Loss Function for Class Imbalanced Data

Yujie Wu
University of Toronto
1003968904
yujie.wu@mail.utoronto.ca

Abstract

Current lost functions for image segmentation using convolutional neural networks are mainly regional integral based and may yield unexpected results with medical image data, where the occurrence of target pixels and background pixels are usually highly imbalanced. In this paper, I proposed some modifications to the current boundary loss function to address the issue. This modified loss function keeps tracking each pixel and the distance to its nearest contour rather than computing the overlapping regions between prediction and ground truth. The weighted ratio between correctly predicted pixels and incorrectly predicted pixels is computed as the final output of this loss function. Both quantitatively and qualitatively evaluations are reported to compare the loss with several baselines showing that the new loss function outperforms traditional regional integral based loss functions and the original boundary loss.

Keywords: Semantic segmentation, machine learning, medical image, class imbalanced data

1 Introduction

Computer aids to medical diagnosis, specifically medical image segmentation using deep learning networks, are becoming popular over the years. People have witnessed the growth of the deep learning architectures designed specifically for medical image segmentation tasks, such as U-Net [1] and SCAN [2]. In every neural network, loss functions play vital roles. They provide numerical evaluations of the current model and show the direction for optimization and minimization. Conventional loss functions for image segmentation, for example, cross-entropy and dice similarity coefficient, are mainly based on measuring the overlapping regions between predictions and ground truths. These conventional loss functions can often perform well with the normal dataset, but maybe sometimes problematic with the medical image dataset. The key difference between the medical image dataset and other datasets is the fact that the majority of medical image datasets have background pixels occupying nearly the entire image with only a few of the pixels belonging to the foreground (the target). For example, Figure 1 (a) is the segmentation of spine vertebrae from my work, the image only contains approximately 13% of pixels belonging to the foreground (spine vertebrae) and 87% being the background. Even more, Figure 1 (b) is an example from the WMH segmentation challenge dataset [3], more than 99% of pixels, in this case, are background pixels and only 8 locations labelled with colour is white matter hyperintensities.

The high imbalance of the medical images data can lead to unexpected results in segmentation tasks if they are optimized over regional integral based loss functions, such as cross-entropy and dice. The results will be biased towards the majority due to the fact that conventional loss functions assume

Submitted to 30th Conference on Neural Information Processing Systems (NIPS 2016). Do not distribute.

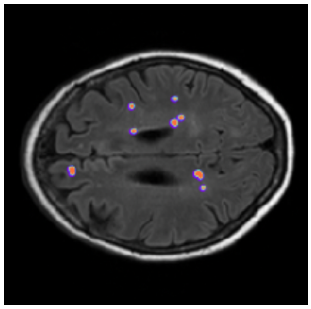

Figure 1: (a) Segmentation of spine vertebrae. (b) Segmentation of white matter hyperintensities.

both foreground and background pixels are equally important. Therefore, it is necessary to develop non regional integral based, weighted loss functions to produce more accurate and reliable outcomes on highly class imbalanced dataset.

1.1 Related work(s)

This project is inspired by the boundary loss, which innovatively encodes distance information into the loss function to encounter the issues with the class imbalanced dataset[4]. The boundary loss function is evaluated by the features in the border between different classes rather than measuring the regions within each class. However, experiments suggest that the boundary loss may sometimes be trapped into local minimum and result in trivial outcomes, and thus need guidance from the traditional loss function. The main task of this project is to avoid the boundary loss from trapping into a local minimum, and the modified loss function also includes more details that the original boundary loss chose to discard (i.e. target regions missing from the prediction). Such tuning involves complete modification to the original loss function, only the idea of distance information encoding remains the same. Other works include refining broken boundaries. This is done by detecting pixels that have similar probabilities of being two or more different classes, which will cause the discontinuity of the boundaries[5]. As well as assign weights inversely proportional to the availability of each class in the computation of standard cross-entropy[6].

