Journal Club 08/07/2019

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*

Presented by Jun Ma

MIDL
London 2019

8 - 10 July 2019

- HOME
- LIVE STREAM
- PROGRAM
- SOCIAL EVENTS
- PROCEEDINGS
- OPENREVIEW
- REGISTRATION
- IMPORTANT DATES
- INFORMATION FOR AUTHORS

International Conference on

Medical Imaging with Deep Learning

London, 8 – 10 July 2019

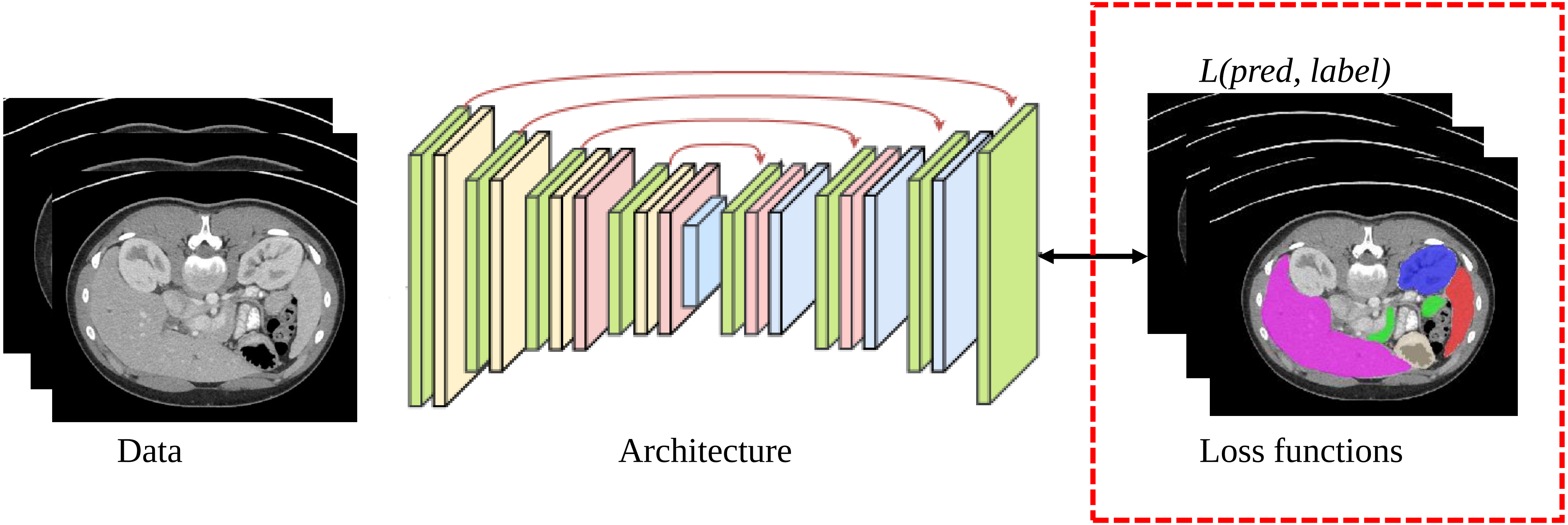
info@midl.io | @midl_conference

Full program as Google Calendar

Outline

1. Background & Motivation
2. Boundary Loss
3. Experiments
4. Beyond Boundary loss

1. Background & Motivation



CNN for medical image segmentation

Recent researches on data augmentation: https://twitter.com/AtoAndyKing/status/1147189669462970369
A collection of SOTA segmentation methods: https://github.com/JunMa11/SOTA-MedSeg

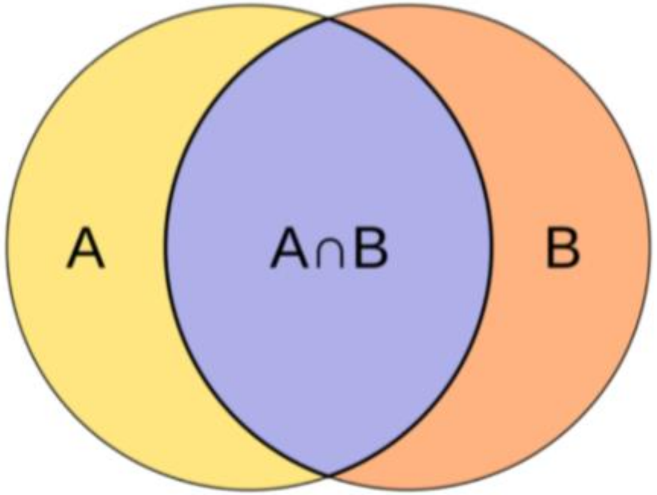
1. Background & Motivation

Two commonly used loss functions

Cross Entropy (CE)

