Innovative temporal loss function for segmentation of fine structures in ultrasound images

Innovative temporal loss function for segmentation of fine structures in ultrasound images

Francesco Marzola
Biolab, PolitoBIOMedLab, Department of Electronics and Telecommunications
Politecnico di Torino
Turin, Italy
francesco.marzola@polito.it

Kristen M. Meiburger
Biolab, PolitoBIOMedLab, Department of Electronics and Telecommunications
Politecnico di Torino
Turin, Italy
kristen.meiburger@polito.it

Massimo Salvi
Biolab, PolitoBIOMedLab, Department of Electronics and Telecommunications
Politecnico di Torino
Turin, Italy
massimo.salvi@polito.it

***Abstract*— Over the past few years, there have been significant advancements in deep learning architectures for semantic segmentation. However, the performance of these models heavily relies on the loss function (LF) used during network training. The LF is a crucial component that enables the network to weigh the errors differently based on the segmentation task to be performed. Despite the progress made in designing increasingly complex and deep architectures for semantic segmentation, the LFs used in these models have remained almost unchanged. Accurately segmenting small and fine objects, such as vessel walls (e.g., intima-media complex, IMC) or nerves (e.g., optic nerve), in ultrasound (US) images is still a challenging task. One of the main difficulties is pixel imbalance between the object and the background, which can result in inaccurate segmentation. Additionally, precise and accurate segmentation along the object's edge is crucial for medical diagnosis and treatment. To address these challenges, this paper proposes a new, temporal loss function for semantic segmentation in US images. The idea behind a temporal loss is to enable the network to learn from multiple sources of information simultaneously and to give more emphasis to losses that are more informative at different stages of the training process. The proposed LF considers pixel imbalance between the object and background and enables precise and accurate segmentation along the object's edge. The study aims to demonstrate the effectiveness of the proposed LF by evaluating its performance in segmenting vessel walls in US images.**

Keywords*— *ultrasound imaging; artificial intelligence; fine-structures segmentation; loss function.

I. INTRODUCTION

The segmentation of fine structures in ultrasound (US) images is a challenging task due to the presence of noise, speckles, and other artifacts that can affect the quality of the images. To address this challenge, deep learning models have been proposed for the automated segmentation of US images. However, the performance of these models heavily relies on the loss function (LF) used during training. In recent years, there has been a growing interest in developing innovative deep-learning architectures with hybrid mechanisms to improve their performance in various applications. However, less attention has been paid to the development of novel loss functions that can improve the performance of these models. Traditional loss functions work by measuring the difference between the predicted output of a deep learning model and the true target output. This difference, also known as the "loss", is used to optimize the model parameters during the training process. Traditional loss functions have several limitations, including static weights of individual losses throughout the training process, optimization of a single objective that may not capture complex relationships between input and output, and lack of consideration for the difficulty of the task being performed, treating all parts of the image equally. These limitations can result in suboptimal performance in tasks where certain parts of the input data are more important than others. Dynamic loss functions have emerged as a promising approach to enhance the performance of deep learning models [1]. Unlike traditional static loss functions, dynamic loss functions allow the network to adaptively adjust the weight of different losses during training. This adaptability can lead to improved accuracy and faster convergence during the training process.

In the context of US image segmentation, the application of dynamic loss functions is still relatively unexplored [2]. This study aims to investigate the effectiveness of a novel temporal loss function specifically designed for the segmentation of fine structures in US images. The objective is to improve the accuracy and robustness of the segmentation network by better handling the challenges of US image segmentation, such as pixel imbalance and precise edge detection.

II. BACKGROUND AND RELATED WORKS

In traditional loss functions, several limitations arise. First, they use static weights for individual losses throughout the training process, which may not be optimal for capturing the complex relationships between input and output. Second, they typically optimize a single objective, which may limit the model's ability to handle the intricacies of the segmentation task. Finally, traditional loss functions treat all parts of the image equally, which can result in failures when certain regions are more challenging to segment than others. Recent literature focusing on ultrasound imaging has proposed various loss functions that address these limitations. These functions expand the optimization problem from the region-based losses like Dice, focusing on different aspects of the segmentation like shape [3], structure [4], uncertainty [5], class-imbalance [2], and boundary awareness [6],[7].

