On power losses for semantic segmentation

David DUQUE

david.duque@mines-paristech.fr

Under supervision of:
Beatriz MARCOTEGUI
Santiago VELASCO-FORERO
Jean-Emmanuel DESCHAUD
Francois GOULETTE

Summary

1. Introduction
2. Loss functions
3. Power losses
4. Experiments
5. Conclusions and future work

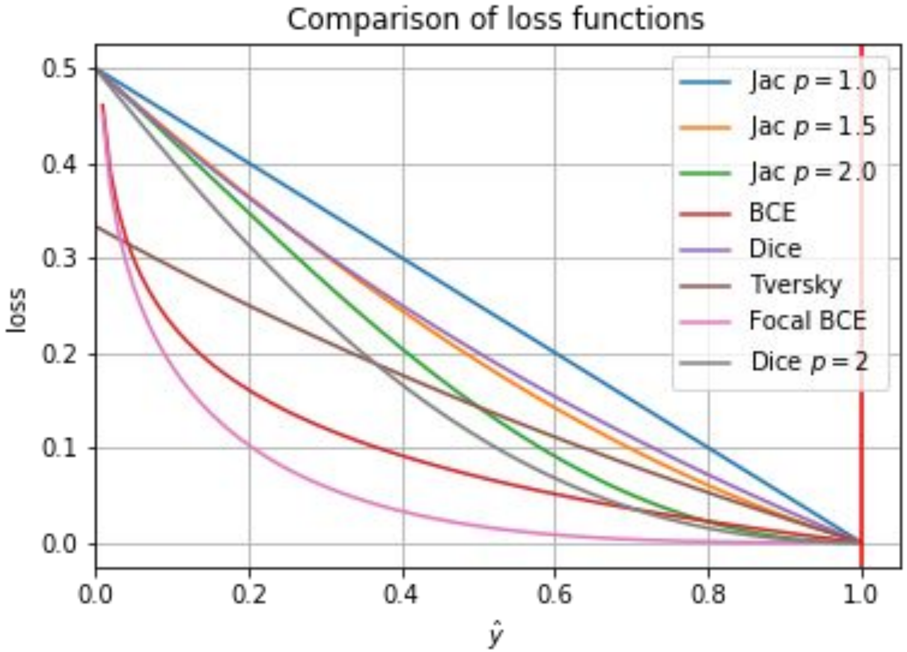
Summary

Introduction

1. Segmentation of **highly unbalanced datasets** is an active research topic.
2. Most common strategies:
   a. Data augmentation.
   b. **Loss function.**

Goal: To penalize in different ways the incorrect labels during training.

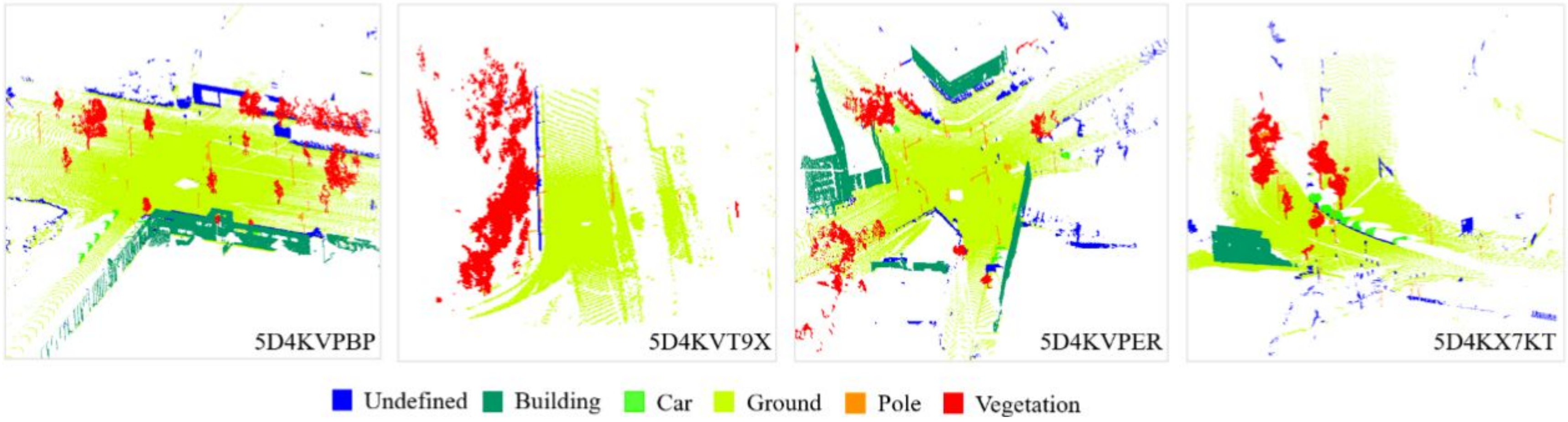
Introduction



Comparison of some classical loss functions: binary crosssentropy (BCE), Dice, Jaccard, Tversky, Focal BCE and our proposed loss functions.