2 Formulation

In this section, I will discuss the formula for the modified boundary loss in detail, as well as how it differs from the original boundary loss.

2.1 Recap of the original boundary loss

The original boundary loss[4]) takes the form of:

$$\mathcal{L}_B(\theta) = \int_\Omega \phi_G(q) s_\theta(q) dq \tag{1}$$

Where Ω represents the image domain, i.e., every pixel in the image. $\phi_G(q)$ is the function returns the encoded distance information, $\phi_G(q) = -D_G(q)$ if pixel q is in the target region of the ground truth and $\phi_G(q) = D_G(q)$ if pixel q is in the background region of the ground truth. (Notice that $D_G(q)$ represents the distance between pixel q and its nearest point on the contour, which can be precomputed from the ground truth label) $s_\theta(q)$ represents the θ th iteration softmax probability output of the neural network $\in [0, 1]$[4].

By analyzing the formula, we find four potential cases depending on the location of a pixel. These four cases are displayed in Figure 2:

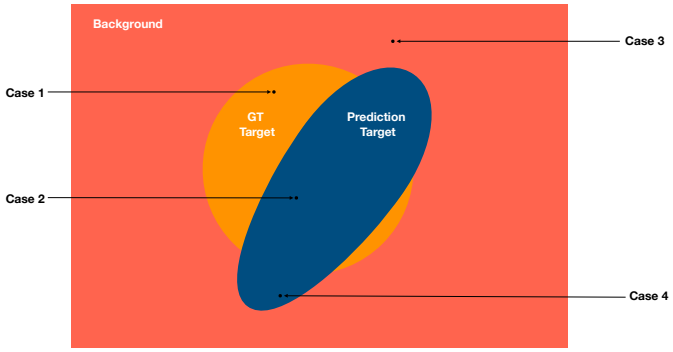

Figure 2: Four cases need to consider

Case 1: where the pixel is in the target region of GT but classified as background in current prediction. In this case, $\phi_G(q) = -D_G(q)$ but $s_\theta(q) \approx 0$, thus this pixel $\approx$ neglected by the formula.

Case 2: where the pixel is in the target region of both the GT and current prediction. In this case, $\phi_G(q) = -D_G(q)$ and $s_\theta(q) \approx 1$, thus this pixel contributes negatively to the result.

Case 3: where the pixel is in the background region of both the GT and current prediction. In this case, $\phi_G(q) = D_G(q)$ but $s_\theta(q) \approx 0$, thus this pixel $\approx$ neglected by the formula.

Case 4: where the pixel is in the background region of GT but classified as target in current prediction. In this case, $\phi_G(q) = D_G(q)$ and $s_\theta(q) \approx 1$, thus this pixel contributes positively to the result.

From the four cases analyzed previously, we summarize the boundary loss as approximately the sum of the encoded distance of all false positives **(Case 4)** minus all correctly predicted target pixels **(Case 2)**. Whereas, the negligence of false negatives yields a problem that the loss function can easily be trapped into a local minimum. This trivial solution with very low gradients can be generated by outputting a probability map with all pixels classified as background[4]. Therefore, the original boundary loss needs to incorporate with a traditional loss function to avoid such a situation.

2.2 Modifications

Here, I proposed several modifications to the existing boundary loss function. The principle of the modifications is enforcing non-trivial outcomes by taking false negatives into account, and computing a ratio between correct predicted targets and incorrectly predicted targets rather than the difference. The new formula takes the form of:

$$\mathcal{L}_{modified}(\theta) = \int_\Omega -\frac{TruePositive}{FalsePositive + FalseNegative + \epsilon} dq \tag{2}$$

This (2) may look a bit different from (1), but we can transform it into a similar fashion:

$$\mathcal{L}_{modified}(\theta) = \int_\Omega -\frac{\phi_{G_t}(q) s_\theta(q)}{\phi_{G_{bg}}(q) s_\theta(q) + [\phi_{G_t}(q) - \phi_{G_t}(q) s_\theta(q)] + \epsilon} dq \tag{3}$$