This work aims to investigate the effectiveness of a dynamic loss function for the segmentation of fine structures in US images. Specifically, we propose a novel common-support Dice loss and a temporal loss function that allows the network to dynamically adjust the weight of different losses based on the current state of the training and the location of the object to be segmented. Our hypothesis is that a dynamic loss function can help the network better handle the challenges of US image segmentation, such as pixel imbalance and precise edge detection. By utilizing a temporal loss function, we aim to improve the accuracy and robustness of the segmentation of fine structures in US images. We believe that the proposed approach can have significant clinical implications for medical diagnosis and treatment, where accurate segmentation of small structures is crucial.

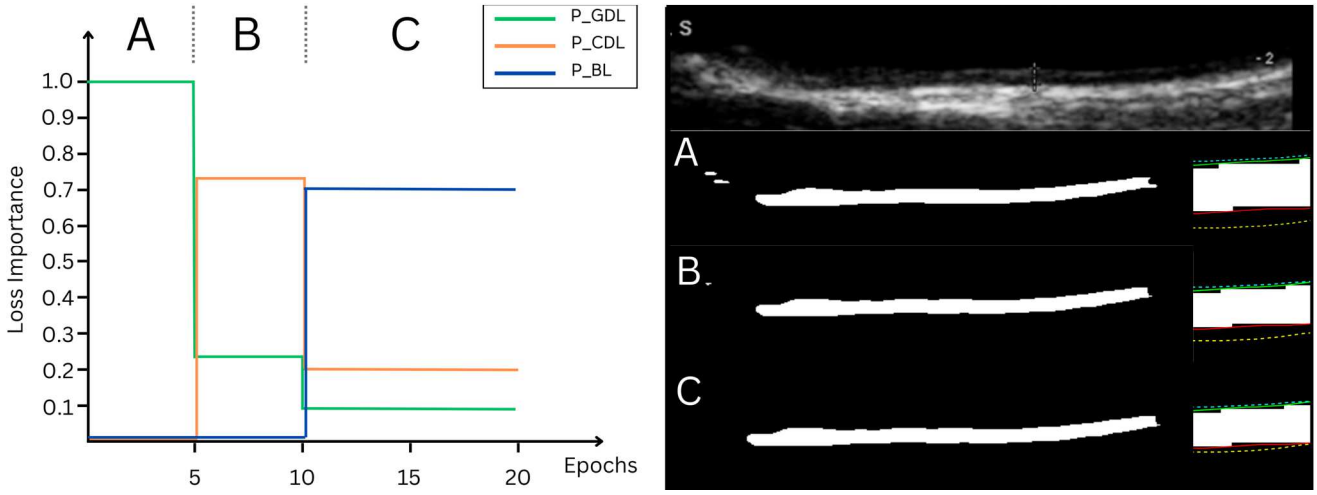

Figure 1 On the left, target value for the importance of each loss during training. On the right, in the top panel the original image, then the segmentation of the network at different stages of the training process identified with letters A, B, and C. On the right of each segmentation there is a zoom on the predicted LI (dotted cyan) and MA (dotted yellow) profiles.

