# Boundary loss for highly unbalanced segmentation

**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*

**Editors:** Under Review for MIDL 2019

arXiv:1812.07032v1 [cs.CV] 17 Dec 2018

## Abstract

Widely used loss functions for convolutional neural network (CNN) segmentation, e.g., Dice or cross-entropy, are based on integrals (summations) over the segmentation regions. Unfortunately, it is quite common in medical image analysis to have highly unbalanced segmentations, where standard losses contain regional terms with values that differ considerably – typically of several orders of magnitude – across segmentation classes, which may affect training performance and stability. The purpose of this study is to build a *boundary* loss, which takes the form of a distance metric on the space of contours (or shapes), not regions. We argue that a boundary loss can mitigate the difficulties of regional losses in the context of highly unbalanced segmentation problems because it uses integrals over the boundary (interface) between regions instead of unbalanced integrals over regions. Furthermore, a boundary loss provides information that is complimentary to regional losses. Unfortunately, it is not straightforward to represent the boundary points corresponding to the regional softmax outputs of a CNN. Our boundary loss is inspired by discrete (graph-based) optimization techniques for computing gradient flows of curve evolution. Following an integral approach for computing boundary variations, we express a non-symmetric $L_2$ distance on the space of shapes as a regional integral, which avoids completely local differential computations involving contour points. Our boundary loss is the sum of linear functions of the regional softmax probability outputs of the network. Therefore, it can easily be combined with standard regional losses and implemented with any existing deep network architecture for N-D segmentation.

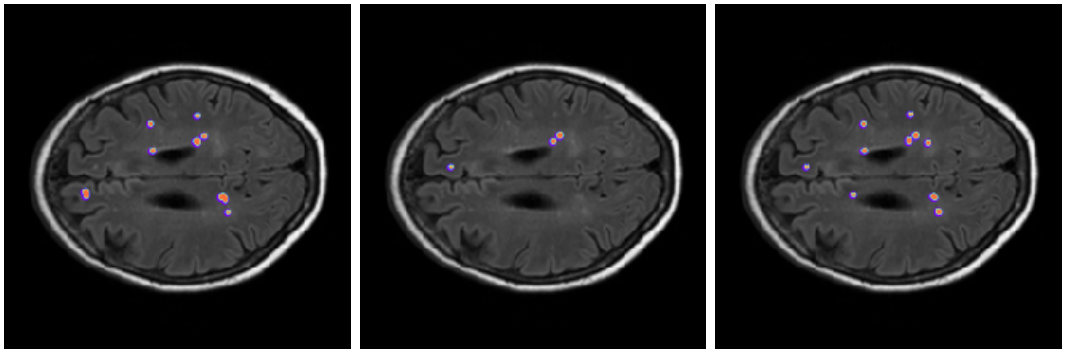
Our boundary loss has been validated on two benchmark datasets corresponding to difficult, highly unbalanced segmentation problems: the ischemic stroke lesion (ISLES) and white matter hyperintensities (WMH). Used in conjunction with the region-based generalized Dice loss (GDL), our boundary loss improves performance significantly compared to GDL alone, reaching up to 8% improvement in Dice score and 10% improvement in Hausdorff score. It also yielded a more stable learning process. Our code is publicly available[1].

**Keywords:** Boundary loss, unbalanced data, semantic segmentation, deep learning

[*] Contributed equally

1. https://github.com/LIVIAETS/surface-loss

© 2019 H. Kervadec, J. Bouchtiba, C. Desrosiers, E. Granger, J. Dolz & I.B. Ayed.



($a$) Ground truth ($b$) GDL ($c$) GDL w/ boundary loss

Figure 1: A visual comparison that shows the positive effect of our boundary loss on a validation data from the WMH dataset. Our boundary loss helped recovering small regions that were otherwise missed by the generalized Dice loss (GDL). Best viewed in colors.

