Study of Loss Functions for Semantic Segmentation

Ved Prakash Dubey

Abstract

Image Segmentation has been an active field of research as it has a wide range of applications, ranging from automated disease detection to self driving cars. Various papers have come up with different objective loss functions used in different cases such as sparse data, biased data etc. In this poster, I have summarized some of the well-known loss functions widely used for Semantic Segmentation and listed out the cases where their usage can help in fast and better convergence of a model. The dataset used is a combination of the NLM Montgomery County and China Set - The Shenzhen set - Chest X-ray Database upon which we have to perform the task of binary semantic segmentation [1].
The associated code can be found here: https://github.com/VedPDubey/Study-of-Loss-Functions-for-Semantic-Segmentation
Index Terms—Computer Vision, Image Segmentation, Medical Image, Loss Function, Healthcare, Lung Segmentation, Deep Learning

Introduction

The choice of loss/objective function is extremely important while designing complex image segmentation based deep learning architectures as they instigate the learning process of algorithm. Therefore, researchers have experimented with various domain specific loss function to improve results for their datasets.
In this paper we have summarized fifteen such segmentation based loss functions that have been proven to provide state of art results in different domains. These loss function can be categorized into 4 categories: Distribution-based, Region-based, Boundary-based, and Compounded [2]. I personally got invested into surveying these loss functions when I noticed in a previous research project of mine that using a region-based loss function was considerably my model's accuracy and IoU score.

Of these I covered three categories and five loss functions namely – Dice-Binary Cross-Entropy Loss, Jaccard (IoU) Loss, Focal Loss, Tversky Loss, Focal-Tversky Loss. The performance of the aforementioned loss functions on the NML Chest X-ray datasets has been compared to and stated in the poster in the form of Intersection-over-Union (IoU) score, and pixel accuracy metrics.

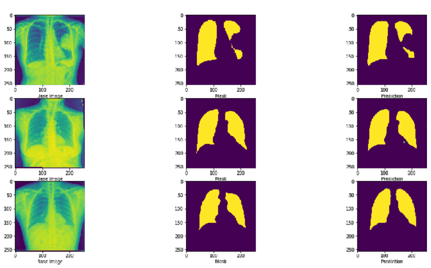

Fig 1: Sample Lung Segmentation from X-ray. [2]
You can see the lungs have been segmented out as masks.

Methodology

I implemented a simple 2D UNet architecture for segmentation with 10 convolutional layers as the encoder and 8 convolutional transpose layers for the decoder [3]. The two datasets used were part of this analysis [1]. The datasets come from Shenzhen and Montgomery respectively. The China Set consists of 336 cases with manifestation of tuberculosis, and 326 normal cases. The other dataset is the Montgomery dataset where X-ray images have been acquired from the tuberculosis control program of the Department of Health and Human Services of Montgomery County, MD, USA. This set contains 138 posterior-anterior x-rays, of which 80 x-rays are normal and 58 x-rays are abnormal with manifestations of tuberculosis. All images are de-identified and available in DICOM format. For training I have used a batch size of 16 and the Adam optimizer with a learning rate of 2e-4 for 50 epochs. I started out with an initial split of 60-40 but settled on 80-10-10 as train,validation and test set respectively due to better scores and more optimal convergence. Ultimately the model was trained using 5 loss functions - Dice-BCE, Jaccard Loss , Focal Loss, Tversky Loss and Focal-Tversky Loss and I evaluated them on the basis of well known evaluation metrics: Intersection-over-Union score and Binary Accuracy.

Table 1: Types of Semantic Segmentation Loss Functions

Type	Loss Function
Distribution-based Loss	Focal Loss
Region-based Loss	Tversky Loss Focal-Tversky Loss Jaccard (IoU) Loss
Compound Loss	Dice-BCE Loss

Loss functions define how neural network models calculate the overall error from their residuals for each training batch. The default choice of loss function for segmentation and other classification tasks is Binary Cross-Entropy (BCE). In situations where a particular metric, like the Dice Coefficient or Intersection over Union (IoU), is being used to judge model performance, competitors will sometimes experiment with loss functions that derive from these metrics - typically in the form 1 - f(x) where f(x) is the metric in question.