Same as (1), Ω represents the image domain, i.e., every pixel in the image. $s_\theta(q)$ represents the θ th iteration softmax probability output of the neural network $\in [0, 1]$[4]. Whereas, $\phi_{G_t}(q)$ is the function returns the encoded distance information for GT target pixels only, $\phi_{G_t}(q) = D_G(q)$ if the pixel q is in the target region of the ground truth and $\phi_{G_t}(q) = 0$ otherwise. $\phi_{G_{bg}}(q)$ is the function

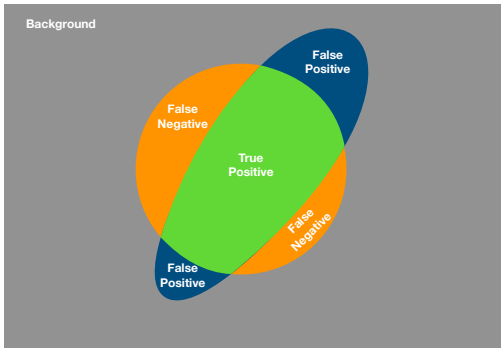

Figure 3: Regions corresponding to False Positive, False Negative and True Positive

returns the encoded distance information for GT background pixels only, $\phi_{G_{bg}}(q) = G_G(q)$ if the pixel q is in the background region of the ground truth and $\phi_{G_{bg}}(q) = 0$ otherwise. ϵ is added to avoid division by zero error. Overall, $\phi_{G_t}(q)s_\theta(q)$ corresponds to $TruePositive$ region, $\phi_{G_{bg}}(q)s_\theta(q)$ corresponds to $FalsePositive$ region and $[\phi_{G_t}(q) - \phi_{G_t}(q)s_\theta(q)] = GT_{target} - TruePositive$, corresponds to $FalseNegative$ region.

It is worth noting that the new loss function takes False Negative regions into account as well, while the original formula only pays attention to the True Positive and False Positive regions in the image. This new loss function will penalize both incorrectly predicted cases (False Negative and False Positive) and award correctly predicted foreground pixels (True Positive) simultaneously. With this modification, the new loss function is way more sensitive to certain weights and biases in the neural network than the original boundary loss, thus the chance of being trapped by a local minimum is reduced. Moreover, the modified loss function also switched to report a ratio instead of the difference between correctly and incorrectly predicted pixels in the original formula, since it would then further increase the sensitivity of weights and biases in the neural network to the output of the loss function without the chance being cancelled out (the numerator and denominator of the modified loss function are inversely related to each other). Finally, the new loss function is set to be correctly predicted pixels over incorrectly predicted pixels, its value increases as the prediction becomes better. A negative sign is added to the front to adapt to the minimization problem and the outcome of the modified loss function will never go above 0.

In the worst-case scenario, none of the pixels are predicted correctly. The modified loss function then has 0 in the numerator and ∞ in the denominator with an overall value being slightly less than 0. As the prediction gets better and better over iterations, in the best-case scenario, all pixels are predicted correctly. The modified loss function will reach the global minimum with ∞ in the numerator and 0 in the denominator. The overall value is now being $-\infty$ (ϵ is added to the denominator to avoid division by zero error). This confirms the meaningfulness of the modified loss function.

3 Experiments

In this section, I perform segmentation tasks with the newly proposed modified boundary loss, as well as several baselines. I compare both qualitative and quantitative results to evaluate the performance of the modified boundary loss.

3.1 Datasets

Two medical image datasets were used to perform segmentation tasks, both of which are highly class imbalanced.

ACDC Dataset:[7] This dataset is taken from the ACDC segmentation challenge. The goal of this challenge is to segment the left ventricular endocardium and epicardium as the right ventricular endocardium for both end diastolic and end systolic phase instances. The ACDC dataset includes 150 patients with 3D cine-MR acquired in clinical routine, 100 of them are assigned as training data and available to the public. During the experiment phase, available data are further split into 75 and 25 for training and testing purposes. Each piece of the data contains 3 foreground classes with the vast majority of pixels being the background.

