Boundary Loss for Breast Ultrasound Segmentation

Ambroise Odonnat*
Master MVA
ENS Paris-Saclay
ambroise.odonnat@ens-paris-saclay.fr

Manh-Dan Pham*
Master MVA
ENS Paris-Saclay
manh-dan.pham@ens-paris-saclay.fr

Abstract

Deep Learning methods have led to great improvements in medical image segmentation tasks. Most state-of-the-art methods in segmentation rely on Convolutional Neural Networks (CNN) trained through loss functions such as the Dice loss or the cross-entropy loss. However, as these functions are based on regional integrations, they struggle with highly unbalanced segmentation. Hence, we are reviewing the paper - written by Kervadec et al. - presenting the *Boundary Loss*, a new function loss based on a distance metric on the space of contours, which alleviates this issue. Furthermore, we have implemented several parts of the paper and experimented the effect of the boundary loss for an unbalanced segmentation of noisy ultrasound images.

1 Introduction

Medical image segmentation is the process of extracting an object of interest, such as an organ or a tumor, from a medical image, in 2D, 3D or 4D. This process can be done manually, semi-automatically or fully-automatically ([10]). More precisely, it consists in assigning every pixel of an image to a class, often chosen as the foreground (object of interest) and the background (unrelevant information for the specific task). The desired property of the segmentation is to assign all pixels of the objects of interest to the same class, namely the foreground.

Recently, Deep Learning architectures have shown great performances in segmentation task and, more importantly, in medical image segmentation where illumination, volume effect and modality dependent artifacts are additional issues to tackle. The current state-of-the-art model in a number of 2D and 3D segmentation tasks is the UNet (or its more recent versions) proposed by Ronneberger et al. in [12]. It heavily relies on Convolution Neural Network (CNN) layers for their capacity to extract local features and regions of interest in an image. Deep Learning architectures are trained through backpropagation by using loss functions such as Dice loss or cross-entropy loss that are based on regional integration. However, performances and stability of the training are affected when dealing with highly imbalanced segmentation tasks. In particular, when the size of the background is several orders of magnitude larger than the size of the foreground, computation of the loss functions involves ground terms with very different orders of magnitude which biases the gradient of the loss. A partial solution to solve this issue was to weight the terms in the loss, rarer terms (e.g. small foreground) having a higher weights.

To alleviate this issue, Kervadec et al. propose in [8] a boundary loss that is a metric distance in the space of contours, not on regions. Rather than computing unbalanced integrals, the boundary loss integrates on the boundary between regions and thus does not depend on the unbalance between foreground and background. It also provides additional information complementary to the regional

*Equal contribution

Preprint. Under review.

loss. Moreover, the authors mention as future work the evaluation of the regularizer effect of the boundary loss on segmentation task of noisy images.

Contributions The main purpose of this study is to review the paper [8] by Kervadec et al. We will describe the motivation and mathematical background of their work. We will then implement most parts of the paper to study the effect of the boundary loss in an application with challenging imaging noise such as ultrasound imaging (US) [1]. Finally, we will discuss about the limitations of the paper and future work.

2 Mathematical Background

2.1 Preliminaries

Common loss functions for CNN-based segmentation tasks include the Dice loss ([3]) and the cross-entropy ([14]). These functions rely on regional summation over the segmentation regions of differentiable functions. In practise, those segmentation regions are defined by probability softmax outputs which enables the use of standard optimizers such as the Stochastic Gradient Descent (SGD) proposed in [11].

Cross-Entropy Loss The cross-entropy loss is a measure of the difference between two probability distributions for a given set or random variable. In binary classification, denoting y the softmax output of the model for the input x, representing the probability that x is in class 1 and $\bar{y}$ the true label of x, a standard formulation of the binary cross-entropy $\mathcal{L}_{CE}$ is $\mathcal{L}(y, \bar{y}) = -y \log(\bar{y}) + (1 - y) \log(1 - \bar{y}))$. It can be generalized to multi-class classification in a straightforward manner. For segmentation tasks, y is typically the set of pixels predicted as foreground or more precisely the probability distribution of membership to the foreground while $\bar{y}$ is the ground-truth. An implicit assumption in the CE loss is the equal importance of every sample and class. It can be easily seen that in the case of highly unbalanced segmentation, it leads to finding a decision boundary biased towards the dominant class ([4]). A solution would be to use a weighted CE loss instead defined by: $\mathcal{L}_{CE} = \mathcal{L}(y, \bar{y}) = -w_{-} y \log(\bar{y}) + w_{+}(1 - y) \log(1 - \bar{y}))$ where the weights w_{-} and w_{+} are used to give more or less importance to a class. Several methods have been developed to alleviate those limitations to the detriment of the training stability and efficiency ([13],[12]).