Motivation

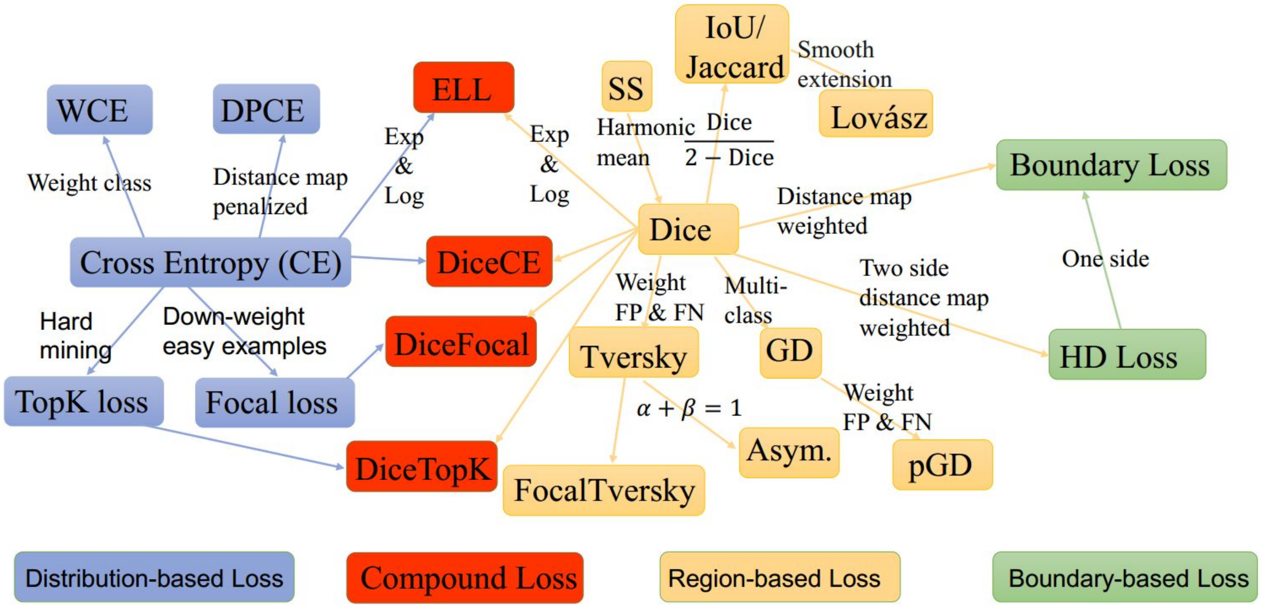
SHREC'20 challenge in 3D point cloud semantic segmentation for street scenes [1]



[1] T. Ku, R.. Veltkamp, B. Boom, D. Duque-Arias, S. Velasco-Forero, J-E. Deschaud, F. Goulette, B. Marcotegui, S. Ortega, A. Trujillo, J. Suárez, J. Santana, C. Ramírez, K. Akadas, S. Gangisetty, SHREC 2020 Track: 3D Point Cloud Semantic Segmentation for Street Scenes, Computers & Graphics, 2020,

Summary

Loss functions



Taken from *https://github.com/JunMa11/SegLoss*

Jaccard index

Jaccard index→ Metric of **similarity** between two sets (A.K.A IoU)

$$J_i = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|}$$

Jaccard distance → Can be implemented as a loss function

$$J_d = 1 - J_i = 1 - \frac{|A \cap B|}{|A \cup B|}$$

$$J_l(y, \hat{y}) = 1 - \frac{(y \cdot \hat{y}) + \boxed{\epsilon}}{(y + \hat{y} - y \cdot \hat{y}) + \boxed{\epsilon}}$$

Dice index

Dice index→ Metric of **similarity** between two sets.

$$D_i = \frac{2|A \cap B|}{|A| + |B|}$$

Dice distance → Can be implemented as a loss function

$$D_l(y, \hat{y}) = 1 - \frac{2 \cdot (y \cdot \hat{y}) + \boxed{\epsilon}}{(y + \hat{y}) + \boxed{\epsilon}}$$