III. MATERIALS AND METHODS

The proposed approach has been implemented for the segmentation of the Intima-Media Complex (IMC) in B-mode US images. In this task, the focus is on the correct segmentation of the Lumen Intima (LI) and Media Adventitia (MA) boundaries to measure their distance, commonly known as the intima-media thickness (IMT), which is measured for cardiovascular risk assessment. For the training phase of the network and the development of the weighting strategy, two open-access datasets [8]-[9] have been used. Together they comprise 2576 B-mode ultrasound image-mask pairs. This dataset was split randomly using an 80/20 ratio to define training and validation sets. In this phase, the output of a semi-automatic segmentation method was adopted as ground truth, which proved to be the best option in [10]. During the evaluation phase of the algorithm, an external multi-centric dataset of 448 images was used to test the proposed approach. The latter dataset was presented in [11]. The segmentation task was modeled as a binary classification problem with classes "Background" (BG) and "Intima-Media-Complex" (IMC). It was performed using a UPerNet [12] with a ConvNeXt-Small encoder [13]. For the training process, the Adam optimizer was set with an initial learning rate of 10^{-5} and the batch size at 4 samples. During training, samples were augmented using geometric and intensity transformations. The network was trained for 11400 iterations across 20 epochs, the metrics calculated on the validation set were logged for each epoch, and the best-performing model was chosen as the final model. The final goal of this work is to understand the optimal loss function for this task; hence the same network was trained with different objective functions. Dice loss, Focal loss [14], and Lovasz loss [15] were used as a reference to compare the performance of the following loss functions:

A. Generalized Dice Loss (GDL)

Modified implementation of the loss function presented in [16]. This loss was first introduced to tackle highly unbalanced segmentation problems in 2D and 3D medical images. Differently from the original implementation, the weighting of the two classes (w_k) was performed for each batch. In this case, N is the total number of pixels in each batch, with the subscript i representing the i-th pixel. C is the number of classes with k indicating the k-th class. The target segmentation is denoted with the letter T, while the network prediction after the SoftMax operation is expressed with the letter P.

$$L_{gdl} = 1 - 2\frac{\sum_{k=1}^{C} w_k \sum_{i=1}^{N} T_i^k P_i^k}{\sum_{k=1}^{C} w_k \sum_{i=1}^{N} (T_i^k)^2 + (P_i^k)^2} \quad (1)$$

$$w_k = \frac{1}{(\sum_{i=1}^{N} T_i^k)^2} \quad (2)$$

B. Common-support Dice Loss (CDL)

The previous loss function was modified by applying a mask, denoted as M, to both the target and the predicted tensor. This mask is only positive for the pixels in the columns where both the tensors have at least one pixel with an IMC label. In this way, we aim to focus on the calculation of the loss in the area of the image where most of the information is present. Moreover, the semi-automatic annotation can be conservative, segmenting only the areas where the profiles are clearly visible. Masking the segmentation avoids penalizing the network when it predicts a segmentation on a wider area with respect to the ground truth.

$$L_{cdl} = 1 - 2\frac{\sum_{k=1}^{C} w_k \sum_{i=1}^{N} M \cdot T_i^k M \cdot P_i^k}{\sum_{k=1}^{C} w_k \sum_{i=1}^{N} (M \cdot T_i^k)^2 + (M \cdot P_i^k)^2} \quad (3)$$

C. Common-support Boundary Loss (BL)

In the same way, the loss first presented in [17] was adapted using the same common-support approach described for the previous objective function, where D_i^k and P_i^k are the result of the distance transform of the ground truth and predicted segmentation respectively.

$$L_{bl} = \frac{1}{NC} \sum_{k=1}^{C} \sum_{i=1}^{N} M \cdot D_i^k M \cdot P_i^k \quad (4)$$

D. Temporal Implementation

Since different loss functions drive the optimization problem toward different local minima a weighting scheme was adopted to favor differently the three different loss functions

TABLE I. SEGMENTATION RESULTS

Loss	Dice	HD	HD LI	HD MA	IMT Abs. Bias
Dice	0.900 ± 0.057	0.210 ± 0.510	2.804 ± 6.686	3.006 ± 6.497	0.811 ± 0.842
Focal	0.899 ± 0.040	0.184 ± 0.101	2.292 ± 1.723	2.488 ± 1.591	0.830 ± 0.841
Lovasz	0.899 ± 0.040	0.187 ± 0.102	2.452 ± 1.850	2.718 ± 1.839	0.840 ± 0.889
GDL	0.900 ± 0.040	0.189 ± 0.104	2.478 ± 1.969	2.621 ± 1.913	0.843 ± 0.882
CDL	**0.903 ± 0.036**	**0.182 ± 0.097**	**2.136 ± 1.348**	**2.342 ± 1.415**	**0.805 ± 0.901**
Temporal	0.896 ± 0.064	0.205 ± 0.159	2.526 ± 2.258	2.566 ± 2.060	0.852 ± 0.987