Dice Coefficient The Dice overlap coefficient was later introduced as a regional segmentation loss. Outperforming the CE loss on unbalanced segmentation tasks, it reflects the size and the localization of segmented regions compared to the pixel-wise accuracy computed by the CE loss. Denoting y the set of pixels predicted as foreground and $\bar{y}$ the true foreground pixels, a standard formulation of the Dice coefficient is $\frac{2|y \cap \bar{y}|}{|y| + |\bar{y}|}$. Due to this formulation, the Dice loss may encounter issues with very small structures. Indeed, in case of highly unbalanced segmentation, pixel mis-classification has huge impact on the loss and hence degrade the stability of the training. Several methods have been developed to alleviate those limitations, coming with a high computational cost or a need to tune parameters ([7], [1]).

2.2 Boundary Loss

In their paper, Kervadec et al. propose a boundary loss as a distance metric on the space of contours instead of the space of regions. They show that it can mitigate the issues encountered with usual CNN losses in highly unbalanced segmentation tasks. Moreover, they argue that it provides complementary information to regional losses. However, the computation of such a loss is challenging as the representation of boundary points is not immediate on regional softmax outputs of a CNN. Inspired from graph-based optimization techniques, the authors avoid differential computation on contour space and express the boundary loss as a sum of linear functions on the regional softmax probability outputs of the model. The boundary loss can thus be combined with any deep learning networks for medical segmentation.

[1] Link to code: https://github.com/AmbroiseOdonnat/BoundaryLoss

2.2.1 Formulation

In this part, we will detail the mathematical formulations of the proposed loss. We keep the same notations introduced in [8].

- $I : \Omega \subset \mathbb{R}^{2,3} \mapsto \mathbb{R}$ training image with domain Ω.
- $g : \Omega \mapsto \{0, 1\}$ a binary ground-truth segmentation of the image with $g(p) = 1$ if the pixel/voxel p is in the target region $G \subset \Omega$ (the foreground), $g(p) = 0$ otherwise i.e $p \in \Omega \setminus G$ (the background)
- $s_\theta \mapsto [0, 1]$ the softmax probability output of a deep segmentation network
- $S_\theta := \{p \in \Omega | s_\theta(p) \geq \delta\}$ the corresponding segmentation region for some threshold δ

The idea of the paper is to build a distance metric $\mathcal{D}(\partial G, \partial S_\theta)$ where ∂G is the boundary of ground-truth region, ∂S_θ is the boundary of the segmentation region output of the model. This distance would have the nice property to be a distance metric on the space of contours in Ω.

Differential Formulation The distance between two contours $\partial G, \partial S$ can be denoted as follows:

$$\text{Dist}(\partial G, \partial S) = \int_{\partial G} ||y_{\partial S}(p) - p||_2 \tag{1}$$

where $y_{\partial S}(p)$ is the projection of $p \in \partial G$ on the contour ∂S. The point $y_{\partial S}(p)$ can be seen as the intersection between ∂S and the line normal to ∂G at the point $p \in \partial G$ (see Figure 1a).

Integral Formulation The distance on the space of contours is a differential variation as it involves evaluation of a boundary change. Unfortunately, this differential framework is not well suited to ∂S_θ, the boundary segmentation region defined by the CNN output. Indeed, to obtain ∂S_θ, an *argmax* or thresholding operation is needed, which breaks the differentiability of the operation such that the network can't be optimized anymore through backpropagation. However, the differential formulation can be approximated in the integral framework as follows:

$$\text{Dist}(\partial G, \partial S) \approx 2 \int_{\Delta S} \mathcal{D}_G(q) dq \tag{2}$$

where ΔS is the regions between the two contours ∂S and ∂G, and $\mathcal{D}_G : \Omega \mapsto \mathbb{R}^+$ is a distance map with respect to boundary ∂G, i.e $\mathcal{D}_G(q) = ||q - z_{\partial G}(q)||_2$ with $z_{\partial G}(q) \in \partial G$ the nearest point from $q \in \Omega$ (see Figure 1b). The relation 2 comes from the fact that, using the definition of $y_{\partial S}$, for all $p \in \partial G$, for all q in the line $[p, y_{\partial S}(p)]$, we have $z_{\partial G}(q) = p$ and $\mathcal{D}_G(q) = ||q - p||_2$. Hence, using the variable change $q \mapsto u = \mathcal{D}_G(q)$, we have the following equation linking $\mathcal{D}_G$ and $y_{\partial S}$:

$$\int_p^{y_{\partial S}(p)} 2\mathcal{D}_G(q) dq = \int_0^{||y_{\partial S}(p) - p||_2} 2u du = ||y_{\partial G}(p) - p||_2^2$$

The relationship between differential and integral formulations of distance between contours is illustrated in the Figure 1.