$$-(y\log(p) + (1-y)\log(1-p))$$

Dice

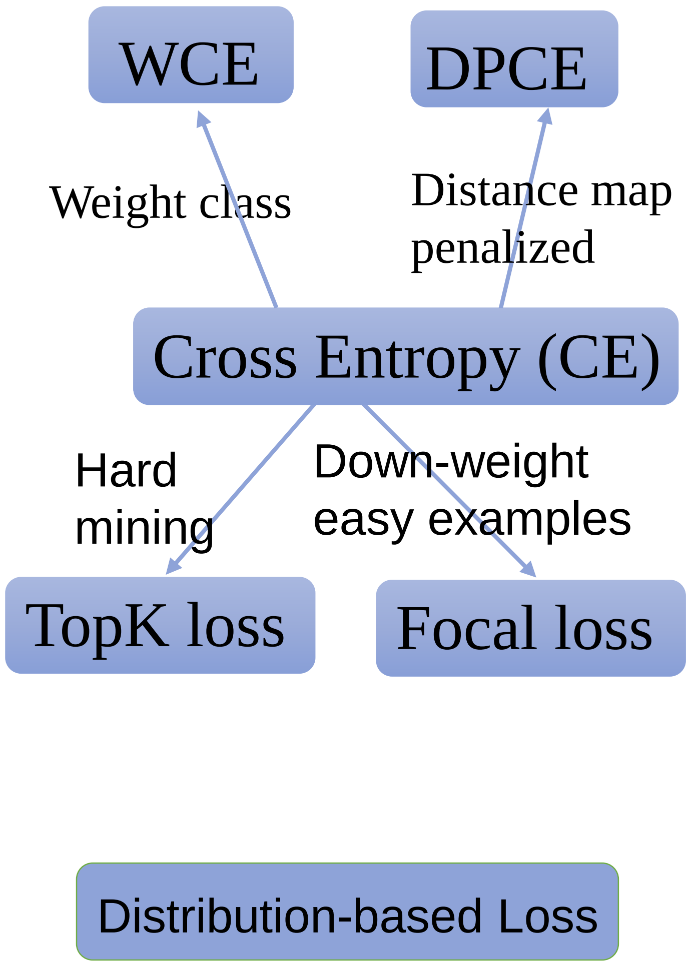


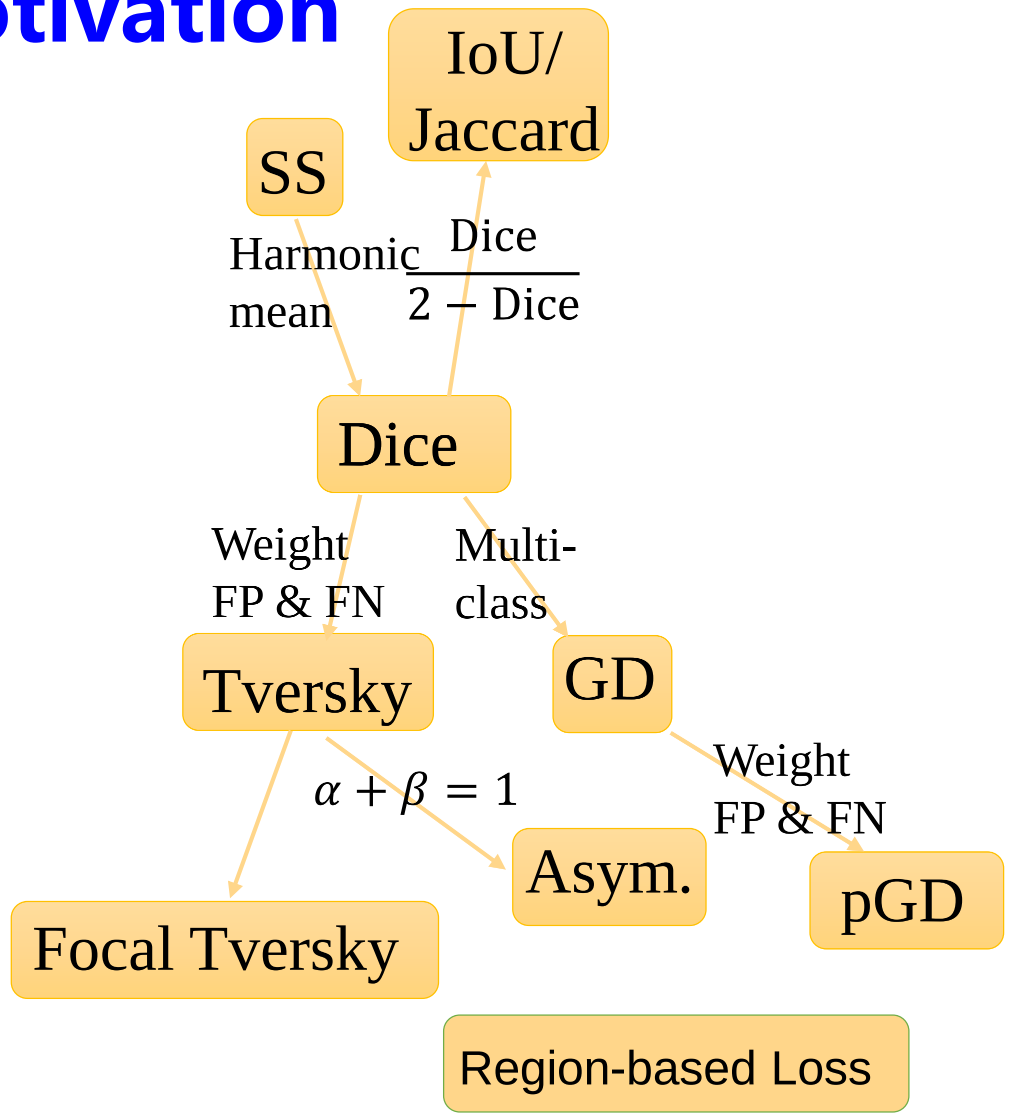
$$\text{Dice loss} = 1 - \frac{2|A \cap B|}{|A|+|B|}$$