## 1. Introduction

Recent years have witnessed a substantial growth in the number of deep learning methods for medical image segmentation (Litjens et al., 2017; Shen et al., 2017; Dolz et al., 2018; Ker et al., 2018). Widely used loss functions for segmentation, e.g., Dice or cross-entropy, are based on *regional* integrals, which are convenient for training deep neural networks. In practice, these regional integrals are summations over the segmentation regions of differentiable functions, each invoking directly the softmax probability outputs of the network. Therefore, standard stochastic optimizers such SGD are directly applicable. Unfortunately, difficulty occurs for highly unbalanced segmentations, for instance, when the size of target foreground region is several orders of magnitude less than the background size. For example, in the characterization of white matter hyperintensities (WMH) of presumed vascular origin, the foreground composed of WMH regions may be 500 times smaller than the background (see the typical example in Fig. 1). In such cases, quite common in medical image analysis, common regional losses contain foreground and background terms that have substantial differences in their values, typically of several orders of magnitude, which might affect performance and training stability (Milletari et al., 2016; Sudre et al., 2017).

Segmentation approaches based on convolutional neural networks (CNN) are typically trained by minimizing the cross-entropy (CE), which measures an affinity between the regions defined by probability softmax outputs of the network and the corresponding ground-truth regions. The standard regional CE has well-known drawbacks in the context of highly unbalanced problems. It assumes identical importance distribution of all the samples and classes. To achieve good generalization, it requires a large training set with balanced classes. For unbalanced data, CE typically results in unstable training and leads to decision boundaries biased towards the majority classes. Class-imbalanced learning aims at mitigating learning bias by promoting the importance of infrequent labels. In medical image segmentation, a common strategy is to re-balance class prior distributions by down-sampling frequent labels (Havaei et al., 2017; Valverde et al., 2017). Nevertheless, this strategy limits

the information of the images used for training. Another common practice is to assign weights to the different classes, inversely proportional to the frequency of the corresponding labels (Brosch et al., 2015; Ronneberger et al., 2015; Kamnitsas et al., 2017; Long et al., 2015; Yu et al., 2017). In this scenario, the standard cross-entropy (CE) loss is modified so to assign more importance to rare labels than to common ones. Although effective for some unbalanced problems, such weighting methods may undergo serious difficulties when dealing with highly unbalanced datasets, as seen with WMH segmentation. The CE gradient computed over the few pixels of infrequent labels is typically noisy, and amplifying this noise with a high class weight may lead to instability.

The well-known Dice overlap coefficient was also adopted as a regional loss function, typically outperforming CE in unbalanced medical image segmentation problems (Milletari et al., 2016, 2017; Wong et al., 2018). Sudre et al. (Sudre et al., 2017) generalized the Dice loss (Milletari et al., 2016), by weighting according to the inverse of class label frequency. Despite these improvements over CE (Milletari et al., 2016; Sudre et al., 2017), regional Dice losses may undergo difficulties when dealing with very small structures. In such highly unbalanced scenarios, mis-classified pixels may lead to large decreases of the loss, resulting in unstable optimization. Furthermore, Dice corresponds to the harmonic mean between precision and recall, implicitly using the arithmetic mean of false positives and false negatives. False positives and false negatives are therefore equally important when the true positives remain the same, making this loss mainly appropriate when both types of errors are equally bad. The recent research in (Salehi et al., 2017; Abraham and Khan, 2018) investigated losses based on the Tversky similarity index in order to provide a better trade-off between precision and recall. It introduces two parameters that control the importance of false positives and false negatives. Other recent advances in class-imbalanced learning for computer vision problems have been adopted in medical image segmentation. For example, inspired by the concept of focal loss (Lin et al., 2018), Dice and Tvserky losses have been extended to integrate a focal term, which is parameterized by a value that controls the importance between easy and hard training samples (Abraham and Khan, 2018; Wong et al., 2018). The main objective of these losses is to balance the classes not only by their relative class sizes, but also by the level of segmentation difficulty.

### Contributions

All the above-mentioned losses are *region-based*. In this paper, we propose a *boundary* loss that takes the form of a distance metric on the space of contours (or shapes), not regions. We argue that a boundary loss can mitigate the issues related to regional losses in highly unbalanced segmentation problems. Rather than using unbalanced integrals over the regions, a boundary loss uses integrals over the boundary (interface) between the regions. Furthermore, a boundary loss provides information that is complimentary to regional losses. It is, however, challenging to represent the boundary points corresponding to the regional softmax outputs of a CNN. This difficulty may explain why boundary losses have been avoided in the context of deep segmentation networks. Our boundary loss is inspired by techniques in discrete (graph-based) optimization for computing gradient flows of curve evolution (Boykov et al., 2006). Following an integral approach for computing boundary variations, we express a non-symmetric $L_2$ distance on the space of shapes as a regional

integral, which avoids completely local differential computations involving contour points. Our boundary loss is the sum of linear functions of the regional softmax probability outputs of the network. Therefore, it can be easily combined with standard regional losses and implemented with any existing deep network architecture for N-D segmentation.