Loss Formulation Using 2, we have:

$$\frac{1}{2}\text{Dist}(\partial G, \partial S) = \int_{\Delta S} \mathcal{D}_G(q) dq \tag{3}$$

$$\frac{1}{2}\text{Dist}(\partial G, \partial S) = \int_G \mathcal{D}_G(q) dq + \int_S \mathcal{D}_G(q) dq - \int_{S \cap G} \mathcal{D}_G(q) dq \tag{4}$$

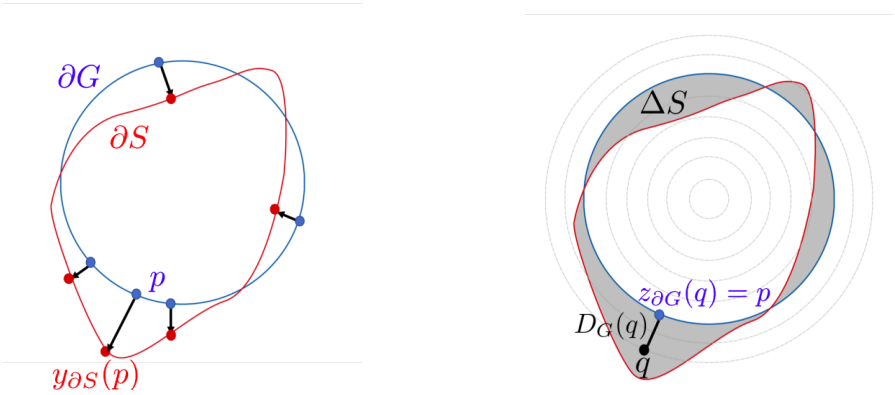
$$\frac{1}{2}\text{Dist}(\partial G, \partial S) = \int_S \Phi_G(q) dq - \int_G \Phi_G(q) dq \tag{5}$$

$$\frac{1}{2}\text{Dist}(\partial G, \partial S) = \int_\Omega \Phi_G(q) s(q) dq - \int_\Omega \Phi_G(q) g(q) dq \tag{6}$$

with $\Phi_G(q) = -\mathcal{D}_G(q)$ if $q \in G$, $\Phi_G(q) = \mathcal{D}_G(q)$ otherwise, $s(q) = 1$ if $q \in S$, 0 otherwise, $g(q) = 1$ if $q \in G$, 0 otherwise.

Replacing the function s by the softmax probability outputs of the network s_θ, and ignoring the right term in 6 as it is independent of network parameters, we obtain the boundary loss which approximates $\text{Dist}(\partial G, \partial S_\theta)$ up to a constant:

$$\boxed{\mathcal{L}_B(\theta) = \int_\Omega \Phi_G(q) s_\theta(q) dq} \tag{7}$$



(a) Differential

(b) Integral

Figure 1: Relationship between differential and integral approaches for evaluating boundary variations

Analysis The authors precise that it is important to use the boundary loss in conjunction with a regional loss as the global optimum in the boundary loss is achieved when softmax predictions correspond exactly to the ground truth. It corresponds to a strictly negative value and very low gradients thus causing the loss to be near a saddle point and impact the training stability. This is the reason why they also use a regional loss such as the cross-entropy: at the beginning, the regional loss guides training and avoid trivial solutions and saddle points while after several epochs, the boundary loss has more impact and influence.