The types of loss functions used in the experiment have been stated as follows :

Dice-BCE Loss –
This loss combines Dice loss with the standard binary cross-entropy (BCE) loss that is generally the default for segmentation models. Combining the two methods allows for some diversity in the loss, while benefitting from the stability of BCE. The equation for multi-class BCE by itself will be familiar to anyone who has studied logistic regression [2]:

$$J(\mathbf{w}) = \frac{1}{N}\sum_{n=1}^{N} H(p_n, q_n) = -\frac{1}{N}\sum_{n=1}^{N}\left[y_n \log \hat{y}_n + (1-y_n)\log(1-\hat{y}_n)\right],$$

Jaccard/Intersection over Union (IoU) Loss –
The IoU metric, or Jaccard Index, is similar to the Dice metric and is calculated as the ratio between the overlap of the positive instances between two sets, and their mutual combined values [4]:

$$J(A,B) = \frac{|A\cap B|}{|A\cup B|} = \frac{|A\cap B|}{|A|+|B|-|A\cap B|}.$$

Focal Loss –
Focal Loss was introduced by *Lin et al* of Facebook AI Research in 2017 as a means of combating extremely imbalanced datasets where positive cases were relatively rare. In practice, the researchers used an alpha-modified version of the function and that is what has been used in this implementation [5].

$$\mathrm{FL}(pt) = -\alpha t(1-pt)\ \gamma\ \log(pt)$$

Here, $\gamma > 0$ and when $\gamma = 1$ Focal Loss works like Cross-Entropy loss function. Similarly, α generally range from [0,1], It can be set by inverse class frequency or treated as a hyper parameter.

Tversky Loss –
Tversky index (TI) can also be seen as an generalization of Dices coefficient. It adds a weight to FP (false positives) and FN (false negatives) with the help of β coefficient [6].

$$TI(p,\hat{p}) = \frac{p\hat{p}}{p\hat{p} + \beta(1-p)\hat{p} + (1-\beta)p(1-\hat{p})}$$

Focal-Tversky Loss -
Similar to Focal Loss, which focuses on hard example by down-weighting easy/common ones. Focal Tversky loss also attempts to learn hard-examples such as with small ROIs(region of interest) with the help of γ coefficient as shown below [7]:

$$\mathrm{FTL} = \sum_c (1 - \mathrm{TI}_c)^{\gamma}$$

here, TI indicates Tversky Index, and γ can range from [1,3].

Results

By using roughly 700 annotated segmented examples, I achieved an optimal IoU score of 0.914 and a Binary Accuracy score of 0.982 using Focal Tversky Loss. Dice-BCE and Jaccard Loss with an IoU score of 0.907 came second to the best results. As of Binary Accuracy, Focal Tversky Loss outperformed all other loss functions at 0.9828, however the scores for all loss functions in this metric were roughly similar to around 0.98, I have also observed similar outcomes in other research where Focal Tversky loss and Tversky loss generally gives optimal results with right parameter values. However what's noticeable here is that Tversky Loss performed slightly worse than most other metrics probably owing to the fact that the dataset was inherently balanced leading to better performance in other metrics namely Dice-BCE and Jaccard Loss.

Table 2: Comparison of mentioned Loss Functions on basis of Intersection-Over-Union and Binary Accuracy scores

Loss Functions	Evaluation Metrics	
	IoU Score	Binary Accuracy
Dice-BCE Loss	0.907812499999999	0.982588768005371
Jaccard Loss	0.907812499999999	0.981801509857177
Focal Loss	0.906249999999999	0.982077360153198
Tversky Loss	0.895312499999999	0.980249404907226
Focal-Tversky Loss	**0.914062499999999**	**0.982855081558227**

Conclusion

Loss functions play an essential role in determining the model performance. For tasks such as segmentation, it's not possible to decide on a universal loss function. The majority of the time, it depends on the dataset properties used for training, such as distribution, skewness, boundaries, etc. None of the mentioned loss functions have the best performance in all the use cases. However, we can say that highly imbalanced segmentation works better with focus based loss functions. Similarly, a compound loss such as dice binary-cross entropy works best across most use cases making it a well-rounded off loss function, whereas mildly skewed data-sets can work around smoothed or generalized dice coefficient.

Acknowledgements

[1] Jaeger S, Candemir S, Antani S, Wáng YX, Lu PX, Thoma G. Two public chest X-ray datasets for computer-aided screening of pulmonary diseases. Quant Imaging Med Surg 2014;4(6):475-477. doi: 10.3978/j.issn.2223-4292.2014.11.20
[2] Ma Jun. Segmentation loss odyssey. arXiv preprint arXiv:2005.13449, 2020
[3] Olaf Ronneberger, Philipp Fischer, Thomas Brox. U-Net: Convolutional Networks for Biomedical Image Segmentation. arXiv:1505.04597
[4] Jeroen Bertels, Tom Eelbode, Maxim Berman, Dirk Vandermeulen, Frederik Maes, Raf Bisschops, Matthew Blaschko Optimizing the Dice Score and Jaccard Index for Medical Image Segmentation: Theory & Practice. arXiv:1911.01685
[5] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, Piotr Dollár. Focal Loss for Dense Object Detection. arXiv:1708.02002
[6] Seyed Sadegh Mohseni Salehi, Deniz Erdogmus, Ali Gholipour. Tversky loss function for image segmentation using 3D fully convolutional deep networks. arXiv:1706.05721
[7] Nabila Abraham, Naimul Mefraz Khan. A Novel Focal Tversky loss function with improved Attention U-Net for lesion segmentation. arXiv:1810.07842