We evaluated our boundary loss in conjunction with the region-based generalized Dice loss (GDL) (Sudre et al., 2017) on two challenging and highly unbalanced segmentation problems – the Ischemic Stroke Lesion (ISLES) and the White Matter Hyperintensities (WMH) benchmark datasets. Results indicate that the proposed boundary loss provides a more stable learning process, and significantly outperforms GDL alone, yielding up to 8% improvement in Dice score, and 10% improvement in Hausdorff score.

## 2. Formulation

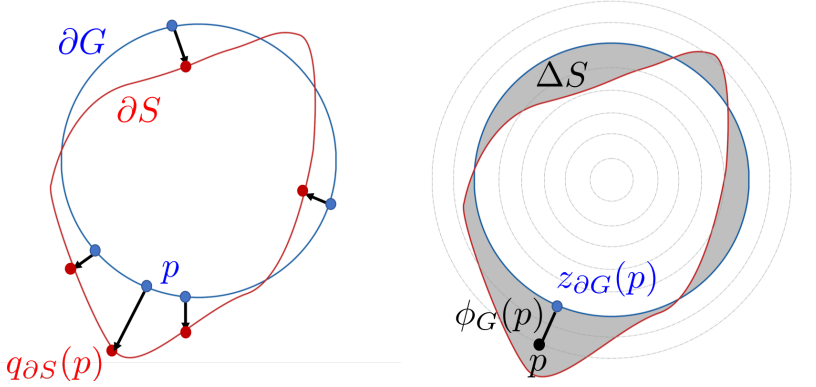
Let $I : \Omega \subset \mathbb{R}^{2,3} \to \mathbb{R}$ denotes a training image with spatial domain $\Omega$, and $g : \Omega \to \{0, 1\}$ a binary ground-truth segmentation of the image: $g(p) = 1$ if pixel/voxel $p$ belongs to the target region $G \subset \Omega$ (foreground region) and 0 otherwise, i.e., $p \in \Omega \setminus G$ (background region)[2]. Let $s_\theta : \Omega \to [0, 1]$ denotes the softmax probability output of a deep segmentation network, and $S_\theta \subset \Omega$ the corresponding segmentation region: $S_\theta = \{p \in \Omega \,|\, s_\theta(p) \geq \delta\}$ for some threshold $\delta$. Widely used segmentation loss functions involve a *regional integral* for each segmentation region in $\Omega$, which measures some similarity (or overlap) between the region defined by the probability outputs of the network and the corresponding ground-truth. In the two-region case, we have an integral of the general form $\int_\Omega g(p) f(s_\theta(p)) dp$ for the foreground, and of the form $\int_\Omega (1-g(p)) f(1-s_\theta(p)) dp$ for the background. For instance, the standard two-region cross-entropy loss corresponds to a summation of these two terms for $f = -\log(\cdot)$. Similarly, the generalized Dice loss (GDL) (Sudre et al., 2017) involves regional integrals with $f = 1$, subject to some normalization, and is given as follows for the two-region case:

$$\mathcal{L}_{GD}(\theta) = 1 - 2\frac{w_G \int_{p\in\Omega} g(p)s_\theta(p)dp + w_B \int_{p\in\Omega}(1 - g(p))(1 - s_\theta(p))dp}{w_G \int_\Omega [s_\theta(p) + g(p)]dp + w_B \int_\Omega [2 - s_\theta(p) - g(p)]dp} \tag{1}$$

where coefficients $w_G = 1/\left(\int_{p\in\Omega} g(p)dp\right)^2$ and $w_B = 1/\left(\int_\Omega (1 - g(p))dp\right)^2$ are introduced to reduce the well-known correlation between the Dice overlap and region size.