3 Materials and Methods

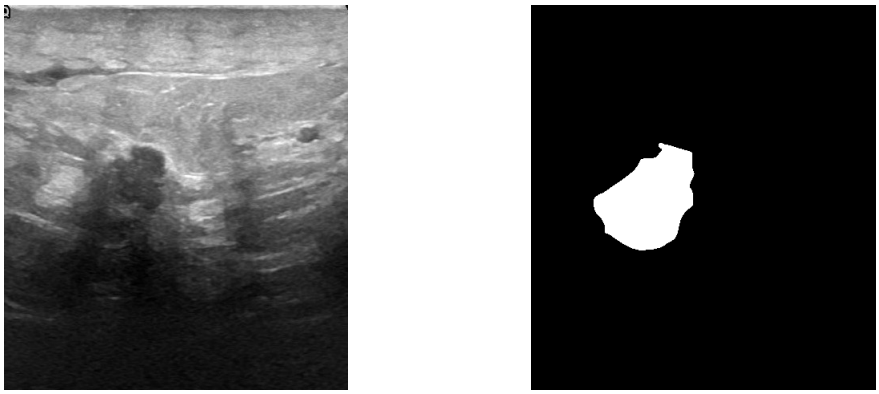
3.1 Datasets and Preprocessing

In their paper, Kervadec et al. ([8]) evaluate the effect of the boundary loss on MRI images for an highly unbalanced task. However, MRI images tend to be low noise images. In this work, we investigate the behaviour of the boundary loss in a noisy setting such as ultrasound imaging. For this purpose, we use the Breast Ultrasound Dataset proposed by Al-Dhabyani et al. [2]. It consists in 780 breast ultrasound images collected in 2018 from 600 women between 25 and 75 years old collected. The average image size is of 500×500 pixels. In the Figure 2, an example of a breast ultrasound image can be seen. In this image (right), there is a malignant tumor which is segmented in the ground truth segmentation mask (left).

In order to stay in the unbalanced framework described in the original paper, we only keep images with less than 5% of tumor surface, which resulted in a total of 425 images for our experiments. The distribution of the proportion of tumorous surface in the 425 images used in our experiment can be seen in Figure 3. We can see that the tumorous surface represent a very small amount of the surface of most images which corresponds to a highly unbalanced segmentation situation: the foreground (tumors) is underrepresented compared to the background (non tumor).

3.2 Methods

To be able to evaluate the effect of the boundary loss on our noisy segmentation task, we compared several strategies. As stated in the paper, the boundary loss presented in 2.2 must be combined with a traditional regional loss for optimization purposes. As we relied on CNN architectures in binary



(a) Input Image (b) Corresponding Ground Truth Mask

Figure 2: Input Image and Ground Truth Mask Annotation from Breast Ultrasound Dataset

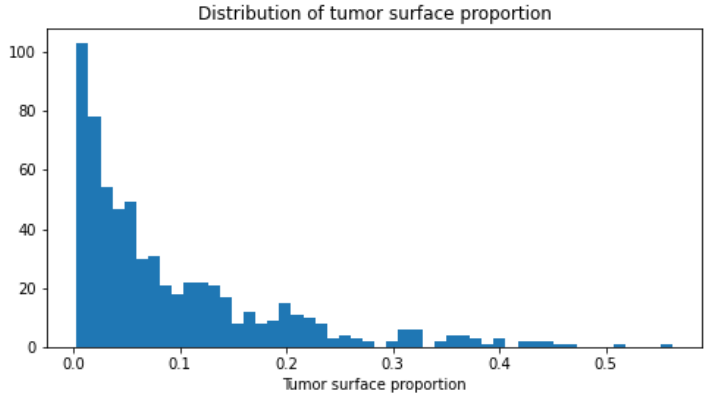


Figure 3: Distribution of the Proportion of Tumorous Surface in the Breast Ultrasound Dataset

segmentation, we used the binary cross-entropy loss (2.1) denoted $\mathcal{L}_{CE}$. Then, we implemented three types of combination:

- **Linear combination** $\mathcal{L} = \mathcal{L}_{CE} + \alpha\mathcal{L}_B$ with a fixed α
- **Linear scheduling** $\mathcal{L} = \mathcal{L}_{CE} + \alpha\mathcal{L}_B$ with α linearly increasing from 0 to 1
- **Rebalance** $\mathcal{L} = (1-\alpha)\mathcal{L}_{CE} + \alpha\mathcal{L}_B$ with α linearly increasing from 0 to 1

3.3 Implementation Details

In order to evaluate the boundary loss on an unbalanced segmentation task with challenging noise, we use the Breast Ultrasound Dataset which contains, after preprocessing, 425 breast ultrasound images. First, images and their corresponding mask annotations are resized to 256×256.

Dataset Split The dataset is then split into 80% of images for training (340 US images) and 20% of images for validation (85 US images).

Architecture and Training We used a U-net-based model ([12]) taken from the *segmentation-models-pytorch* library ([6]). The encoder network is a ResNet-34 backbone [5]. The network is trained on 20 epochs using a batch size of 16 and the Adam optimizer ([9]) with a learning rate of 10^{-3}. We used the losses described in 3.2.