Distribution-based Loss

Region-based Loss

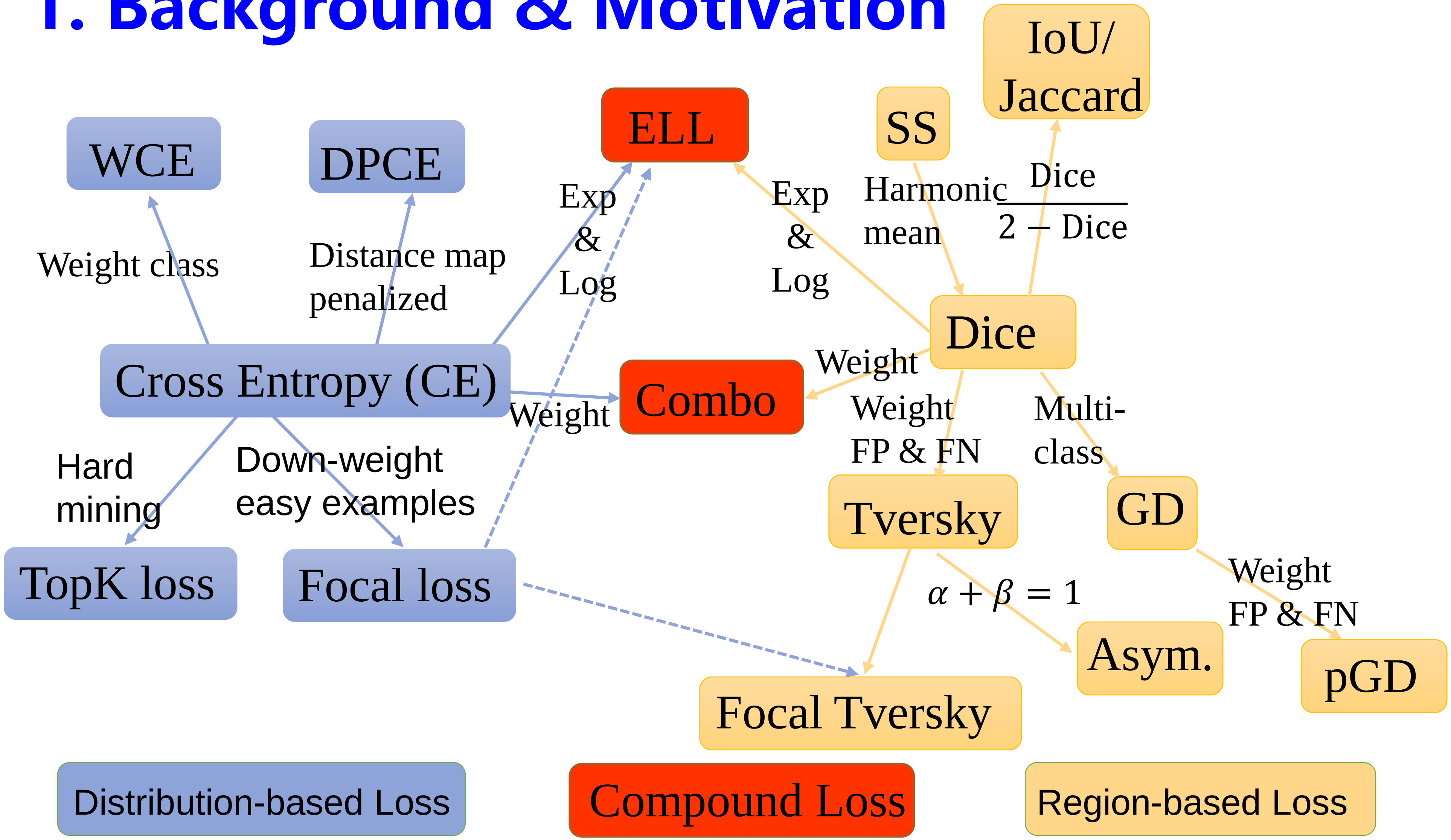
1. Background & Motivation





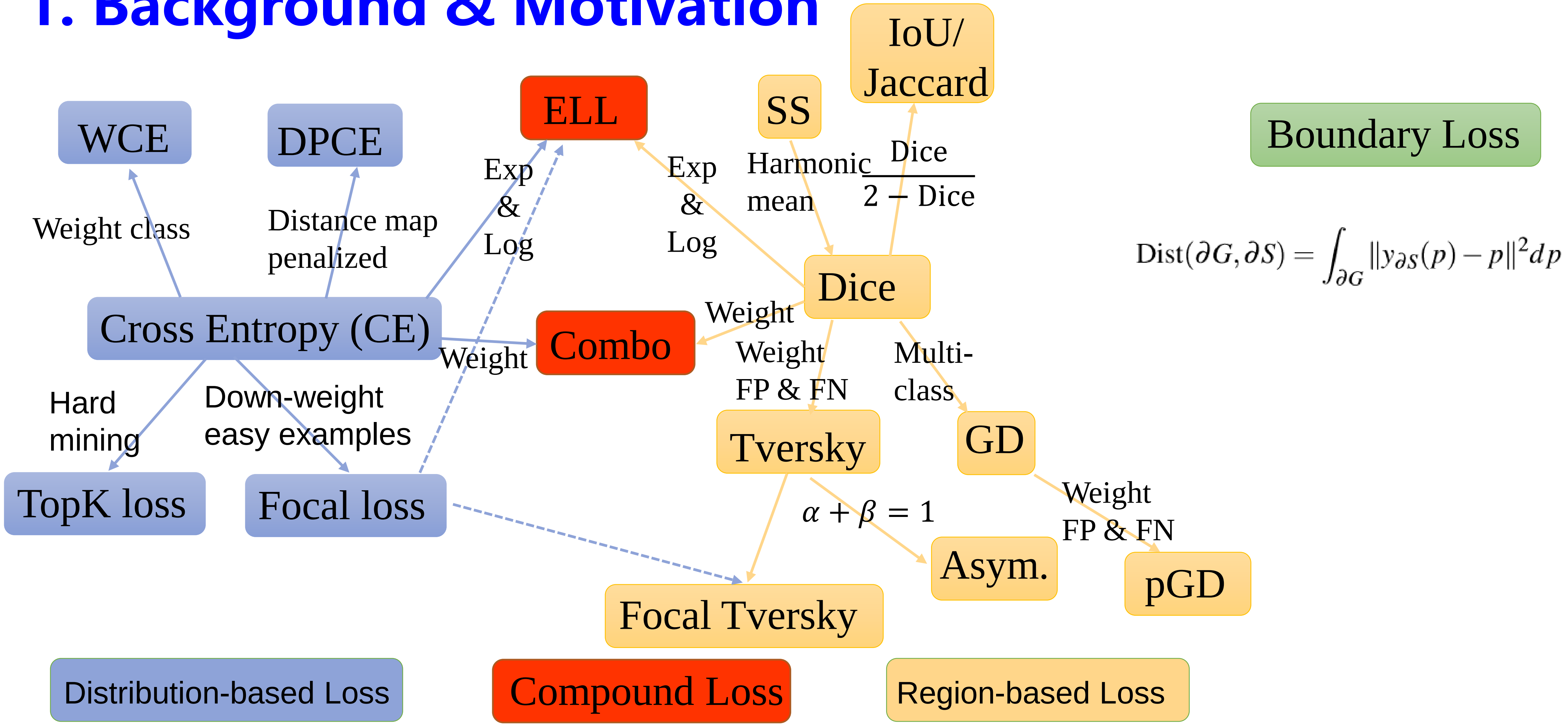
A lot of variants…

1. Background & Motivation



A lot of variants…

1. Background & Motivation

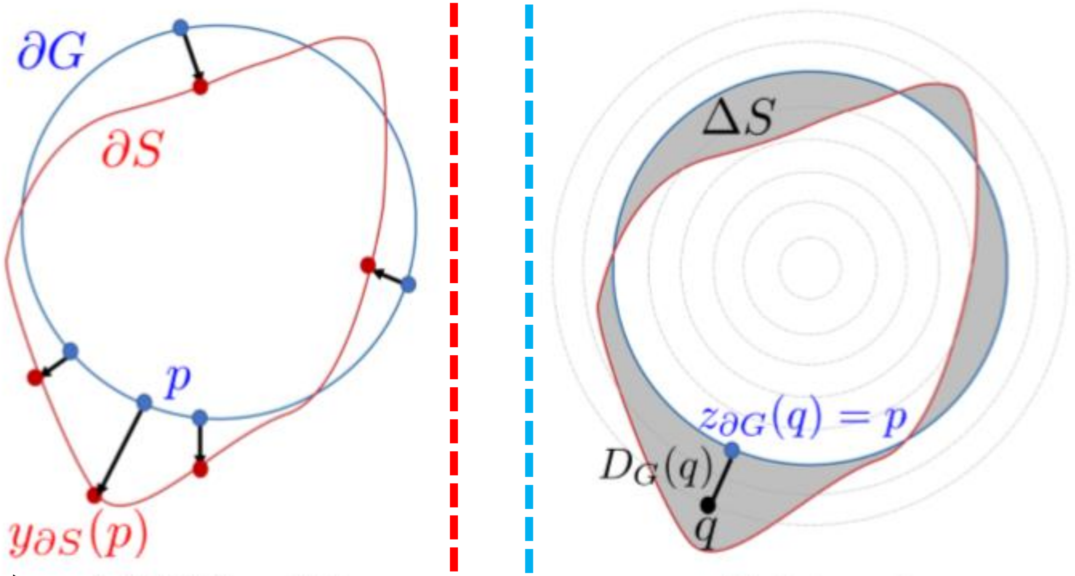


A lot of variants…

2. Boundary Loss

Aim: minimize the distance between two boundaries

$$\mathrm{Dist}(\partial G, \partial S) = \int_{\partial G} \|y_{\partial S}(p) - p\|^2 dp \quad (1)$$



(*a*) Differential

(*b*) Integral

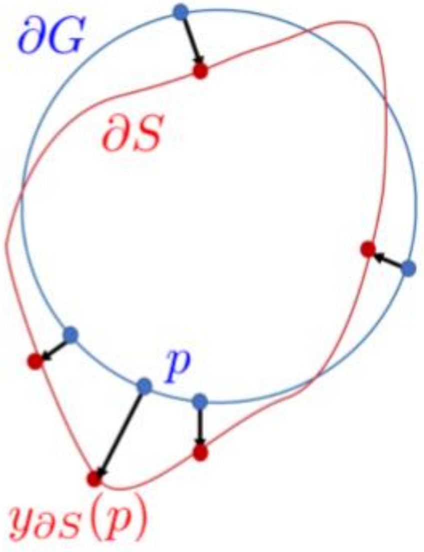
Figure 2: The relationship between *differential* and *integral* approaches for evaluating boundary change (variation).