Table 1. Metrics calculated on the Test set for the same network trained with different objective functions. GDL: Generalized Dice Loss. CDL: Common-support Loss. Temporal: dynamically weighted loss. Hausdorff Distances (HD) and Intima Media Thickness (IMT) are expressed in pixels.

during training. The relative importance *P* of each loss over the total loss was fixed and the weights changed each epoch to match the relative importance. To avoid sharp transitions in the total loss function thus preventing exploding or vanishing gradients, the sum of relative importance was fixed to 1.

Empirically, in the initial part of training, the focus was set on the loss with a wider scope to achieve an initial convergence. To do this only the L_{gdl} was used to guide the training of the network ($P_{L_{gdl}} = 1$). Then the weight of the L_{cdl} was raised to focus the training only on the labeled part of the image ($P_{L_{gdl}} = 0.25$; $P_{L_{cdl}} = 0.75$) . This should bring the network to a lower minimum and should bring better performances on the test set since the results are evaluated on the common support as well. Finally, the L_{bl} weight was increased to guide the network towards the minima that better represent the LI and MA boundaries hence leading to a more precise IMT measurement ($P_{L_{gdl}} = 0.1$; $P_{L_{cdl}} = 0.2$; $P_{L_{bl}} = 0.7$). L_{cdl} was introduced at epoch 6, while L_{bl} at epoch 11. Weights were set to tune the contribution of each loss over the total loss to a certain percentage (i.e., the relative importance *P*). The pseudocode for this weighting scheme is the following:

Algorithm: calculation of loss weights

Input: loss values (GDL, CDL, BL)
Output : loss weights

```
On epoch start:
if n_epoch > 10 then
  Tot_Loss= w_GDL * GDL + w_CDL * CDL + w_BL * BL
  w_GDL = GDL * Tot_Loss * P_gdl
  w_CDL = CDL * Tot_Loss * P_cdl
  w_BL  = BL  * Tot_Loss * P_bl

else if n_epoch > 5 then
  Tot_Loss= GDL + CDL
  w_GDL = GDL * Tot_Loss * P_gdl
  w_CDL = CDL * Tot_Loss * P_cdl
  w_BL  = 0
else if n_epoch <= 5 then
  w_GDL = 1; w_CDL = 0; w_BL = 0
  Log GDL, CDL, BL
```

After introducing the boundary loss, the weights were adjusted at each epoch with the same method to preserve the relative importance of each loss.

IV. RESULTS

A. Performance metrics

To validate the performance of the proposed approach we adopted the same approach already used in several works related to IMC segmentation and IMT measurement [9],[10],[18]. For segmentation performance specifically, Dice score and Hausdorff Distance were calculated on the held-out test set. After extracting the LI and MA profile coordinates and measuring the IMT, for each image the IMT absolute bias was calculated as well as the Hausdorff Distance for both the LI and MA profiles separately. These metrics were computed on the common support between all the methods in the comparison. Results are presented in Table 1, distance-based metrics are expressed in pixels. Results of the network trained only with the common support boundary loss are not shown since the network couldn't reach a convergence point.

The network trained with the CDL showed improvements in all the metrics taken into consideration. These improvements are more conclusive in distance-based metrics where the Hausdorff Distance for LI and MA profile is more than 15% lower on average with respect to the second-best performing loss function. The Temporal loss shows decreased performance in all the computed metrics, this result will be further discussed in the next section.