Regional integrals are widely used because they are convenient for training deep segmentation networks. In practice, these regional integrals are summations of differentiable functions, each invoking directly the softmax probability outputs of the network, $s_\theta(p)$. Therefore, standard stochastic optimizers such SGD are directly applicable. Unfortunately, extremely unbalanced segmentations are quite common in medical image analysis, where, e.g., the size of the target foreground region is several orders of magnitude smaller than the background size. This represents challenging cases because the foreground and background terms have substantial differences in their values, which affects segmentation performance and training stability (Milletari et al., 2016; Sudre et al., 2017).

2. We focus on two-region segmentation to simplify the presentation. However, our formulation extends to the multi-region case in a straightforward manner.




($a$) Differential ($b$) Integral

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

Our purpose is to build a boundary loss $\text{Dist}(\partial G, \partial S_\theta)$, which takes the form of a distance metric on the space of contours (or region boundaries) in $\Omega$, with $\partial G$ denoting a representation of the boundary of ground-truth region $G$ (e.g., the set of points of $G$, which have a spatial neighbor in background $\Omega \setminus G$) and $\partial S_\theta$ denoting the boundary of the segmentation region defined by the network output. On the one hand, a boundary loss should be able to mitigate the above-mentioned difficulties for unbalanced segmentations: rather than using unbalanced integrals within the regions, it uses integrals over the boundary (interface) between the regions. Furthermore, a boundary loss provides information that is different from and, therefore, complimentary to regional losses. On the other hand, it is not clear how to represent boundary points on $\partial S_\theta$ as a differentiable function of regional network outputs $s_\theta$. This difficulty might explain why boundary losses have been, to the best of our knowledge, completely avoided in the context of deep segmentation networks.

Our boundary loss is inspired from discrete (graph-based) optimization techniques for computing gradient flows of curve evolution (Boykov et al., 2006). Similarly to our problem, curve evolution methods require a measure for evaluating boundary changes (or variations). Consider the following non-symmetric $L_2$ distance on the space of shapes, which evaluates the change between two nearby boundaries $\partial S$ and $\partial G$ (Boykov et al., 2006):

$$\text{Dist}(\partial G, \partial S) = \int_{\partial G} \|q_{\partial S}(p) - p\|^2 dp \tag{2}$$

where $p \in \Omega$ is a point on boundary $\partial G$ and $q_{\partial S}(p)$ denotes the corresponding point on boundary $\partial S$, along the direction normal to $\partial G$, i.e., $q_{\partial S}(p)$ is the intersection of $\partial S$ and the line that is normal to $\partial G$ at $p$ (See Fig. 2.a for an illustration). In fact, this *differential* framework for evaluating boundary change is in line with standard variational curve evolution methods (Mitiche and Ben Ayed, 2011), which compute the motion of each point $p$ on the evolving curve as a velocity along the normal to the curve at point $p$. Similarly to any contour distance invoking directly points on contour $\partial S$, expression (2) cannot be used

directly as a loss for $\partial S = \partial S_\theta$. However, it is easy to show that the differential boundary variation in (2) can be expressed using an *integral* approach (Boykov et al., 2006), which avoids completely local differential computations involving contour points and represent boundary change as a regional integral:

$$\text{Dist}(\partial G, \partial S) = 2 \int_{\Delta S} \phi_G(p) dp \tag{3}$$

where $\Delta S$ denotes the region between the two contours and $\phi_G : \Omega \to \mathbb{R}$ is a *level set* representation of boundary $\partial G$: $\phi_G(p)$ evaluates a signed distance between point $p \in \Omega$ and the nearest point $z_{\partial G}(p)$ on contour $\partial G$: $\phi_G(p) = -\|p - z_{\partial G}(p)\|$ if $p \in G$ and $\phi_G(p) = \|p - z_{\partial G}(p)\|$ otherwise. Fig. 2.b illustrates this integral framework for evaluating the boundary distance in (2). To show that (3) holds, it suffices to notice that integrating the distance function $2\phi_G(p)$ over the normal segment connecting $p$ and $q_{\partial S}(p)$ yields $\|q_{\partial S}(p) - p\|^2$. Thus, the non-symmetric $L_2$ distance between contours in Eq. (2) can be expressed as a sum of regional integrals:

$$\int_S \phi_G(p) dp - \int_G \phi_G(p) dp = \int_\Omega \phi_G(p) s(p) dp - \int_\Omega \phi_G(p) g(p) dp \tag{4}$$

