# Boundary loss for highly unbalanced segmentation

MIDL 2019, London

Hoel Kervadec*, Jihene Bouchtiba*, Christian Desrosiers, Éric Granger, Jose Dolz, Ismail Ben Ayed

ÉTS Montréal
*Contributed equally
hoel.kervadec.1@etsmtl.net
https://github.com/LIVIAETS/surface-loss

# Presentation overview

- Difficulty of highly-imbalanced segmentation

# Presentation overview

- Difficulty of highly-imbalanced segmentation
- Formulation: inspired from curve/level-set evolution/differentiation

# Presentation overview

- Difficulty of highly-imbalanced segmentation
- Formulation: inspired from curve/level-set evolution/differentiation
- Evaluation

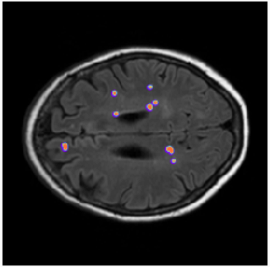
# Presentation overview

- Difficulty of highly-imbalanced segmentation
- Formulation: inspired from curve/level-set evolution/differentiation
- Evaluation
- Future works on other applications

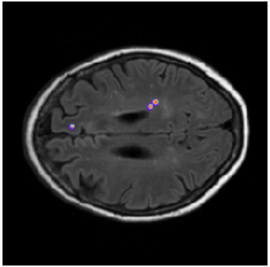
# Problem statement & motivations

# Introduction

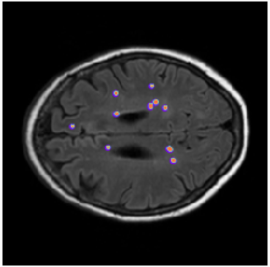
Regional losses (cross-entropy, dice loss, …) are sensitive to high class-imbalance



Ground truth



**Regional** loss



Our **boundary** loss

# Motivations

$$\mathcal{L}_R(\theta) = \int_\Omega -g(p)\log s_\theta(p)dp + \int_\Omega -(1-g(p))\log(1-s_\theta(p))dp$$

$\Omega$: image set, $\theta$: network parameters, $S_\theta$: softmax predictions, $G$: ground truth.

# Motivations

$$\mathcal{L}_R(\theta) = \int_{\Omega} -g(p) \log s_\theta(p) dp + \int_{\Omega} -(1 - g(p)) \log (1 - s_\theta(p)) dp$$

$\Omega$: image set, $\theta$: network parameters, $S_\theta$: softmax predictions, $G$: ground truth.

# Motivations

$$\mathcal{L}_R(\theta) = \int_{\Omega} -g(p) \log s_\theta(p) dp + \int_{\Omega} -(1 - g(p)) \log (1 - s_\theta(p)) dp$$

$\Omega$: image set, $\theta$: network parameters, $S_\theta$: softmax predictions, $G$: ground truth.

# Motivations

$$\mathcal{L}_R(\theta) = \int_{\Omega} -g(p) \log s_\theta(p) dp + \int_{\Omega} -(1 - g(p)) \log(1 - s_\theta(p)) dp$$

$\Omega$: image set, $\theta$: network parameters, $S_\theta$: softmax predictions, $G$: ground truth.

# Motivations

$$\mathcal{L}_R(\theta) = \int_\Omega -g(p) \log s_\theta(p) dp + \int_\Omega -(1 - g(p)) \log (1 - s_\theta(p)) dp$$

$\Omega$: image set, $\theta$: network parameters, $S_\theta$: softmax predictions, $G$: ground truth.

# Motivations

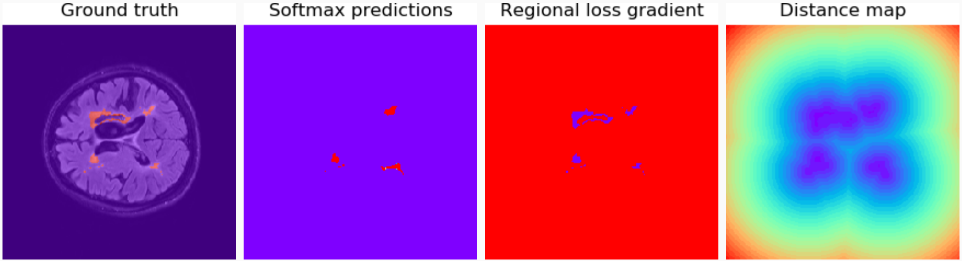
$$
\begin{aligned}
\mathcal{L}_R(\theta) &= \int_{\Omega} -g(p) \log s_\theta(p) dp + \int_{\Omega} -(1 - g(p)) \log (1 - s_\theta(p)) dp \\
&= \int_{G} - \log s_\theta(p) dp + \int_{\Omega \setminus G} - \log (1 - s_\theta(p)) dp
\end{aligned}
$$

$\Omega$: image set, $\theta$: network parameters, $S_\theta$: softmax predictions, $G$: ground truth.