V. DISCUSSION

This work aims to address the challenges of accurately segmenting complex fine structures in ultrasound (US) images, such as the intima-media complex. To achieve this, a new optimized temporal loss function was developed and evaluated against several state-of-the-art loss functions using a publicly available dataset of US images.

The method validation shows mixed results. On one hand, the CDL outperformed all the other methods for each of the metrics taken into consideration. On the other, the implementation of the temporal loss suffers from a decrease in performance with respect to the CDL. Introducing the boundary loss during training allowed the network to effectively minimize the error associated with the loss; however, it led to an increase in CDL and GDL. This highlights the sensibility of deep networks to the choice of the objective function, emphasizing the importance of finding the most suitable one for each specific task. This is particularly crucial for medical applications, where accurate segmentation

of small structures can have significant clinical implications. The proposed modification to the generalized Dice loss represents a step forward in this direction.

A possible explanation for the reduced performances of the temporal loss resides in the nature of the LI and MA boundaries. Especially when dealing with multi-centric and multi-device datasets, these boundaries can exhibit different echogenicity levels and present hypo or hyper-echogenic areas depending on the subject's physiology and the chosen acquisition angle and settings. This can lead to stronger overfitting on the training set, which is difficult to spot even using a validation set drawn from the same distribution. This issue would have a smaller impact when using a more relaxed loss function, such as Dice loss and its variants.

The results of this study have significant implications for other medical imaging applications that require accurate and precise segmentation of fine complex structures for diagnosis and treatment planning. Several clinical applications are dependent on a similar task, like optic nerve sheaths diameter measurement to monitor intracranial pressure or pathophysiological evaluation of muscle function segmenting fascicles and aponeuroses.

In conclusion, this paper presents a new optimized loss function for semantic segmentation in US images that addresses the challenges of pixel imbalance and object edge segmentation. The proposed loss function demonstrates superior performance in segmenting small and fine objects compared to existing ones, highlighting its potential for medical diagnosis and treatment. Future developments will focus on incorporating additional losses into the overall objective functions without compromising the training process of the network.

REFERENCES

[1] S. Lu, F. Gao, C. Piao, and Y. Ma, "Dynamic Weighted Cross Entropy for Semantic Segmentation with Extremely Imbalanced Data," in *2019 International Conference on Artificial Intelligence and Advanced Manufacturing (AIAM)*, Dublin, Ireland: IEEE, Oct. 2019, pp. 230–233. doi: 10.1109/AIAM48774.2019.00053.

[2] K. Lee, J. Y. Kim, M. H. Lee, C.-H. Choi, and J. Y. Hwang, "Imbalanced Loss-Integrated Deep-Learning-Based Ultrasound Image Analysis for Diagnosis of Rotator-Cuff Tear," *Sensors*, vol. 21, no. 6, p. 2214, Mar. 2021, doi: 10.3390/s21062214.

[3] Z. Sobhaninia *et al.*, "Fetal Ultrasound Image Segmentation for Measuring Biometric Parameters Using Multi-Task Deep Learning," in *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Berlin, Germany: IEEE, Jul. 2019, pp. 6545–6548. doi: 10.1109/EMBC.2019.8856981.

[4] Y. Fu, J. Chen, and K. Li, "Structure-aware Loss Function for Ultrasound Image Segmentation," in *2021 IEEE International Ultrasonics Symposium (IUS)*, Xi'an, China: IEEE, Sep. 2021, pp. 1–4. doi: 10.1109/IUS52206.2021.9593773.

[5] Y. Xie, H. Liao, D. Zhang, and F. Chen, "Uncertainty-aware Cascade Network for Ultrasound Image Segmentation with Ambiguous Boundary," in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2022*, L. Wang, Q. Dou, P. T. Fletcher, S. Speidel, and S. Li, Eds., in Lecture Notes in Computer Science, vol. 13434. Cham: Springer Nature Switzerland, 2022, pp. 268–278. doi: 10.1007/978-3-031-16440-8_26.