2. Boundary Loss

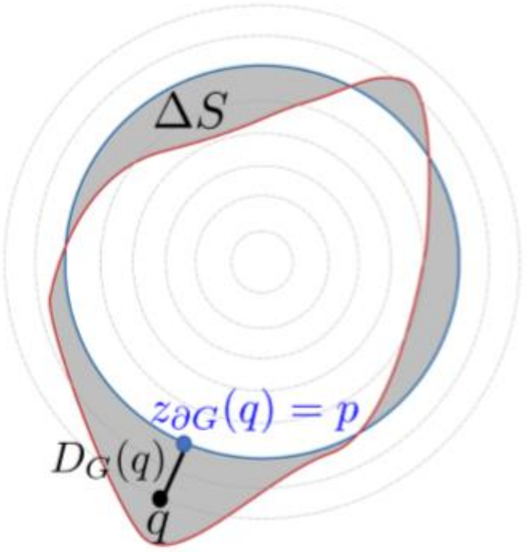
Aim: minimize the distance between two boundaries

It can not be used directly as a loss.

$$\mathrm{Dist}(\partial G, \partial S) = \int_{\partial G} \| y_{\partial S}(p) - p \|^2 dp \quad (1)$$



(*a*) Differential



(*b*) Integral

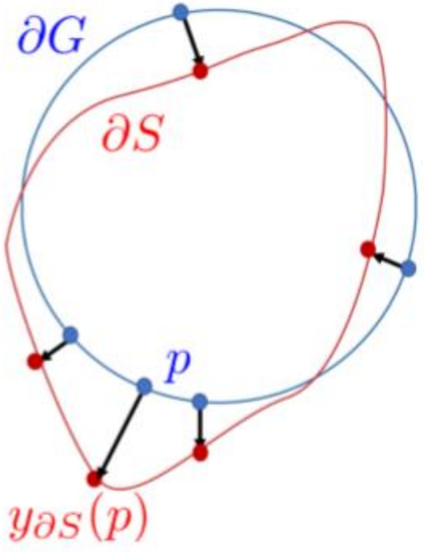
Figure 2: The relationship between *differential* and *integral* approaches for evaluating boundary change (variation).

2. Boundary Loss

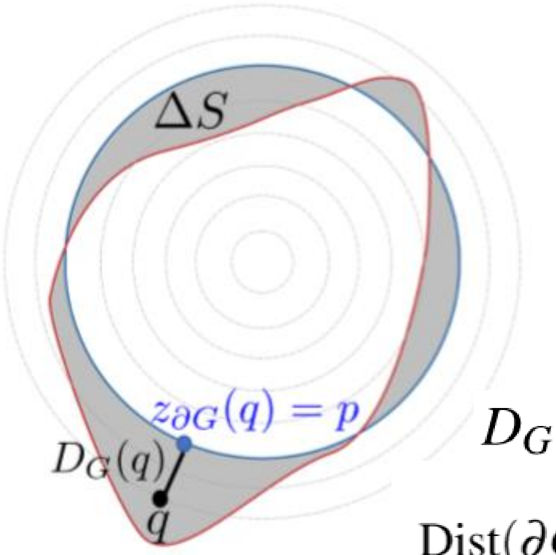
Aim: minimize the distance between two boundaries

It can not be used directly as a loss.

$$\mathrm{Dist}(\partial G, \partial S) = \int_{\partial G} \|y_{\partial S}(p) - p\|^2 dp \quad (1)$$



(a) Differential



(b) Integral

[Boykov et al. 2006]

$$D_G(q) = \|q - z_{\partial G}(q)\|$$

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

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

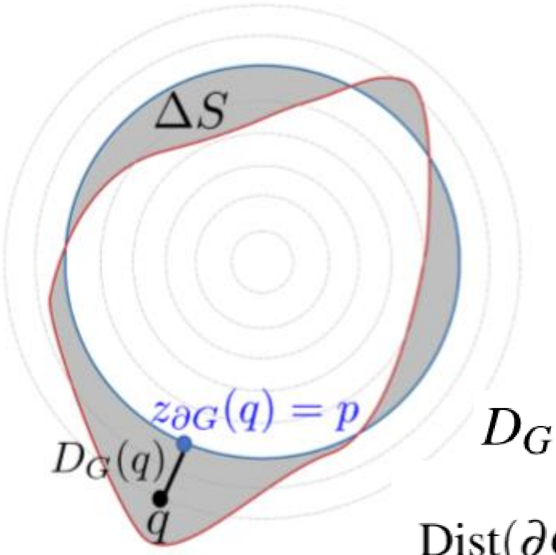
2. Boundary Loss

Aim: minimize the distance between two boundaries

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

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

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



(*b*) Integral

$$D_G(q) = \|q - z_{\partial G}(q)\|$$

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

2. Boundary Loss

CNN outputs $s_\theta(q)$ Omit

$$\frac{1}{2}\mathrm{Dist}(\partial G, \partial S) = \int_S \phi_G(q)dq - \int_G \phi_G(q)dq = \int_\Omega \phi_G(q)\boxed{s(q)}dq - \boxed{\int_\Omega \phi_G(q)g(q)dq} \quad (4)$$

$$\phi_G(q) = \begin{cases} -D_G(q) & \text{if } q \in G \\ D_G(q) & \text{otherwise} \end{cases}$$

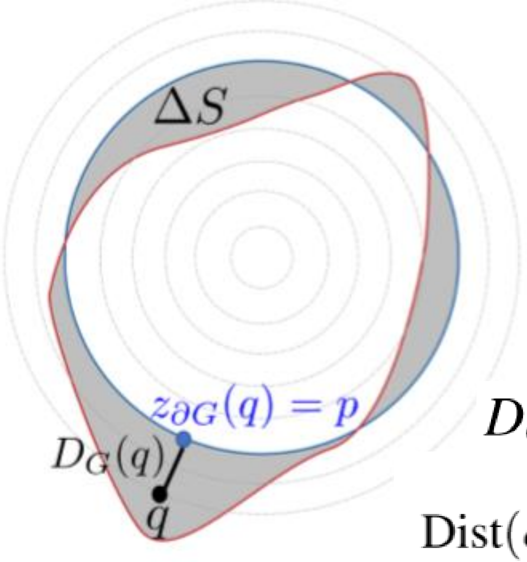
$$s(q) = 1 \text{ if } q \in S$$

boundary loss

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

In the experiments, boundary loss is used in conjunction with generalized dice loss.