Several orders of magnitude of difference.

# Motivations

Regional losses discard spatial/boundary information:

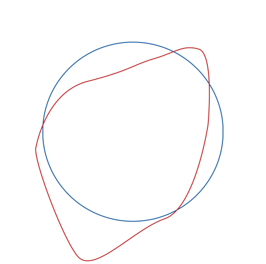



$G$ $\qquad$ $S_{\boldsymbol{\theta}}$ $\qquad$ $\frac{\partial \mathcal{L}_R(\theta)}{\partial S_\theta}$ $\qquad$ $\frac{\partial \mathcal{L}_B(\theta)}{\partial S_\theta}$

## A boundary loss?

A loss minimizing the distance between two boundaries could:

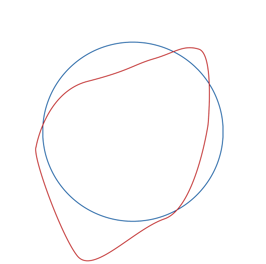
- Use that extra spatial information
- Be less sensitive to the class imbalance



# A boundary loss?

A loss minimizing the distance between two boundaries could:

- Use that extra spatial information
- Be less sensitive to the class imbalance



But challenging to represent the boundary of $S_\theta$

# Formulation

# Formulation

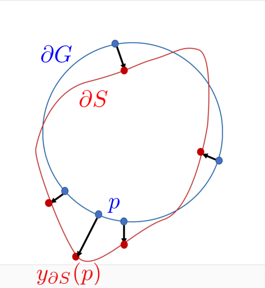
Inspired by discrete optimization techniques in curve evolution.

# Formulation

Inspired by discrete optimization techniques in curve evolution.

When two boundaries are close enough [Boykov et al., 2006], their distance can be written as:

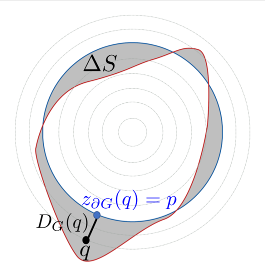
$$\mathrm{Dist}(\partial G, \partial S) \approx \int_{\partial G} \|y_{\partial S}(p) - p\|^2 dp$$



# Formulation

Switch to an integral over $\Delta S$

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




$D_G(q)$: distance map corresponding to $\partial G$

# Formulation

$$\frac{1}{2}\mathrm{Dist}(\partial G, \partial S) \quad \approx \int_{\Delta S} D_G(q)dq$$

# Formulation

$$\begin{aligned} \tfrac{1}{2}\mathrm{Dist}(\partial G, \partial S) &\approx \int_{\Delta S} D_G(q)dq \\ \Delta S &= (S \setminus G) \cup (G \setminus S) \end{aligned}$$

# Formulation

$$
\begin{aligned}
\frac{1}{2}\mathrm{Dist}(\partial G, \partial S) &\approx \int_{\Delta S} D_G(q)dq \\
\Delta S &= (S \setminus G) \cup (G \setminus S) \\
\phi_G(q) &= \begin{cases} -D_G(q) & \text{if } q \in G \\ D_G(q) & \text{otherwise} \end{cases}
\end{aligned}
$$

# Formulation

$$\begin{cases} \frac{1}{2}\mathrm{Dist}(\partial G, \partial S) & \approx \int_{\Delta S} D_G(q)dq \\ \Delta S & = (S \setminus G) \cup (G \setminus S) \\ \phi_G(q) & = \begin{cases} -D_G(q) & \text{if } q \in G \\ D_G(q) & \text{otherwise} \end{cases} \end{cases}$$

Can rewrite $\mathrm{Dist}(\partial G, \partial S)$ as regional integrals of level-set functions:

$$\frac{1}{2}\mathrm{Dist}(\partial G, \partial S) \approx \int_S \phi_G(q)dq - \int_G \phi_G(q)dq$$

# Formulation

$$\begin{cases} \frac{1}{2}\text{Dist}(\partial G, \partial S) & \approx \int_{\Delta S} D_G(q)dq \\ \Delta S & = (S \setminus G) \cup (G \setminus S) \\ \phi_G(q) & = \begin{cases} -D_G(q) & \text{if } q \in G \\ D_G(q) & \text{otherwise} \end{cases} \end{cases}$$

