Combining loss functions for deep learning bladder segmentation on dynamic MRI

Marc-Adrien Hostin, Augustin C. Ogier, Nicolas Pirró, Marc-Emmanuel Bellemare

Combination of loss functions for bladder segmentation on dynamic MRI using U-Net

1st Hostin Marc-Adrien
Aix Marseille Université
Université de Toulon, CNRS, LIS
Marseille, France
0000-0003-3649-2054

2nd Ogier Augustin C.
Aix Marseille Université
Université de Toulon, CNRS, LIS
Marseille, France
0000-0001-9178-9964

3rd Pirró Nicolas
AP-HM
Digestive Surgery - La Timone Hospital
Marseille, France

4rd Bellemare Marc-Emmanuel
Aix Marseille Université
Université de Toulon, CNRS, LIS
Marseille, France
0000-0001-9913-3097

***Abstract*—Pelvic floor disorders mainly affect women and turn to be a public health issue although their pathophysiology is still poorly understood. As the main concerned organs undergo motions and deformations that are specific to the pathology, dynamic MRI is a now gold standard for radiologists. Unclear organ boundaries, inter-subject variability of organ shapes and pathological deformities make segmentation difficult to perform. To develop an imaging software towards pathology gradation, accuracy of the MRI segmentation of the organ borders is a critical criterion. Automatic methods are not yet accurate enough to replace the mandatory manual segmentation step. Automatic segmentation methods using a fully convolutional neural network (FCN) have been developed, but usually the loss functions used for their training are not sufficiently accurate for organ boundary detection. We propose a loss function dedicated to organ boundary detection to enhance training and therefore improve results accuracy. The method was evaluated on bladder segmentation from dynamic 2D MRI, using a baseline U-Net architecture [1]. The results have shown that the use of our boundary loss function coupled with the Dice loss outperformed existing methods and improved the segmentation accuracy in comparison with the widely used Dice loss.**

***Index Terms*—Image segmentation, Fully convolutional network, Hybrid loss, Distance loss, dynamic MRI, pelvis**

I. Introduction

Pelvic floor disorders affect 50% of women over the age of 50 [2]. Pathophysiology is poorly known, causes are multiple and symptoms may vary from incontinence to sexual dysfunction or pelvic organ prolapse. The current clinical practice for the study of these pathologies is dynamic 2D MRI on the sagittal plane showing the movement of the organs during straining exercises. The segmentation of these images is mandatory for the computer aided quantitative analysis of organ deformations, beyond a simple visual investigation [3]. To date, segmentations are mainly performed manually and are a complex and time-consuming task, as they must be performed on each image of a sequence. In addition, 2D dynamic MRI segmentation is a challenging task because: (1) pelvic organ boundaries are not well contrasted (2) organ deformation along the sequence is significant due to stress exercise. Automatic segmentation methods such as active-contour methods have been developed to perform the segmentation of pelvic organs [4] but these methods require user-intervention which is a time-consuming and subjective process. More recently, deep learning methods (DL) have been developed for medical image segmentation with the emergence of FCN such as the U-Net architecture [1]. U-Net has provided convincing results for pelvic organ segmentation without user-intervention. However, with respect to the pelvic region, studies have mostly been performed using computed tomography (CT) images for the study of prostate cancer [5]. Fewer studies were dedicated to MRI, dealing mainly with axial rather than sagittal plane [6], [7]. To the best of the authors knowledge, we are the first to apply a deep learning method to the segmentation of pathological female pelvis on 2D dynamic MRI.

In DL based pelvic segmentation, boundary aware representations were used only on CT images [8], [9], [10]. In these methods, additional networks were used to integrate the boundary-sensitive aspect, which complicated the network training and the hyperparameter tuning. In [9], the authors used a multiclass ground truth representation to implement the boundary representation, making the network training even more complex. In [10], a non-differentiable loss function was used, which is problematic for the gradient descent process.

Instead of using multiple networks, which is yet possible, we investigated on the loss function of the FCN. Research has been conducted on modules that take into account the distance between segmentation boundaries to complement the information provided by the Dice loss function. In a review [11] shows that various studies have been carried out on loss functions using distance maps. Among the boundary-based loss functions a method using average distance between surfaces (ASD) estimation was found to be performant [12], [13]. It does not require additional modules and does not have a high algorithmic cost which made it a good candidate to improve the pelvic organs boundary detection.

The challenge was then to combine the distance function with the Dice loss function, the two functions having different scales and gradients. In the literature, this issue is not generally discussed and only empirical solutions were proposed.