Scoliosis Dataset[8]: This dataset is taken from MICCAI 2019 challenge. The goal of this challenge is to automatically detect scoliosis and my approach to the problem involves segmentation of the vertebrae. The scoliosis dataset contains 609 spine regions Anterior-Posterior X-Ray images with all of the patients showing signs of scoliosis. During the experiment phase, images are split into 481 and 128 for training and testing purposes. Each piece of the data contains 1 foreground class with the vast majority of pixels being background, the foreground pixels (the vertebrae) are discontinued and separated across the entire image.

The comparison and evaluation are mainly focused on the ACDC Dataset[7], as the modified boundary loss generates meaningless outputs on the Scoliosis Dataset[8]. A potential problem that causes the poor performance on the Scoliosis Dataset is due to the discontinuity and separation of the foreground pixels (discussed in the **Limitations** section).

3.2 Baselines

Boundary Loss: The original boundary loss, formula (1) and explanation can be found in section 2.1 Recap of the original boundary loss.

Dice Loss: It is generalized from Dice Similarity Coefficient, which measures the overlapping regions between two sets. The range for Dice Similarity Coefficient is $[0, 1]$ with a larger value corresponding to more overlapping regions. Thus, Dice Loss = 1 - Dice Similarity Coefficient is used for minimization problems. The formula for Dice Loss is:

$$\mathcal{L}_{\boldsymbol{D}}(\theta) = 1 - \frac{2\sum_i^N p_i g_i}{\sum_i^N p_i^2 + \sum_i^N g_i^2} \tag{4}$$

3.3 Comparison

In this section, results from ACDC dataset[7] will be compared with several baselines. The modified boundary loss and all baselines are trained on the same U-Net architecture for 50 epochs. Results from best epochs are shown.

Qualitative evaluation: Qualitative results are displayed in Figure 4. Segmentations from the ground truth label, modified boundary loss, original boundary loss and dice loss are shown respectively. We observe trivial predictions (Example 2 and Example 3) with empty foreground presence frequently when minimizing over the original boundary loss without assistance from a traditional regional integral based loss function. These two examples both have noticeably smaller regions of interest (foreground objects), which correspond to smaller distances to the contours and less sensitivity of the derivatives to the weights and biases. With the enhancement from the modified boundary loss, valid predictions are presented, which verified the intuition behind the modification (take False Negative into account and make it a ratio). It is also notable that the segmentation results from dice loss in most of the cases are noticeably poorly compared with both modified boundary loss and original

boundary loss. They usually have more noise and unclear boundaries. Compare with the baselines, the predictions of the modified boundary loss are closest to the ground truth regardless of some degrees of error still exist. None of them have the noise from the dice loss predictions.