Can rewrite $\text{Dist}(\partial G, \partial S)$ as regional integrals of level-set functions:

$$\begin{aligned} \frac{1}{2}\text{Dist}(\partial G, \partial S) &\approx \int_S \phi_G(q)dq - \int_G \phi_G(q)dq \\ &\approx \int_\Omega \phi_G(q)s(q)dq - \int_\Omega \phi_G(q)g(q)dq \end{aligned}$$

# Formulation

$$\mathrm{Dist}(\partial G, \partial S_\theta) \approx \int_\Omega \phi_G(q) s_\theta(q) dq - \int_\Omega \phi_G(q) g(q) dq$$

# Formulation

$$\min_{\theta} \text{Dist}(\partial G, \partial S_\theta) \approx \int_\Omega \phi_G(q) s_\theta(q) dq - \int_\Omega \phi_G(q) g(q) dq$$

# Formulation

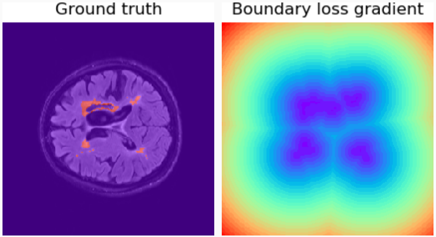
$$\min_{\theta} \mathrm{Dist}(\partial G, \partial S_\theta) \approx \int_\Omega \phi_G(q) s_\theta(q) dq - \int_\Omega \phi_G(q) g(q) dq$$

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

# Formulation

$$\min_{\theta} \mathrm{Dist}(\partial G, \partial S_\theta) \approx \int_\Omega \phi_G(q) s_\theta(q) dq - \int_\Omega \phi_G(q) g(q) dq$$

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




$G$ $\quad$ $\frac{\partial \mathcal{L}_B(\theta)}{\partial S_\theta} = \phi_G$

$\phi_G$ can be efficiently pre-computed

# Experiments

# Datasets

Experiments on two dataset:

- White Matter Hyperintensities (WMH)[1]
- Ischemic stroke lesion (ISLES)[2]

[1] http://wmh.isi.uu.nl
[2] http://www.isles-challenge.org

# Datasets

Experiments on two dataset:

- White Matter Hyperintensities (WMH)[1]
- Ischemic stroke lesion (ISLES)[2]

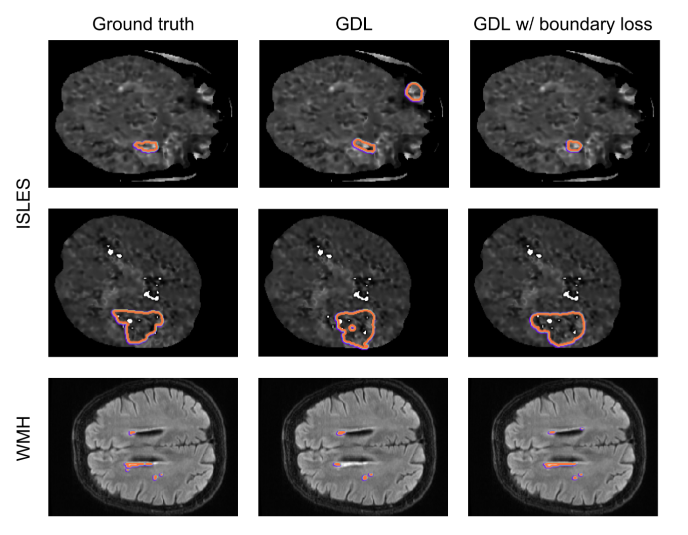
Combine with a GDL for the initial predictions:

$$\alpha \mathcal{L}_{GD}(\theta) + (1 - \alpha)\mathcal{L}_{B}(\theta)$$

We decrease $\alpha$ over time, with the boundary loss taking over.