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



(b) Integral

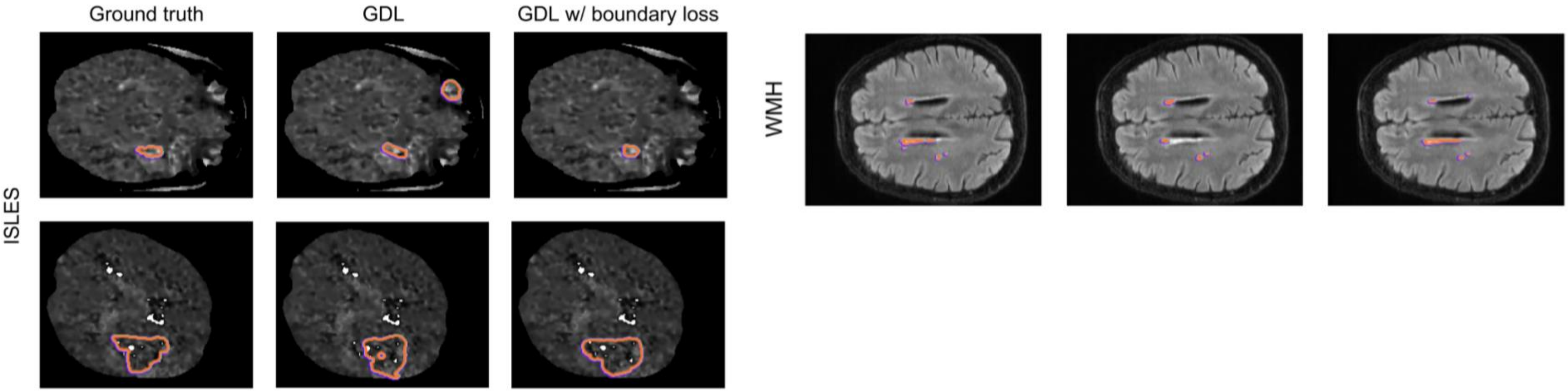
$$D_G(q) = \|q - z_{\partial G}(q)\|$$

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

3. Experiments: datasets

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. Experiments: training protocol

Model: 2D U-Net

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

Optimizer: Adam

Learning rate: 0.001; halved if the validation performances do not improve during 20 epochs

Batch size: 8

Server: NVIDIA GTX 1080 Ti GPU with 11GBs of memory

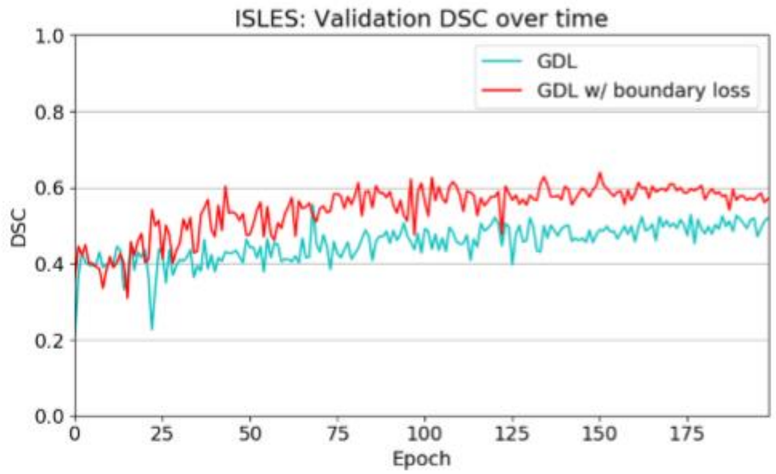
Code: pytorch. https://github.com/LIVIAETS/surface-loss

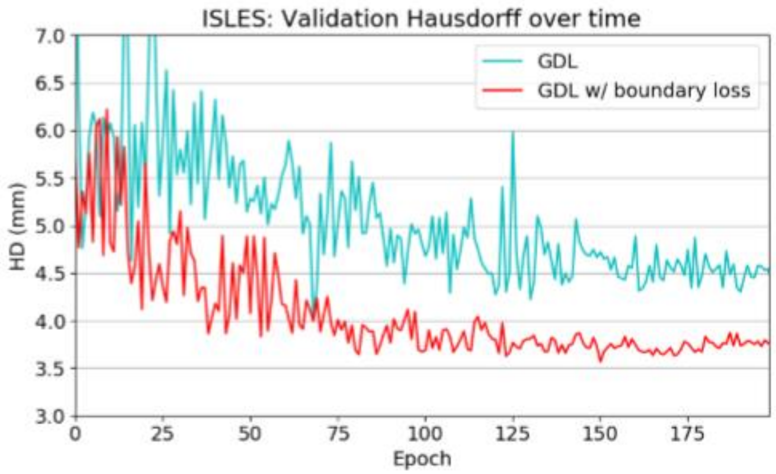
Others: α was initially set to 1, and decreased by 0.01 after each epoch, until it reached the value of 0.01.

3. Experiments: results

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.

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



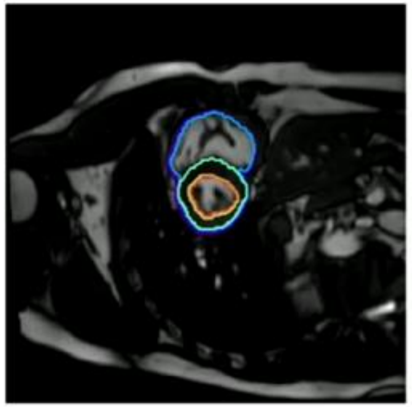


3. Experiments: results

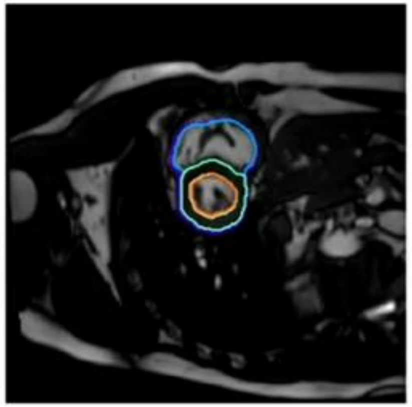
Table 1: DSC and HD values achieved on the validation subset. The values represent the mean performance (and standard deviation) of 2 runs for each setting.

