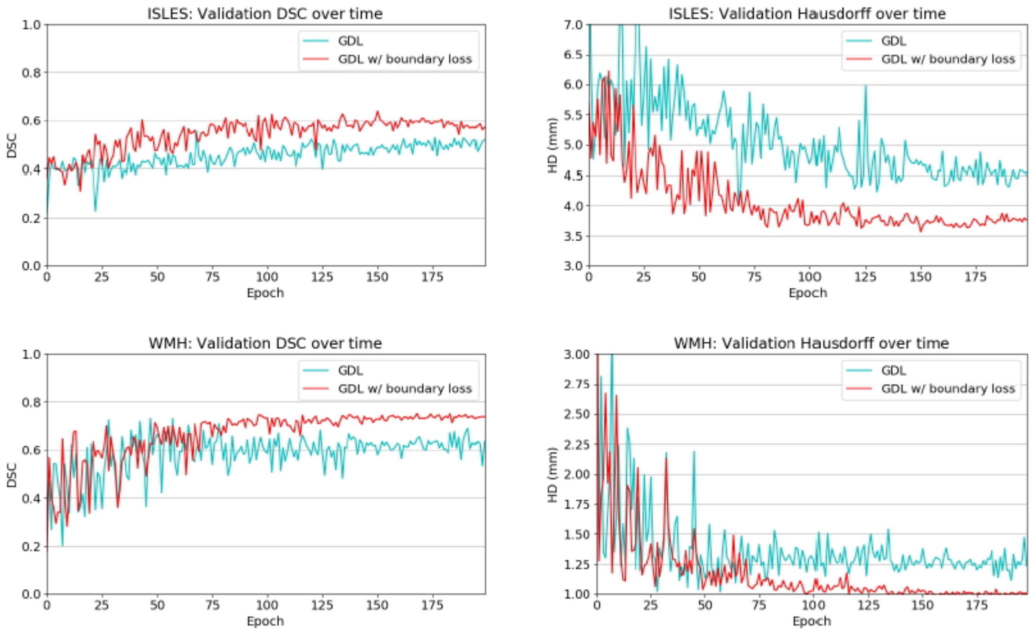

Figure 4: Evolution of DSC and HD values.

4 My thoughts

The proposed boundary loss is meant to address segmentation label imbalances. However, the effect of boundary loss in the field of medical image segmentation is worth validation and discussion. According to the presentation on MIDL19，the choice of dataset has a great impact on the loss proposed in this paper.