where $s : \Omega \to \{0, 1\}$ is binary indicator function of region $S$: $s(p) = 1$ if $p \in S$ belongs to the target and 0 otherwise. Now, for $S = S_\theta$, i.e., replacing binary variables $s(p)$ in Eq. (4) by the softmax probability outputs of the network $s_\theta(p)$, we obtain the following boundary loss which, up to a constant independent of $\theta$, approximates boundary distance $\text{Dist}(\partial G, \partial S_\theta)$:

$$\mathcal{L}_B(\theta) = \int_\Omega \phi_G(p) s_\theta(p) dp \tag{5}$$

Notice that we omitted the last term in Eq. (4) as it is independent of network parameters. The level set function $\phi_G$ is pre-computed directly from the ground-truth region $G$. In practice, our boundary loss in Eq. (5) is the sum of linear functions of the regional softmax probability outputs of the network. Therefore, it can be easily combined with standard regional losses and implemented with any existing deep network architecture for N-D segmentation. In the experiments, we will use our boundary loss in conjunction with the regional generalized Dice loss:

$$\alpha \mathcal{L}_{GD}(\theta) + (1 - \alpha) \mathcal{L}_B(\theta) \tag{6}$$

Finally, it is worth noting that our boundary loss uses ground-truth boundary information via pre-computed level-set function $\phi_G(p)$, which encodes the distance between each point $p$ and $\partial G$. In Eq. (5), the softmax for each point $p$ is weighted by the distance function. Such distance-to-boundary information is omitted in widely used regional losses, where all the points within a given region are treated equally, independently of their distances from the boundary.

## 3. Experiments

### 3.1. Datasets

To evaluate the proposed boundary loss, we selected two challenging brain lesion segmentation tasks, defined by highly unbalanced classes.

**ISLES** The training dataset provided by the ISLES organizers is composed of 94 ischemic stroke lesion multi-modal scans. In our experiments, we split this dataset into training and validation sets containing 74 and 20 examples, respectively. Each scan contains Diffusion maps (DWI) and Perfusion maps (CBF, MTT, CBV, Tmax and CTP source data), as well as the manual ground-truth segmentation. More details can be found in the ISLES website[3].

**WMH** The public dataset of the White Matter Hyperintensities (WMH) [4] MICCAI 2017 challenge contains 60 3D T1-weighted scans and 2D multi-slice FLAIR acquired from multiple vendors and scanners in three different hospitals. In addition, the ground truth for the 60 scans is provided. From the whole set, 50 scans were used for training, and the remaining 10 for validation.

### 3.2. Implementation details

**Data pre-processing.** While the scans are provided as 3D images, we process them as a stack of independent 2D images, which are fed into the network. In fact, the scans in some datasets, such as ISLES, contain between 2 and 16 slices, making them ill-suited for 3D convolutions in those cases. The scans were normalized between 0 and 1 before being saved as a set of 2D matrices, and re-scaled to 256×256 pixels if needed. When several modalities were available, all of them were concatenated before being used as input to the network. We did not use any data augmentation in our experiments.

**Architecture and training.** We employed UNet (Ronneberger et al., 2015) as deep learning architecture in our experiments. To train our model, we employed Adam optimizer, with a learning rate of 0.001 and a batch size equal to 8. The learning rate is halved if the validation performances do not improve during 20 epochs. We did not use early stopping.

To compute the level set function $\phi_G$ in Eq. (5), we used standard SciPy functions [5]. Note that, for slices containing only the background region, we used a zero-distance map, assuming that the GDL is sufficient in those cases. Furthermore, during training, the value of $\alpha$ in Eq. (6) was 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, we give more importance to the regional loss term at the beginning while gradually increasing the impact of the boundary loss term.

For our implementation, we used PyTorch (Paszke et al., 2017), and ran the experiments on a machine equipped with an NVIDIA GTX 1080 Ti GPU with 11GBs of memory. Our code (data pre-processing, training and testing scripts) is publicly available[6].

**Evaluation.** For evaluation purposes, we employ the common Dice Similarity Coefficient (DSC) and Hausdorff Distance (HD) metrics.