Metric For the evaluation, we employ the common Dice Similary Coefficient (DSC) (2.1), also known as F1-score to evaluate the performances of our models which is defined as follow :

$$DSC = 2\frac{|A_{seg} \cap A_{GT}|}{|A_{seg}| + |A_{GT}|}$$

where A_{seg} and A_{GT} are respectively the predicted segmentation mask and the ground truth mask annotation.The Dice Coefficent is particularly adapted for unbalanced segmentation as it measures the pixel-wise agreement between the predicted segmentation and the ground truth.

4 Results

The best Dice Coefficient (DSC) on the validation set of the Breast Ultrasound Dataset are reported for each training strategy in Table 1. We can see that the results are very similar with a slight improvement when it comes to the rebalance strategy. Compared to the Table 2, we can see that our results are similar to those of Kervadec et al.. The fact that the boundary loss does not bring a significant improvement could be explained by the fact that our dataset is too small and less unbalanced than the one the authors used. Indeed, overtfitting occured early in the training process. Moreover, the cross-entropy loss performs well on this dataset which indicates that the segmentation task might be too easy.

Loss		DSC
Cross-entropy	$\mathcal{L}_{CE}$	0.798
Linear combination	$\mathcal{L}_{CE} + \mathcal{L}_B$	0.798
Linear scheduling	$\mathcal{L}_{CE} + \alpha\mathcal{L}_B$	0.796
Rebalance	$(1-\alpha)\mathcal{L}_{CE} + \alpha\mathcal{L}_B$	**0.801**

Table 1: DSC values on Validation Set

Loss	DSC for ISLES	DSC for WMH
$\mathcal{L}_{CE}$	0.608	**0.757**
$(1-\alpha)\mathcal{L}_{CE} + \alpha\mathcal{L}_B$	**0.631**	0.756

Table 2: DSC values achieved **by the authors** on the validation subset of ISLES and WMH datasets

What is more interesting is to visualize the effect of the boundary loss on the predicted segmentation made by the model. The Figure 4 shows some qualitative results of the effect of the boundary loss on validation set. The left columns shows the ground truth segmentation, the middle one shows the predicted segmentation with the cross-entropy training strategy and the right column shows the predicted segmentation with the conjunction of the boundary loss and the cross-entropy during training. We can see that on some examples, the use of the boundary loss enables better segmentation, with less mistake. Our results don't demonstrate a significant benefit of the use of the boundary loss but on some hard examples, it can be useful especially when the cross-entropy fails. Moreover, our experiements demonstrated that the segmented regions have often less connected components when using the boundary loss which could be linked to tumors anatomy.

5 Conclusion and Future Work

Kervadec et al. ([8]) proposed a boundary loss term that can be easily combined with any standard regional loss to deal with highly unbalanced segmentation tasks. Hence, this loss can be implemented with any existing deep learning architecture for N-D segmentation. They showed the benefits of using this new loss in terms of performance and of training stability on two challenging dataset. Moreover, studying the regularizer effect of the boundary loss on applications with challenging imaging noise such as ultrasound imaging was one a the proposed future works of the authors.

In this work, we studied the motivation and mathematical formulation of the proposed loss. We also implemented several parts of the original paper to evaluate the effect of the boundary loss for