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

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



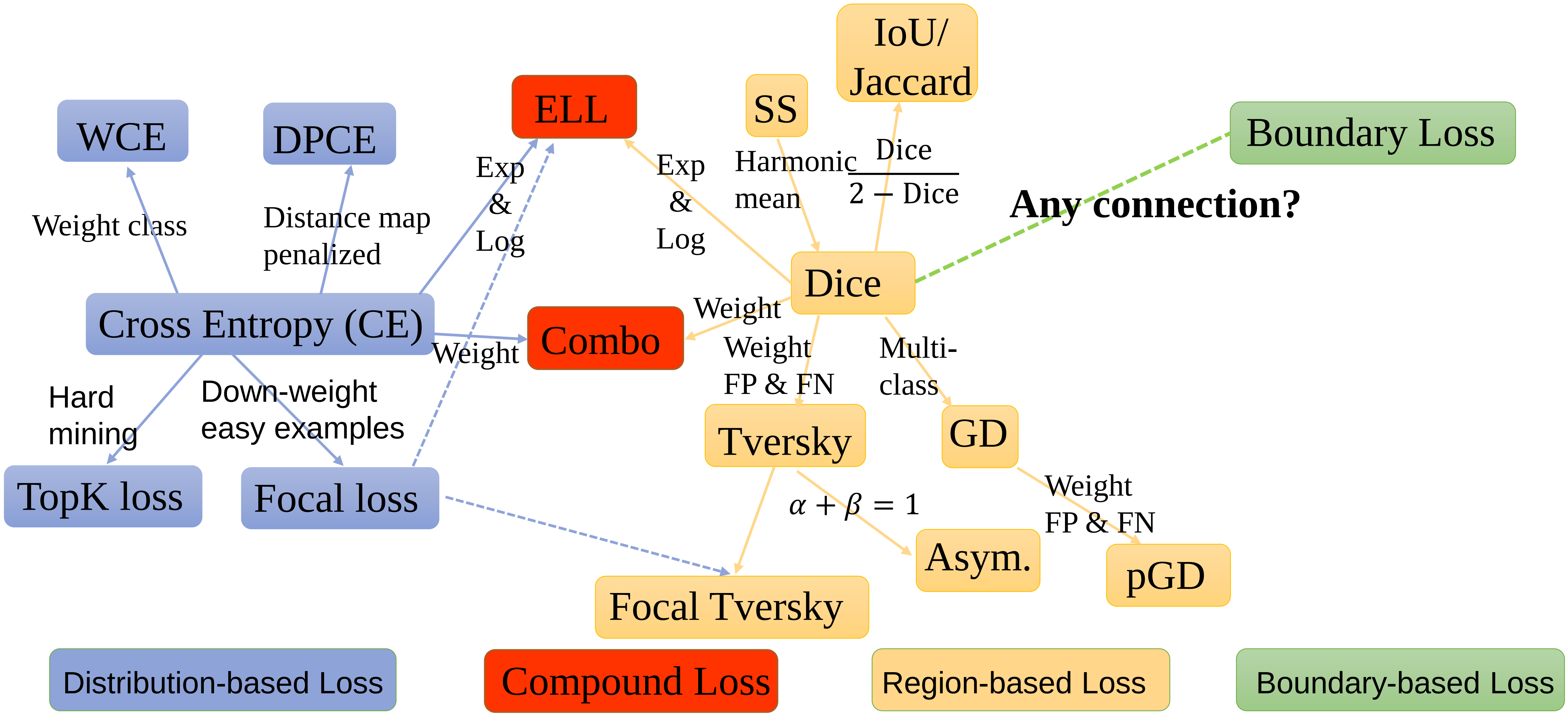
Ground truth



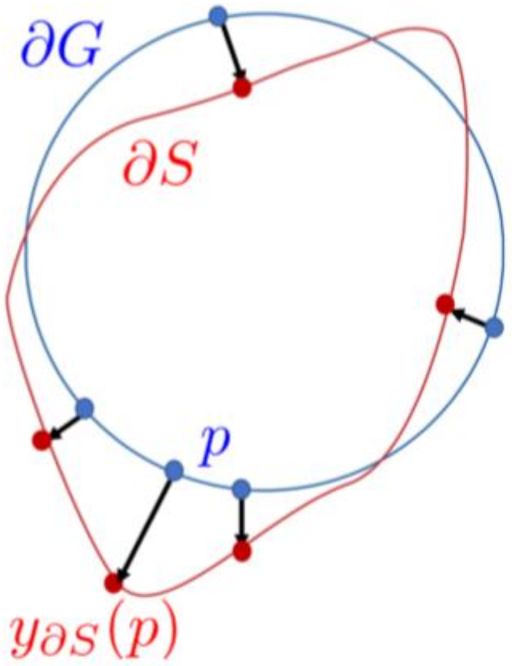
Boundary loss alone

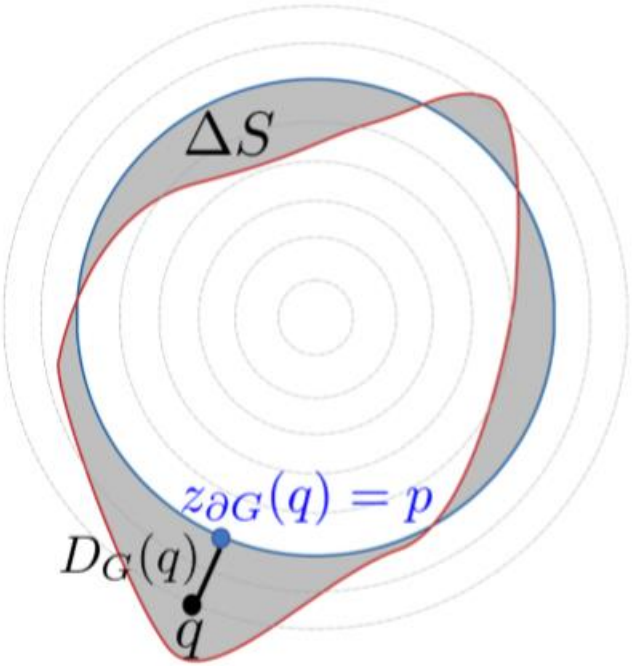
https://www.youtube.com/watch?v=2MWuEoOJ6fo

4. Beyond Boundary Loss



4. Beyond Boundary Loss





Dice loss

$$= 1 - \frac{2|S \cap G|}{|S| + |G|}$$

$$= \frac{|S| - |S \cap G| + |G| - |S \cap G|}{|S| + |G|}$$

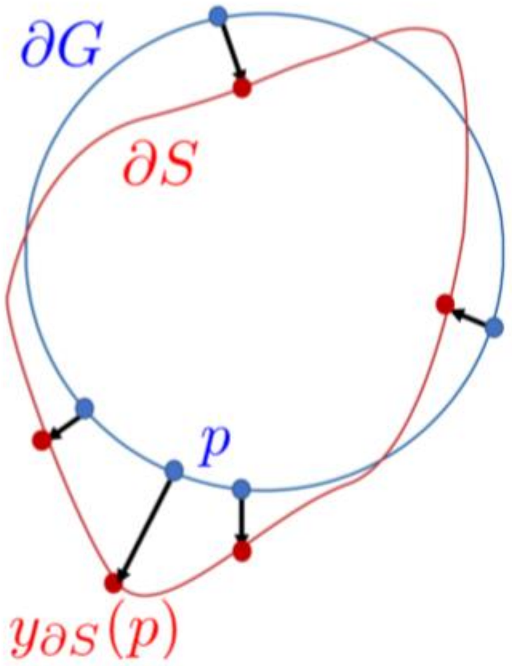
$$= \frac{\boxed{\Delta S}}{|S| + |G|}$$

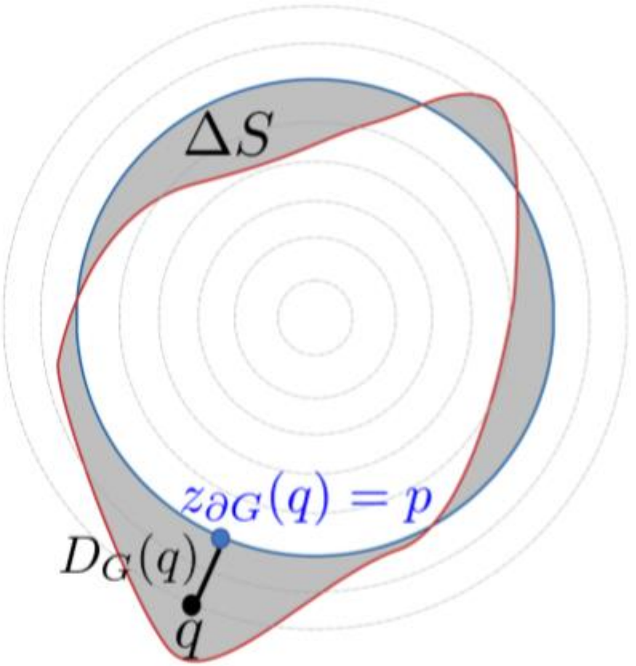
$$D_G(q) = \|q - z_{\partial G}(q)\|$$

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

$$\Delta S = (S \setminus G) \cup (G \setminus S)$$
$$= (S \setminus (G \cap S)) \cup (G \setminus (G \cap S))$$

4. Beyond Boundary Loss





Both dice loss and boundary loss aim to minimize the mismatch region.