The first contribution of our study is the combination of the Dice loss function and a boundary loss function using a gradient rescaling technique. The second contribution consists in applying our method, with a U-Net architecture, to the segmentation of pelvic organs on dynamic 2D sagittal MRI. We show that our approach improved contour accuracy on dynamic images of controls and patients with pathological disorders and outperformed DL methods using standard loss functions and more classical methods such as active contours.

II. Methods

The section is organized as follows. The first part describes the scores used for the evaluation of the results, this explanation being necessary to understand our loss function. Then follows a description of the function and the solutions found to associate it with the Dice loss.

A. Evaluation metrics

Experimental results need to be assessed with quantitative scores to show the accuracy of a segmentation method. The results were evaluated using 3 common indicators in segmentation that quantify the similarity between the network prediction and the manually established ground truth. The first one is the dice similarity coefficient (DSC) which measures the overlap between a segmentation result and the ground truth. Let X be the network's prediction and Y the ground truth mask, where X and Y are binary images. The DSC is defined by (1) with values varying between 0 (no overlap) and 1 (full overlap).

$$DSC(X,Y) = \frac{2|X \cap Y|}{|X| + |Y|} \tag{1}$$

The two other indicators concern the distance from the mask boundary: the ASD (2) and the Hausdorff distance (HD) (3). ASD measures a global distance error whereas HD focuses on the longest difference between prediction and the mask. Each measure is symmetric with respect to its operand. Let $\partial X, \partial Y$ be respectively the boundaries of X and Y, and $\delta(x,y)$ the euclidean distance between two points x and y. $\delta(x, \partial Y)$ denotes the minimal euclidean distance between x and the boundary ∂Y.

$$ASD(\partial X, \partial Y) = \frac{\sum_{x \in \partial X} \delta(x, \partial Y) + \sum_{y \in \partial Y} \delta(\partial X, y)}{|\partial X| + |\partial Y|} \tag{2}$$

$$HD(\partial X, \partial Y) = \max\{ \sup_{x \in \partial X} \delta(x, \partial Y), \sup_{y \in \partial Y} \delta(\partial X, y)\} \tag{3}$$

Although the DSC is a widely used measure, it is not sufficient to properly assess a result quality since it represents an average which is not sensitive to thin shaped errors. For example, in Fig. 1 in the upper right image, although the DSC is high (0.89) so is the HD (6.6 mm), showing a segmentation failure. Overlap and distance between boundaries are complementary to compare efficiency of segmentation algorithms. A well-designed loss function should take into account these 3 metrics to segment effectively.

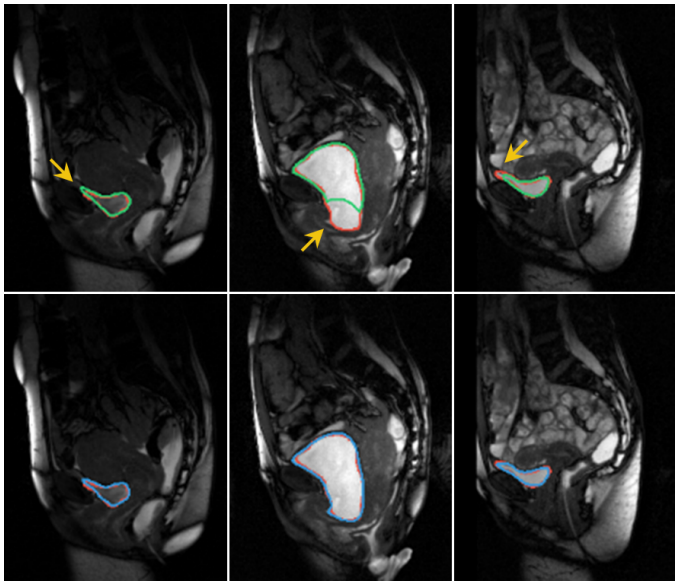

Fig. 1. Examples of 3 bladder image segmentations: ground truth (red), network trained with $\mathcal{L}_O$ (green), network trained with the proposed method $\mathcal{L}_P$ (blue). Yellow arrows show main delineation errors.