Summary

1. Introduction
2. Loss functions
3. **Power losses**
4. Experiments
5. Conclusions and future work

Power losses

Generalized loss functions including a **power term** *p:*

$$J_p(y, \hat{y}, p) = 1 - \frac{(y \cdot \hat{y}) + \epsilon}{(y^{\boxed{p}} + \hat{y}^{\boxed{p}} - y \cdot \hat{y}) + \epsilon}$$

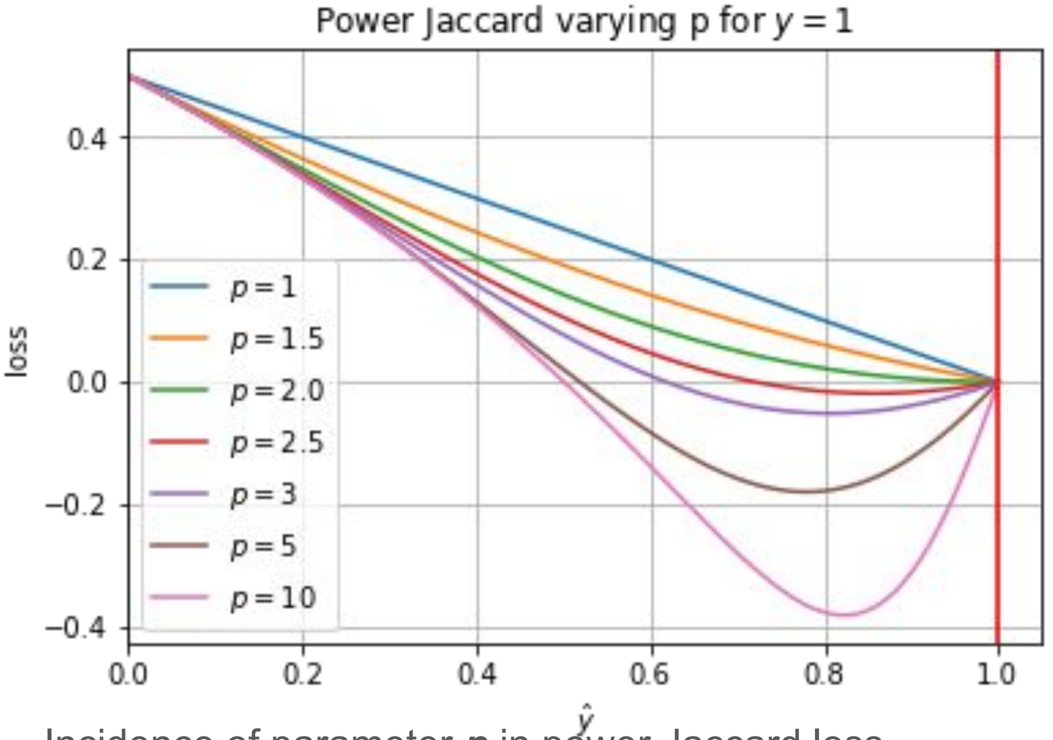
$$D_p(y, \hat{y}, p) = 1 - \frac{2 \cdot (y \cdot \hat{y}) + \epsilon}{(y^{\boxed{p}} + \hat{y}^{\boxed{p}}) + \epsilon}$$

Some previous works have directly used p = 2 in geometrical losses [2-3]

[2] F. Diakogiannis, F. Waldner, P. Caccetta and C Wu. Resunet-a: A deep learning framework for semantic segmentation of re-motely sensed data. ISPRS Journal of Photogrammetry and Remote Sens-ing, 162:94–114, 2020.

[3] E. Decenciere, S. Velasco-Forero, F. Min, J. Chen, H. Burdin, G. Gauthier, B. Lay, T. Bornschloegl and T. Baldeweck. Dealing with topological information within a fullyconvolutional neural network. InInternational Conference on AdvancedConcepts for Intelligent Vision Systems, pages 462–471. Springer, 2018.

Power losses



Incidence of parameter ***p*** in power Jaccard loss.
Vertical red line indicates the ground truth value of ***y* = *1***