3. http://www.isles-challenge.org
4. http://wmh.isi.uu.nl
5. https://docs.scipy.org/doc/scipy-0.14.0/reference/generated/scipy.ndimage.morphology.distance_transform_edt.html
6. https://github.com/LIVIAETS/surface-loss

### 3.3. Results

**Quantative evaluation.** Table 1 reports the DSC and HD performance for our experiments using GDL alone and the loss we proposed in Eq. (6) on the ISLES and WMH datasets. By adding our boundary loss term to the GDL consistently improves segmentation performance, which is reflected in significantly higher DSC and HD values. While this boundary loss term brings a DSC improvement of around 2% on the WMH dataset, it achieves 8% better DSC on the ISLES segmentation task. The same trend is observed with the HD metric, where the gain is larger on the ISLES dataset than on WMH.

| | ISLES | | WMH | |
|---|---|---|---|---|
| Loss | DSC | HD (mm) | DSC | HD (mm) |
| $\mathcal{L}_{GD}$ | 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)** |

Table 1: DSC and HD values achieved on the validation subset. The values represent the mean performance (and standard deviation) of 2 runs for each setting.

The learning curves depicted in Figure 3 show the gap in performance of the GDL with respect to the GDL with boundary loss, where the difference becomes significant at convergence. In addition to outperforming GDL, we can also observe that the boundary loss term helps stabilizing the training process, yielding a much smoother curve as the network

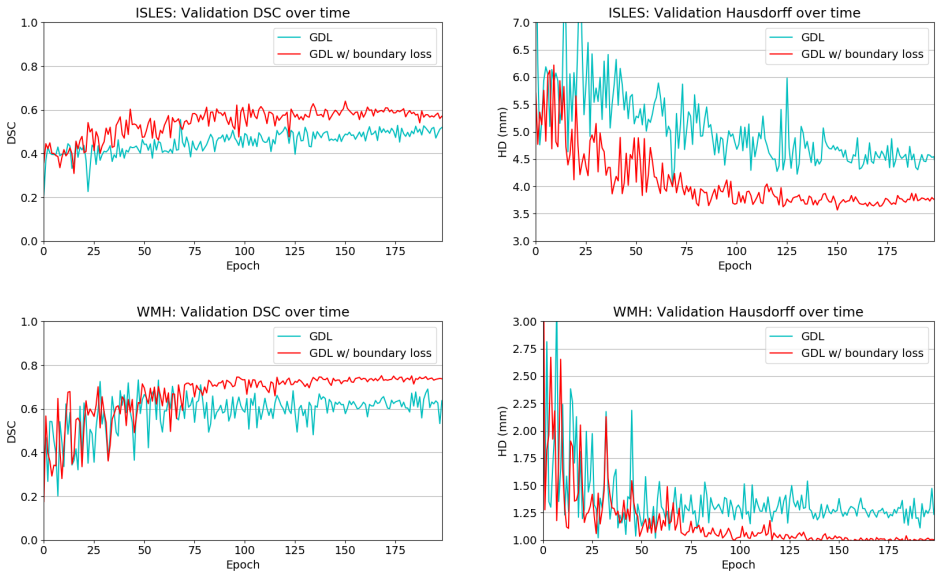



Figure 3: Evolution of DSC and HD values on the validation subset when training on ISLES and WMH dataset. The blue curve shows the performance of the network trained using the GDL loss, while the red curve represents the optimization process with the GDL + our proposed boundary loss term.

training converges. This behaviour is consistent for both metrics and both dataset, which clearly shows the benefits of employing the proposed boundary loss term.

**Qualitative evaluation.** Qualitative results are depicted in Fig. 4. Inspecting these results visually, we can observe that there are two major types of improvements when employing the proposed boundary loss. First, as the methods based on DSC losses, such as GDL, do not use spatial information, prediction errors are treated equally. This means that the errors for pixels/voxels in an already detected object have the same importance as the errors produced in completely missed objects. On the contrary, as our boundary loss is based on the distance map from the ground-truth boundary $\partial G$, it will penalize much more such cases, helping to recover small and far regions. This effect is best illustrated in Fig. 1 and Fig. 4 (third row). False positives (first row in Fig. 4) will be far away from the closest foreground, getting a much higher penalty than with the GDL alone. This helps in reducing the number of false positives.