The key difference is weighting method.

Dice loss

$$= 1 - \frac{2|S \cap G|}{|S| + |G|}$$

$$= \frac{|S| - |S \cap G| + |G| - |S \cap G|}{|S| + |G|}$$

$$= \frac{\boxed{\Delta S}}{|S| + |G|}$$

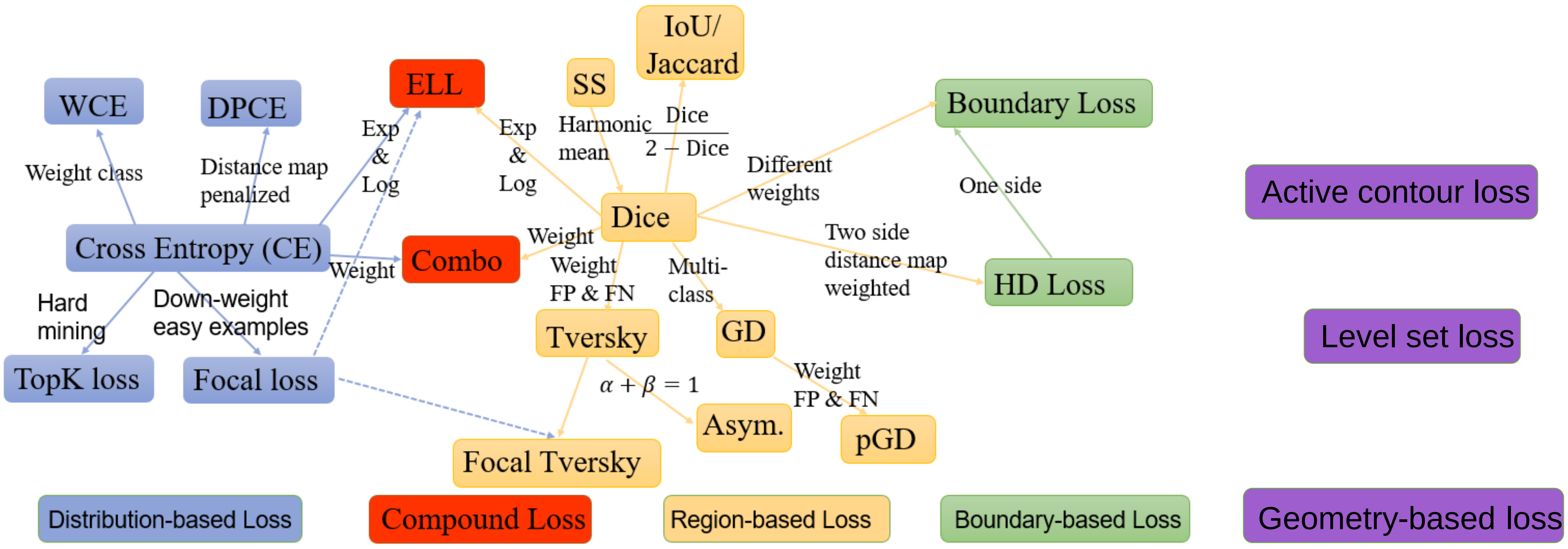
$$D_G(q) = \|q - z_{\partial G}(q)\|$$

$$\text{Dist}(\partial G, \partial S) = 2 \int_{\boxed{\Delta S}} D_G(q) dq$$

$$\Delta S = (S \setminus G) \cup (G \setminus S)$$
$$= (S \setminus (G \cap S)) \cup (G \setminus (G \cap S))$$

4. Beyond Boundary Loss

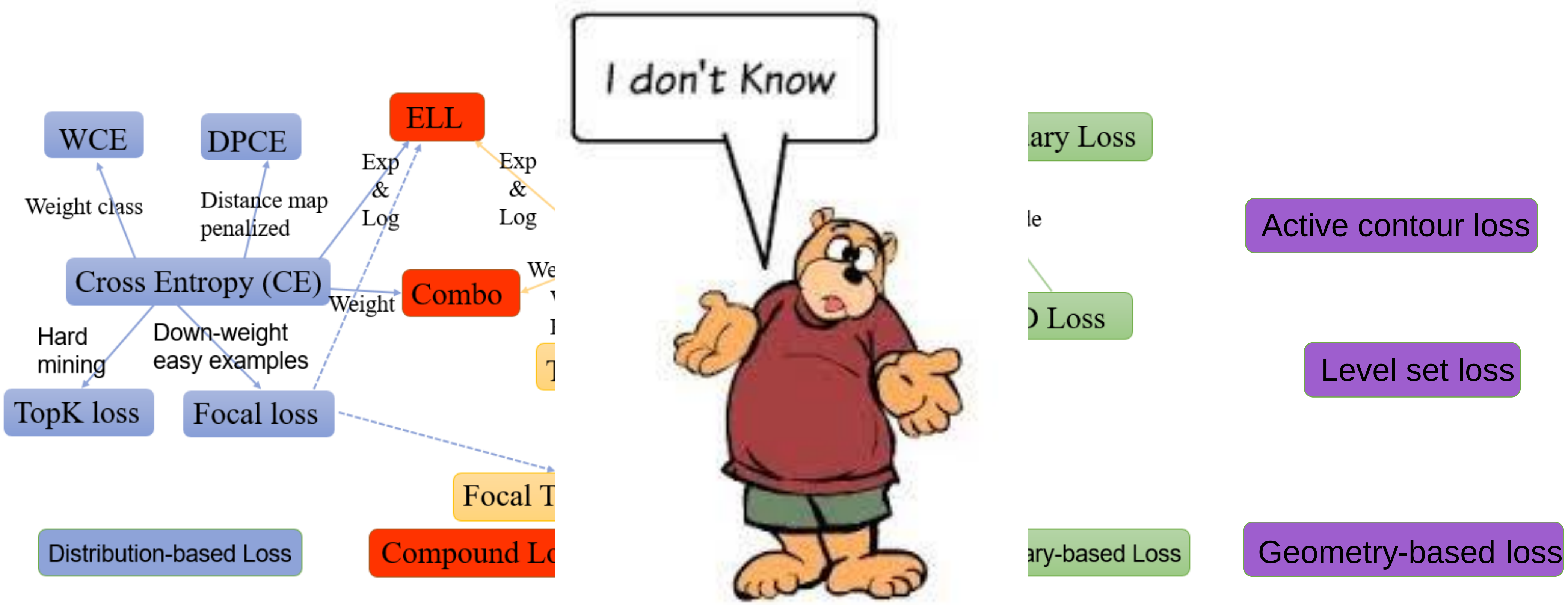
The whole loss function picture



Which one should we use for medical image segmentation tasks?

4. Beyond Boundary Loss

The whole loss function picture



Which one should we use for medical image segmentation tasks?

4. Beyond Boundary Loss

But, there are some side evidences...

(b) Average Dice coefficients (mean±std%) with respective to the ground truth.