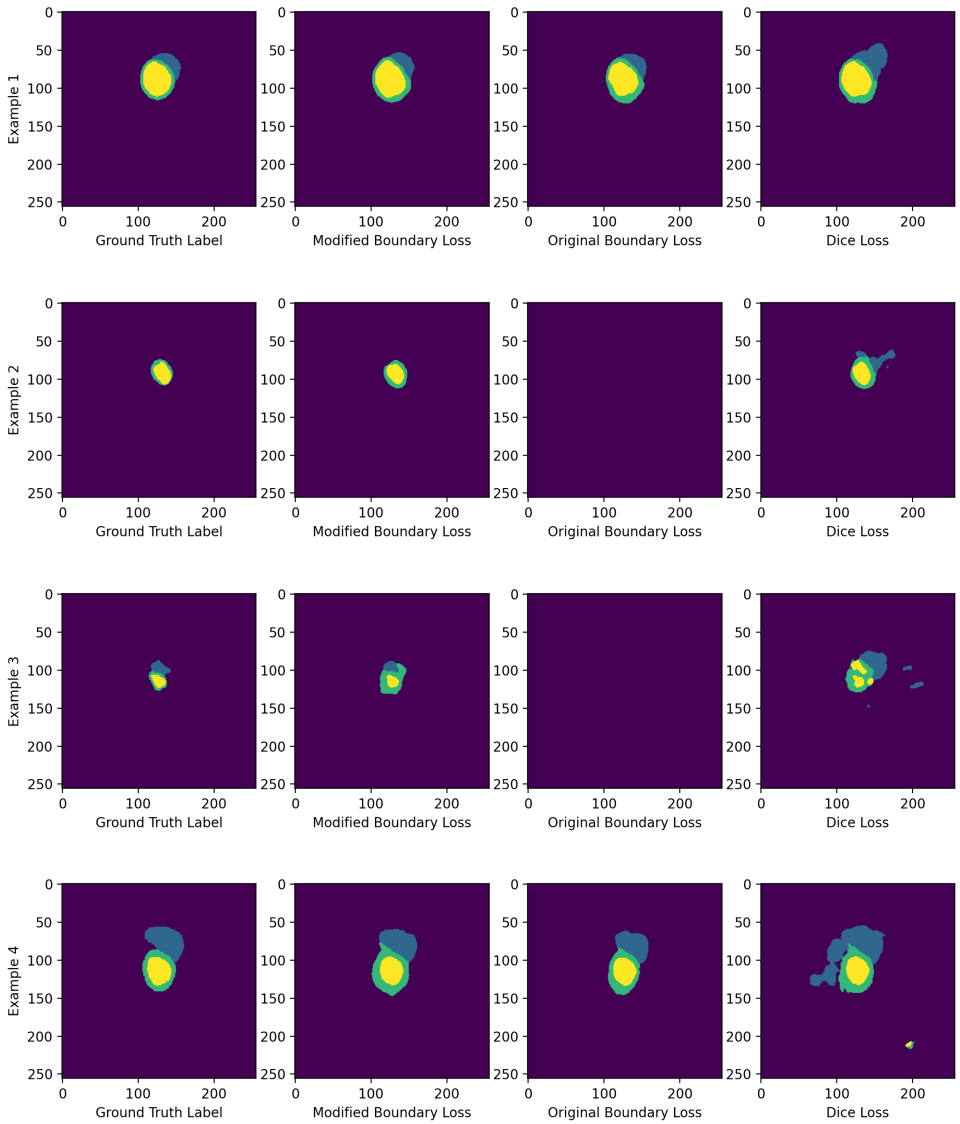

Figure 4: Qualitative results from ACDC dataset[7]

Quantitative evaluation: Table 1 involves quantitative results evaluated in Dice Similarity Coefficient from my experiments using different choices of loss functions. In Table 1, we observe dice loss has the highest overall dice similarity coefficient, with modified boundary loss having slightly less dice similarity coefficient being the second highest and the original boundary loss ranked the last. However, we also observed how noisy and unclear some dice loss predictions are from the qualitative results discussed above. Furthermore, one of the reasons for the overall dice similarity coefficient of the original boundary loss being much lower than the other two loss functions is the frequent presence of empty predictions.

Table 1: Quantitative results from ACDC dataset[7]

	Dice Similarity Coefficient (Train)	Dice Similarity Coefficient (Test)
Modified Boundary Loss	0.8460	0.8379
Original Boundary Loss	0.7981	0.7681
Dice Loss	**0.8657**	**0.8496**

Note: Experiments on the Scoliosis dataset[8] is discussed in the **Limitations** section below.

4 Conclusions

In this paper, I proposed some modifications to the current boundary loss function to address the issue with segmentation tasks on highly imbalanced datasets over traditional unweighted loss functions. Experiments are performed on two highly imbalanced medical image datasets to compare their performance with two baselines (including the original boundary loss and dice loss). The newly proposed modification is able to produce meaningful predictions without the guidance from a traditional loss, while the original boundary loss can be trapped into local minimums and generate trivial solutions in those cases. Qualitative results also suggest the modified boundary loss has higher visual quality and agreement with the ground truth than the other two baselines on extremely biased datasets.