**Computational complexity.** It is worth mentioning that, as the proposed boundary loss term involves an element-wise product between two matrices – i.e., the pre-computed level-set function $\phi_G$ and the softmax output $s_\theta(p)$ – the complexity that it adds is negligible.

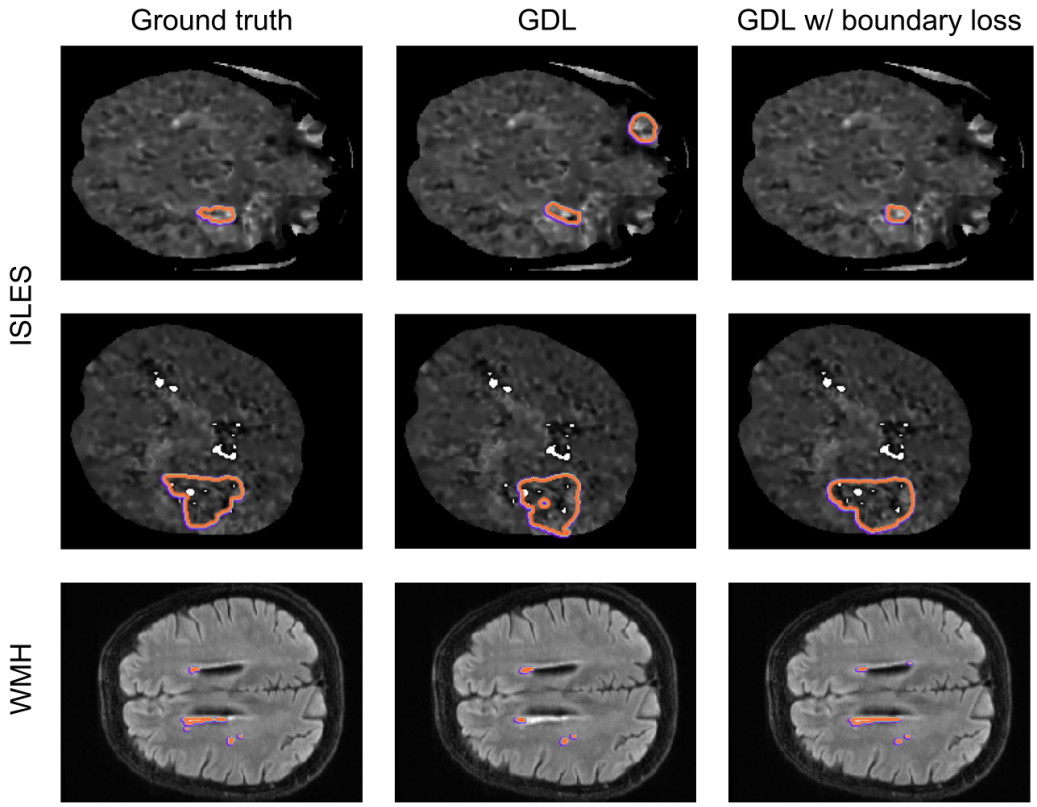



Figure 4: Visual comparison on two different datasets from the validation set.

## 4. Conclusion

We proposed a boundary loss term that can be easily combined with standard regional losses to tackle the segmentation task in highly unbalanced scenarios. Furthermore, the proposed term can be implemented in any existing deep network architecture and for any N-D segmentation problem. Our experiments on two challenging and highly unbalanced datasets demonstrated the effectiveness of including the proposed boundary loss term during training. It consistently improved the performance, with a large margin on one data set, and enhanced training stability. Even though we limited the experiments to 2-D segmentation problems, the proposed framework can be trivially extended to 3-D, which could further improve the performance of deep networks, as more context is analyzed.

## Acknowledgments

This work is supported by the National Science and Engineering Research Council of Canada (NSERC), discovery grant program, and by the ETS Research Chair on Artificial Intelligence in Medical Imaging.

## References

Nabila Abraham and Naimul Mefraz Khan. A novel focal tversky loss function with improved attention U-Net for lesion segmentation. *arXiv preprint arXiv:1810.07842*, 2018.

Yuri Boykov, Vladimir Kolmogorov, Daniel Cremers, and Andrew Delong. An integral solution to surface evolution PDEs via geo-cuts. In *European Conference on Computer Vision*, pages 409–422. Springer, 2006.