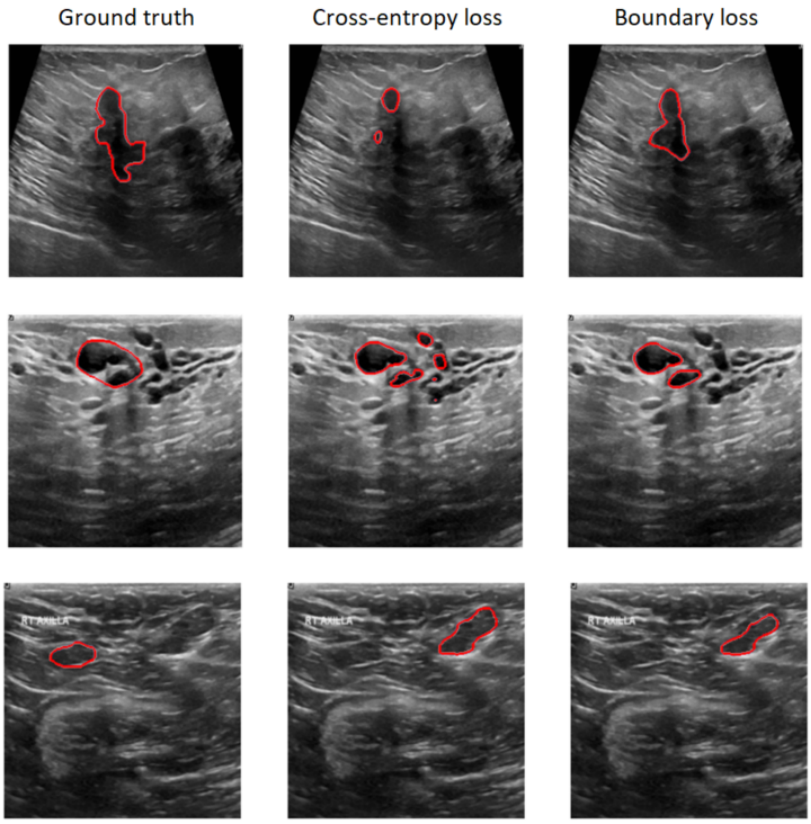


Figure 4: Qualitative Comparison between Predicted Segmentation on Validation Set

unbalanced segmentation of noisy ultrasound images. Our experiments on the breast ultrasound images ([2]) demonstrated qualitatively the benefit of the boundary loss on hard cases. The average results in terms of Dice Similarity Coefficient are similar with a slight advantage for the conjunction of cross-entropy loss and boundary loss during training.

The original work and our study limit to binary classification. An interesting future research direction would be to extend the boundary loss for multi-regions segmentation, with distance maps from multiple complex structures.

References

[1] Nabila Abraham and Naimul Mefraz Khan. A novel focal tversky loss function with improved attention u-net for lesion segmentation, 2018.

[2] Walid Al-Dhabyani, Mohammed Gomaa, Hussien Khaled, and Aly Fahmy. Dataset of breast ultrasound images. *Data in Brief*, 28:104863, 2020.

[3] Jeroen Bertels, Tom Eelbode, Maxim Berman, Dirk Vandermeulen, Frederik Maes, Raf Bisschops, and Matthew B. Blaschko. Optimizing the dice score and jaccard index for medical image segmentation: Theory and practice. In *Lecture Notes in Computer Science*, pages 92–100. Springer International Publishing, 2019.

[4] Jie Cao, Zhe Su, Liyun Yu, Dongliang Chang, Xiaoxu Li, and Zhanyu Ma. Softmax cross entropy loss with unbiased decision boundary for image classification. In *2018 Chinese Automation Congress (CAC)*, pages 2028–2032, 2018.

[5] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Identity mappings in deep residual networks. In *European conference on computer vision*, pages 630–645. Springer, 2016.

[6] Pavel Iakubovskii. Segmentation models pytorch. `https://github.com/qubvel/segmentation_models.pytorch`, 2019.

[7] Davood Karimi and Septimiu E. Salcudean. Reducing the hausdorff distance in medical image segmentation with convolutional neural networks, 2019.

[8] Hoel Kervadec, Jihene Bouchtiba, Christian Desrosiers, Eric Granger, Jose Dolz, and Ismail Ben Ayed. Boundary loss for highly unbalanced segmentation. *Medical Image Analysis*, 67:101851, jan 2021.

[9] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[10] Jianning Li, Marius Erdt, Firdaus Janoos, Ti chiun Chang, and Jan Egger. 1 - medical image segmentation in oral-maxillofacial surgery. In Jan Egger and Xiaojun Chen, editors, *Computer-Aided Oral and Maxillofacial Surgery*, pages 1–27. Academic Press, 2021.

[11] Herbert Robbins and Sutton Monro. A Stochastic Approximation Method. *The Annals of Mathematical Statistics*, 22(3):400 – 407, 1951.

[12] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. *CoRR*, abs/1505.04597, 2015.

[13] Sergi Valverde, Mariano Cabezas, Eloy Roura, Sandra González-Villà, Deborah Pareto, Joan-Carles Vilanova, LLuís Ramió-Torrentà, Àlex Rovira, Arnau Oliver, and Xavier Lladó. Improving automated multiple sclerosis lesion segmentation with a cascaded 3d convolutional neural network approach, 2017.

[14] Ma Yi-de, Liu Qing, and Qian Zhi-bai. Automated image segmentation using improved pcnn model based on cross-entropy. In *Proceedings of 2004 International Symposium on Intelligent Multimedia, Video and Speech Processing, 2004.*, pages 743–746, 2004.