4.1 Limitations

The experiment on the Scoliosis dataset[8] shows the modified boundary loss performs poorly in some settings, i.e., where a single foreground class contains too many relatively large instances separated around the entire image. Figure 5 displayed one example of the ground truth distant map from the Scoliosis dataset. In this example, the target is formed by 17 separated vertebrae corresponding to 17 dark quadrilaterals. The segmentation network ends up producing predictions that have extremely low dice similarity coefficients to the ground truths. One explanation to the failure of the Scoliosis dataset is the distance map is way more complicated in this case, there are significantly more pixels that have short distances to contour compared with other datasets. Therefore, the loss function is less/insensitive to too many of the weights and biases in the neural network and is much easier to be trapped into a local minimum. In these cases, traditional regional loss functions, such as dice loss and binary cross-entropy, can most of the time produce extraordinary outputs (may have unclear boundaries and noises in some of the cases). The alternative solution to this problem is fairly simple, classifying each vertebrae as a different foreground object (class) should be able to resolve the problem.

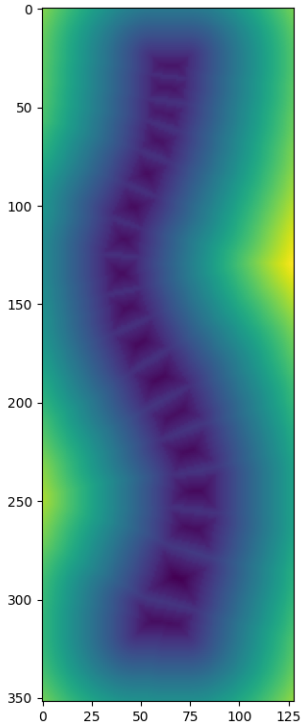

Figure 5: Sample distant map from Scoliosis dataset[8]

Moreover, the modified boundary loss sometimes can still produce trivial results when the foreground objects are extremely tiny, even though the frequency is way lower than the original boundary loss. Dealing with these extreme cases can be a good direction for future work on this topic.

References

[1] 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.

[2] Wei Dai, Nanqing Dong, Zeya Wang, Xiaodan Liang, Hao Zhang, and Eric P Xing. Scan: Structure correcting adversarial network for organ segmentation in chest x-rays. In *Deep learning in medical image analysis and multimodal learning for clinical decision support*, pages 263–273. Springer, 2018.

[3] Hugo J. Kuijf. Wmh segmentation challenge. 2017.

[4] Hoel Kervadec, Jihene Bouchtiba, Christian Desrosiers, Eric Granger, Jose Dolz, and Ismail Ben Ayed. Boundary loss for highly unbalanced segmentation. In *International conference on medical imaging with deep learning*, pages 285–296. PMLR, 2019.

[5] Minh Ôn Vû Ngoc, Yizi Chen, Nicolas Boutry, Joseph Chazalon, Edwin Carlinet, Jonathan Fabrizio, Clément Mallet, and Thierry Géraud. Introducing the boundary-aware loss for deep image segmentation. In *British Machine Vision Conference (BMVC) 2021*, 2021.

[6] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision*, pages 2980–2988, 2017.

[7] Acdc - segmentation. Pierre-Marc Jodoin Videos Images Theory and Analytics Laboratory, University of Sherbrooke, Alain Lalande Pole Imagerie Medicale et Sante, Faculté de Medecine, University of Bourgogne, Olivier Bernard Creatis laboratory, University of Lyon, 2017.

[8] Hongbo Wu, Chris Bailey, Parham Rasoulinejad, and Shuo Li. Automatic landmark estimation for adolescent idiopathic scoliosis assessment using boostnet. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 127–135. Springer, 2017.