[6] Y. Gong *et al.*, "SCCNet: Self-correction boundary preservation with a dynamic class prior filter for high-variability ultrasound image segmentation," *Comput. Med. Imaging Graph.*, vol. 104, p. 102183, Mar. 2023, doi: 10.1016/j.compmedimag.2023.102183.

[7] D. Mishra, S. Chaudhury, M. Sarkar, and A. S. Soin, "Ultrasound Image Segmentation: A Deeply Supervised Network With Attention to Boundaries," *IEEE Trans. Biomed. Eng.*, vol. 66, no. 6, pp. 1637–1648, Jun. 2019, doi: 10.1109/TBME.2018.2877577.

[8] K. M. Meiburger *et al.*, "Carotid Ultrasound Boundary Study (CUBS): An Open Multicenter Analysis of Computerized Intima–Media Thickness Measurement Systems and Their Clinical Impact," *Ultrasound Med. Biol.*, vol. 47, no. 8, pp. 2442–2455, Aug. 2021, doi: 10.1016/j.ultrasmedbio.2021.03.022.

[9] K. M. Meiburger *et al.*, "Carotid Ultrasound Boundary Study (CUBS): Technical considerations on an open multi-center analysis of computerized measurement systems for intima-media thickness measurement on common carotid artery longitudinal B-mode ultrasound scans," *Comput. Biol. Med.*, vol. 144, p. 105333, May 2022, doi: 10.1016/j.compbiomed.2022.105333.

[10] F. Marzola, K. M. Meiburger, F. Molinari, and M. Salvi, "Exploring the Impact of Learning Paradigms on Network Generalization: A Multi-Center IMT Study," in *2023 24th International Conference on Digital Signal Processing (DSP)*, Rhodes (Rodos), Greece: IEEE, Jun. 2023, pp. 1–5. doi: 10.1109/DSP58604.2023.10167892.

[11] F. Molinari *et al.*, "Ultrasound IMT measurement on a multi-ethnic and multi-institutional database: Our review and experience using four fully automated and one semi-automated methods," *Comput. Methods Programs Biomed.*, vol. 108, no. 3, pp. 946–960, Dec. 2012, doi: 10.1016/j.cmpb.2012.05.008.

[12] T. Xiao, Y. Liu, B. Zhou, Y. Jiang, and J. Sun, "Unified Perceptual Parsing for Scene Understanding," 2018, doi: 10.48550/ARXIV.1807.10221.

[13] Z. Liu, H. Mao, C.-Y. Wu, C. Feichtenhofer, T. Darrell, and S. Xie, "A ConvNet for the 2020s," 2022, doi: 10.48550/ARXIV.2201.03545.

[14] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal Loss for Dense Object Detection," 2017, doi: 10.48550/ARXIV.1708.02002.

[15] M. Berman, A. R. Triki, and M. B. Blaschko, "The Lovász-Softmax loss: A tractable surrogate for the optimization of the intersection-over-union measure in neural networks," 2017, doi: 10.48550/ARXIV.1705.08790.

[16] A. Celaya, A. Diaz, A. Balsells, B. Riviere, and D. Fuentes, "A Weighted Normalized Boundary Loss for Reducing the Hausdorff Distance in Medical Imaging Segmentation," 2023, doi: 10.48550/ARXIV.2302.03868.

[17] H. Kervadec, J. Bouchtiba, C. Desrosiers, E. Granger, J. Dolz, and I. B. Ayed, "Boundary loss for highly unbalanced segmentation," 2018, doi: 10.48550/ARXIV.1812.07032.

[18] F. Marzola, K. Meiburger, F. Molinari, and M. Salvi, "Can multiple segmentation methods enhance deep learning networks generalization? A novel hybrid learning paradigm," in *Medical Imaging 2023: Computer-Aided Diagnosis*, K. M. Iftekharuddin and W. Chen, Eds., San Diego, United States: SPIE, Apr. 2023, p. 39. doi: 10.1117/12.2653394.