Tom Brosch, Youngjin Yoo, Lisa YW Tang, David KB Li, Anthony Traboulsee, and Roger Tam. Deep convolutional encoder networks for multiple sclerosis lesion segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 3–11. Springer, 2015.

Jose Dolz, Christian Desrosiers, and Ismail Ben Ayed. 3D fully convolutional networks for subcortical segmentation in MRI: A large-scale study. *NeuroImage*, 170:456–470, 2018.

Mohammad Havaei, Axel Davy, David Warde-Farley, Antoine Biard, Aaron Courville, Yoshua Bengio, Chris Pal, Pierre-Marc Jodoin, and Hugo Larochelle. Brain tumor segmentation with deep neural networks. *Medical image analysis*, 35:18–31, 2017.

Konstantinos Kamnitsas, Christian Ledig, Virginia FJ Newcombe, Joanna P Simpson, Andrew D Kane, David K Menon, Daniel Rueckert, and Ben Glocker. Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. *Medical image analysis*, 36:61–78, 2017.

Justin Ker, Lipo Wang, Jai Rao, and Tchoyoson Lim. Deep learning applications in medical image analysis. *IEEE Access*, 6:9375–9389, 2018.

Tsung-Yi Lin, Priyal Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. *IEEE transactions on pattern analysis and machine intelligence*, 2018.

Geert Litjens, Thijs Kooi, Babak Ehteshami Bejnordi, Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Mohsen Ghafoorian, Jeroen Awm Van Der Laak, Bram Van Ginneken, and Clara I Sánchez. A survey on deep learning in medical image analysis. *Medical image analysis*, 42:60–88, 2017.

Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440, 2015.

Fausto Milletari, Nassir Navab, and Seyed-Ahmad Ahmadi. V-Net: Fully convolutional neural networks for volumetric medical image segmentation. In *3D Vision (3DV), 2016 Fourth International Conference on*, pages 565–571. IEEE, 2016.

Fausto Milletari, Seyed-Ahmad Ahmadi, Christine Kroll, Annika Plate, Verena Rozanski, Juliana Maiostre, Johannes Levin, Olaf Dietrich, Birgit Ertl-Wagner, Kai Bötzel, et al. Hough-CNN: deep learning for segmentation of deep brain regions in MRI and ultrasound. *Computer Vision and Image Understanding*, 164:92–102, 2017.

Amar Mitiche and Ismail Ben Ayed. *Variational and level set methods in image segmentation*. Springer, 2011.

Adam Paszke, Sam Gross, Soumith Chintala, Gregory Chanan, Edward Yang, Zachary DeVito, Zeming Lin, Alban Desmaison, Luca Antiga, and Adam Lerer. Automatic differentiation in pytorch. 2017.

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.

Seyed Sadegh Mohseni Salehi, Deniz Erdogmus, and Ali Gholipour. Tversky loss function for image segmentation using 3D fully convolutional deep networks. In *International Workshop on Machine Learning in Medical Imaging*, pages 379–387. Springer, 2017.

Dinggang Shen, Guorong Wu, and Heung-Il Suk. Deep learning in medical image analysis. *Annual review of biomedical engineering*, 19:221–248, 2017.

Carole H Sudre, Wenqi Li, Tom Vercauteren, Sebastien Ourselin, and M Jorge Cardoso. Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, pages 240–248. Springer, 2017.

Sergi Valverde, Mariano Cabezas, Eloy Roura, Sandra González-Villà, Deborah Pareto, Joan C Vilanova, Lluis Ramio-Torrenta, Àlex Rovira, Arnau Oliver, and Xavier Lladó. Improving automated multiple sclerosis lesion segmentation with a cascaded 3D convolutional neural network approach. *NeuroImage*, 155:159–168, 2017.

Ken CL Wong, Mehdi Moradi, Hui Tang, and Tanveer Syeda-Mahmood. 3D segmentation with exponential logarithmic loss for highly unbalanced object sizes. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 612–619. Springer, 2018.

Lequan Yu, Xin Yang, Hao Chen, Jing Qin, and Pheng-Ann Heng. Volumetric convnets with mixed residual connections for automated prostate segmentation from 3D MR images. In *AAAI*, pages 66–72, 2017.