Power losses

Derivative analysis to find valid values of p

$$J_p(y, \hat{y}) = \frac{y + \hat{y}^p - 2 \cdot y \cdot \hat{y}}{(y + \hat{y}^p - y \cdot \hat{y}) + \epsilon}$$

$$\frac{\partial J_p}{\partial \hat{y}} = \frac{(\epsilon + y \cdot \hat{y})(p \cdot \hat{y}^{p-1} - y)}{((y + \hat{y}^p - y \cdot \hat{y}) + \epsilon)^2} - \frac{y}{(y + \hat{y}^p - y \cdot \hat{y}) + \epsilon} = 0$$

$$\Rightarrow$$

$$\hat{y}^* = \sqrt[p]{\frac{1}{(p-1)}}$$

$\hat{y}^* \in [0, 1]$, so the inequality $1 <= \hat{y}^*$ implies $\hat{y}^* = y = 1$

$$1 \leq \sqrt[p]{\frac{1}{(p-1)}}$$

$$\boxed{1 \leq p \leq 2}$$

Summary

1. Introduction
2. Loss functions
3. Power losses
4. **Experiments**
5. Conclusions and future work

Experimental design

1. Multiclass segmentation → 2D projections of point clouds
2. Binary segmentation → RGB aerial images.
3. **Multiclass segmentation on MNIST → Gray scale images**

MNIST dataset

We generated datasets for semantic segmentation **based on MNIST images.**

Segmentation of digits + background = 11 classes

Parameters:

- Image size
- Rate of overlapping.
- Size of instances.
- Class imbalancement.
- Noise rate.

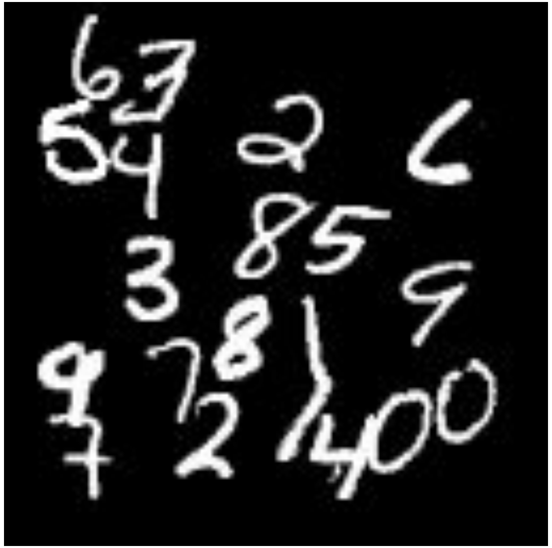
Relation with point clouds:

- Unbalanced datasets (mostly ground)
- "Empty pixels" and "background" in 2D projections.

Goal: Deeper analysis of power losses in well known images.

MNIST dataset

Train set: 1000 images
Validation set: 200 images
Test set: 200 images



Multiclass segmentation

Modified parameters during training

Architecture: Unet with 3 levels of depth.
Input: [W,H,1] (intensity)
Optimizer: Adam with lr=0.001
Losses: [crossentropy, Jaccard and Power Jaccard (p=2)]
Filters: [2, 4, 8, 16]
Kernel size: [3, 5]
Data augmentation: random shifts
Max epochs: 100
Callbacks:

- Patience: 10 epochs.
- Reduce LR on plateau: 7 epochs

Each experiment was repeated 5 times to measure stability

Results

Evaluation in test set composed by 200 images.

Kernel size = 3

Loss \ filters	2	4	8	16
Cat. CE	0.6446 ± 0.0818	0.8039 ± 0.0196	0.8401 ± 0.0209	0.8634 ± 0.0231
Jaccard	0.5527 ± 0.1406	0.8031 ± 0.0480	0.8765 ± 0.0105	**0.8839 ± 0.0093**
Jac. p = 2.0	**0.7407 ± 0.0264**	**0.8308 ± 0.0319**	**0.8870 ± 0.0155**	0.8658 ± 0.0196

Mean IoU and standard deviation for each configuration.

Results

Evaluation in test set composed by 200 images.

Kernel size = 5

Loss \ filters	**2**	**4**	**8**	**16**
Cat. CE	**0.7957 ± 0.0272**	0.8489 ± 0.0087	0.8677 ± 0.0164	0.8828 ± 0.0188
Jaccard	0.7874 ± 0.0518	**0.8560 ± 0.0172**	**0.8750 ± 0.0049**	**0.8951 ± 0.0185**
Jac. p = 2.0	0.7814 ± 0.0346	0.8364 ± 0.0211	0.8715 ± 0.0148	0.8746 ± 0.0214

Mean IoU and standard deviation for each configuration.

Summary

Conclusions and future work

1. In tested scenarios: 2D projections of point clouds, RGB aerial images and MNIST images → Power losses **improved** performance on image segmentation compared against CE and classical Jaccard.
2. Less data and less complex models → Power losses **perform better** than classical Jaccard.
3. In MNIST images, we experimentally found power Jaccard performed better than classical Jaccard and CE with smaller kernel size.

Future work:

- Learn “p” parameter of power losses.

Thanks for your attention