Proposed network with linear Dice loss, logarithmic Dice loss, and weighted cross-entropy							
$\mathbf{E}[1 - \text{Dice}_i]$ (2)	1. 87±1	2. 47±38	3. 32±40	4. 72±36	5. 50±41	6. 30±37	7. 31±38
	8. 0±0	9. 0±0	10. 34±42	11. 88±1	12. 86±1	13. 88±1	14. 32±39
	15. 0±0	16. 54±44	17. 0±0	18. 51±42	19. 35±43	**Average: 43±11**	
$L_{\text{Dice}}(\gamma = 1)$ (2)	1. 84±1	2. 61±30	3. 83±2	4. 90±1	5. 81±2	6. 73±2	7. **78±2**
	8. 68±2	9. 74±2	10. 85±1	11. 87±1	12. 85±1	13. 88±1	14. 79±2
	15. 59±3	16. 89±1	17. 79±1	18. 86±2	19. 88±1	**Average: 80±2**	
$L_{\text{Cross}}(\gamma = 1)$ (3)	1. 87±1	2. 56±5	3. 79±3	4. 86±2	5. 76±3	6. 67±2	7. 73±6
	8. 59±4	9. 65±4	10. 83±2	11. 87±2	12. 85±1	13. 89±1	14. 75±3
	15. 54±6	16. 89±1	17. 76±3	18. 84±1	19. 86±1	**Average: 77±2**	
Proposed network with L_{Exp} at different values of γ							
$L_{\text{Exp}}(\gamma = 1)$ (1)	1. 87±2	2. **78±3**	3. **84±1**	4. 90±1	5. **82±1**	6. 74±2	7. 78±3
	8. 68±3	9. **75±1**	10. 83±3	11. 87±1	12. **86±0**	13. 89±1	14. 80±1
	15. **64±1**	16. 90±1	17. 80±2	18. 86±2	19. 88±1	**Average: 81±1**	
$L_{\text{Exp}}(\gamma = 2)$ (1)	1. 79±7	2. 61±15	3. 74±6	4. 75±10	5. 67±12	6. 62±8	7. 66±10
	8. 52±17	9. 56±15	10. 64±12	11. 78±8	12. 78±7	13. 84±4	14. 64±11
	15. 46±10	16. 77±10	17. 60±16	18. 67±15	19. 67±15	**Average: 67±11**	
$L_{\text{Exp}}(\gamma = 0.3)$ (1)	1. **88±1**	2. 77±2	3. **84±1**	4. **91±1**	5. **82±1**	6. **74±1**	7. **78±2**
	8. **69±2**	9. 75±2	10. **86±1**	11. **89±1**	12. 86±1	13. **89±0**	14. **81±1**
	15. 62±5	16. **91±1**	17. **80±1**	18. **87±1**	19. **89±1**	**Average: 82±1**	
V-Net with the best L_{Exp} at $\gamma = 0.3$							
V-Net	1. 84±2	2. 67±7	3. 80±4	4. 87±4	5. 78±3	6. 67±5	7. 73±6
$L_{\text{Exp}}(\gamma = 0.3)$ (1)	8. 59±7	9. 65±5	10. 72±5	11. 85±2	12. 82±4	13. 86±2	14. 72±7
	15. 48±8	16. 82±6	17. 70±7	18. 75±6	19. 78±6	**Average: 74±4**	

Dice + Cross entropy

$$L_{\text{Exp}} = w_{\text{Dice}} L_{\text{Dice}} + w_{\text{Cross}} L_{\text{Cross}} \quad (1)$$

with w_{Dice} and w_{Cross} the respective weights of the exponential logarithmic Dice loss (L_{Dice}) and the weighted exponential cross-entropy (L_{Cross}):

$$L_{\text{Dice}} = \mathbf{E}\left[(-\ln(\text{Dice}_i))^{\gamma_{\text{Dice}}}\right] \text{ with } \text{Dice}_i = \frac{2\left(\sum_{\mathbf{x}} \delta_{il}(\mathbf{x})\, p_i(\mathbf{x})\right) + \epsilon}{\left(\sum_{\mathbf{x}} \delta_{il}(\mathbf{x}) + p_i(\mathbf{x})\right) + \epsilon} \quad (2)$$

$$L_{\text{Cross}} = \mathbf{E}\left[w_l\, (-\ln(p_l(\mathbf{x})))^{\gamma_{\text{Cross}}}\right] \quad (3)$$

Wong, Ken CL, et al. "3d segmentation with **exponential logarithmic loss** for highly unbalanced object sizes." *International Conference on Medical Image Computing and Computer-Assisted Intervention*. 2018.

4. Beyond Boundary Loss

But, there are some side evidences…

TABLE V. Comparisons of test performances of models trained with different loss functions, evaluated with Dice coefficients.

Anatomy	Dice loss	Exp. Log. Dice	Dice + focal	Dice + cross entropy
Brain Stem	85.1	85.0	**86.1**	85.2
Chiasm	50.1	50.0	**52.2**	48.8
Mand.	**91.5**	89.9	90.0	91.0
Optic Ner L	69.1	67.9	68.4	**69.6**
Optic Ner R	66.9	65.9	**69.1**	67.4
Paro. L	86.6	86.4	87.4	**88.0**
Paro. R	85.6	84.8	86.3	**86.9**
Subm. L	78.5	76.3	**79.6**	77.8
Subm. R	77.7	78.2	**79.8**	78.4
Average	76.8	76.0	**77.7**	77.0

Dice + Focal loss

Zhu, Wentao, et al. "AnatomyNet: Deep learning for fast and fully automated whole-volume segmentation of head and neck anatomy." *Medical physics* 46.2 (2019): 576-589.

4. Beyond Boundary Loss

But, there are some side evidences...

	BraTS	Liver lowres	Liver fullres	Hippocampus	Prostate	Lung nodule	Pancreas
Vanilla nnU-Net	0.72	0.79	0.78	0.89	0.77	0.65	0.65
Batch norm instead of Inst. norm	1.0%	-0.1%	2.9%	-0.1%	-1.3%	-14.2%	-3.7%
No feature map normalization	1.1%	-4.6%	-22.8%	-0.2%	-4.2%	3.0%	-100.0%
ReLU instead of LeakyReLU	0.6%	0.0%	1.0%	-0.1%	-0.2%	-0.4%	0.5%
No data augmentation	-0.8%	-4.9%	1.5%	-1.5%	-0.4%	4.2%	-11.3%
Only cross-entropy loss	-0.6%	-12.0%	-6.3%	0.0%	-1.4%	-25.4%	-8.8%
Only dice loss	0.9%	-2.5%	-10.1%	-0.3%	-3.0%	-11.5%	1.6%

Dice + Cross entropy

Isensee, Fabian, et al. "nnU-Net: Breaking the Spell on Successful Medical Image Segmentation." *arXiv preprint arXiv:1904.08128* (2019).

4. Beyond Boundary Loss

But, there are some side evidences…

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.

	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)**

Dice + Boundary loss

Kervadec, Hoel, et al. "Boundary loss for highly unbalanced segmentation." *International Conference on Medical Imaging with Deep Learning*. 2019.

Take home message

- Boundary loss is compliment to regional losses (e.g, Dice loss). It can be used for any standard segmentation networks.
- Using compound loss functions is a better choice for medical image segmentation tasks.

A collection of loss functions https://github.com/JunMa11/SegLoss

Thanks for your attention!