[1] `http://wmh.isi.uu.nl`
[2] `http://www.isles-challenge.org`

# Results

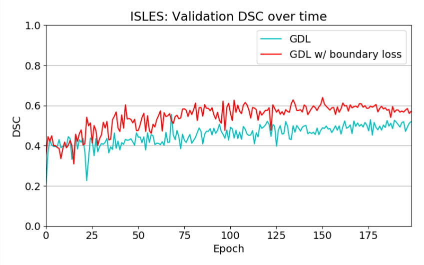


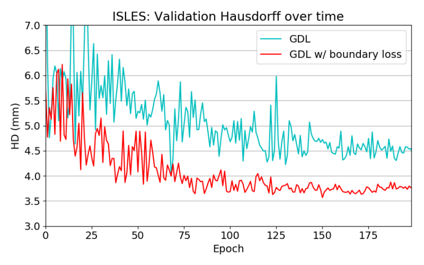
# Results

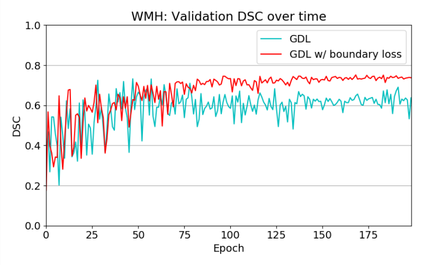
| | ISLES | | WMH | |
|---|---|---|---|---|
| Loss | DSC | HD (mm) | DSC | HD (mm) |
| $\mathcal{L}_B$ | 0.321 (0.000) | NA | 0.569 (0.000) | NA |
| $\mathcal{L}_{GD}$ | 0.575 (0.028) | 4.009 (0.016) | 0.727 (0.006) | 1.045 (0.014) |
| $\mathcal{L}_{GD} + \mathcal{L}_B$ | **0.656 (0.023)** | **3.562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

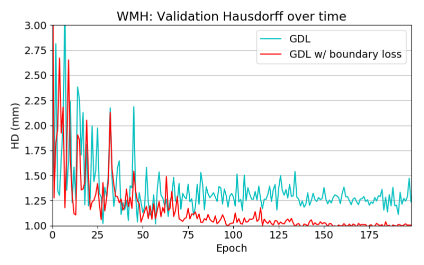
Figure 4: Mean performance (and standard deviation) of 2 runs for each setting.

# Results


ISLES: Validation DSC over time
DSC
Epoch
GDL
GDL w/ boundary loss
1.0
0.8
0.6
0.4
0.2
0.0
0
25
50
75
100
125
150
175

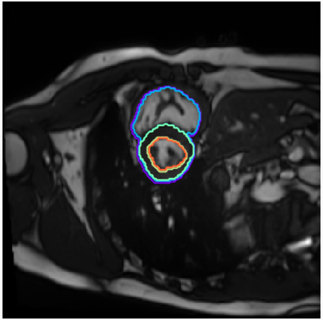

ISLES: Validation Hausdorff over time
HD (mm)
Epoch
GDL
GDL w/ boundary loss
7.0
6.5
6.0
5.5
5.0
4.5
4.0
3.5
3.0
0
25
50
75
100
125
150
175


WMH: Validation DSC over time
DSC
Epoch
GDL
GDL w/ boundary loss
1.0
0.8
0.6
0.4
0.2
0.0
0
25
50
75
100
125
150
175

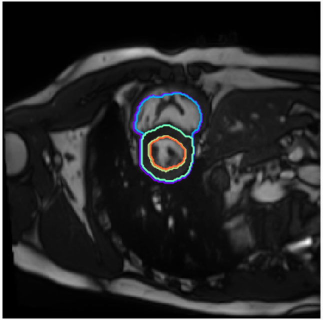

WMH: Validation Hausdorff over time
HD (mm)
Epoch
GDL
GDL w/ boundary loss
3.00
2.75
2.50
2.25
2.00
1.75
1.50
1.25
1.00
0
25
50
75
100
125
150
175

# Other applications

Preliminary results on ACDC (4 classes): might work as stand-alone loss.



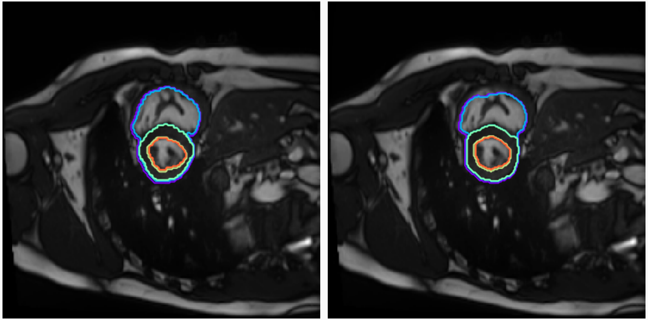
Ground truth



Boundary loss alone

# Other applications

Preliminary results on ACDC (4 classes): might work as stand-alone loss.



Ground truth          Boundary loss alone

Already several feedbacks of improved performances in various applications:
`https://github.com/LIVIAETS/surface-loss`

# Take home message

# Take home message

- New, computationally efficient loss that uses boundary information
- Can be used for any standard regional network
- Improve results and stabilize training on 2 highly-imbalanced datasets
- Many more potential applications to evaluate

# Take home message

- New, computationally efficient loss that uses boundary information
- Can be used for any standard regional network
- Improve results and stabilize training on 2 highly-imbalanced datasets
- Many more potential applications to evaluate

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

# References i

Boykov, Y., Kolmogorov, V., Cremers, D., and Delong, A. (2006).
**An integral solution to surface evolution PDEs via geo-cuts.**
In *European Conference on Computer Vision*, pages 409–422. Springer.