B. Loss functions

Our function is designed as a combination of an overlap based function (Dice loss) and a boundary-aware loss based on ASD. The Dice loss, which is a well known estimation of DSC, is first presented. Let $\Omega \subset \mathbb{R}^2$ be the pixel grid where $p \in \Omega$ represents a pixel coordinates, X is the network output composed of real values in $[0,1]$ (*i.e.* the probability softmax output). Let Y be the binary ground truth (1 inside, 0 outside). $\mathcal{L}_O$, the DSC based loss function is as follows :

$$\mathcal{L}_O(X,Y) = 1 - 2\frac{\sum_{p \in \Omega} X(p)Y(p)}{\sum_{p \in \Omega} X(p)^2 + \sum_{p \in \Omega} Y(p)^2} \tag{4}$$

$\mathcal{L}_O$ values go from 1 (no similarity) to 0 (perfect similarity) with real values in order to be minimized by the optimization process.

The second loss was inspired by the ASD score. However, ASD and HD cannot be used as loss functions for a learning algorithm because they are not differentiable and have a high algorithmic cost. Therefore, estimates of each metric must be found for use as a cost function. The boundary loss function is an estimate of the ASD that can be formulated as a function using a boundary distance map associated with ground truth only, as reported by Karimi *et al.* [13]. Indeed, calculation of the prediction distance map would imply too high an algorithmic cost as it would need to be computed for each predicted image. Let d_Y be the distance map. It assigns to each image pixel a value corresponding to the distance of that pixel from the boundary of the ground truth mask. It is the same size as the image. So, the loss function, $\mathcal{L}_D$ (Distance loss), is asymmetric and defined as follows :

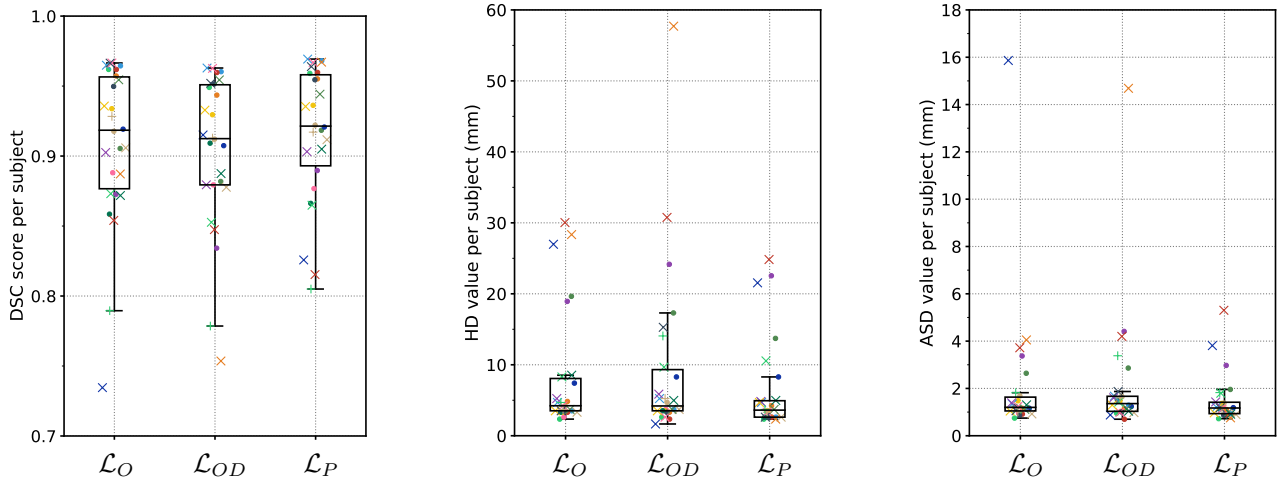

Fig. 2. DSC (Left), HD (Middle) and ASD (Right) scores of segmentation made by network trained with Dice loss ($\mathcal{L}_O$), combined loss with empirical α coefficient ($\mathcal{L}_{OD}$) and our proposed loss function ($\mathcal{L}_P$). Each flier represents the score for a subject.

$$\mathcal{L}_D(X,Y) = \frac{1}{|\Omega|}\sum_{p\in\Omega}(X(p)-Y(p))^2 d_Y(p) \tag{5}$$

The map weights are high if the squared error between ground truth and prediction is far from the ground truth boundary and low when the error approaches it. The loss value depends strongly on the distance from boundary.

C. Loss functions compatibility

The combination of two loss functions raised an issue of compatibility. Theoretically, minimizing one loss function should minimize the second one in the same time. In practice, the loss functions may be minimized at different speeds since they did not have the same order of magnitude. We introduced an α coefficient which plays two roles: (1) ensuring gradient compatibility, (2) smoothing the transition between the training with dice loss only and combined loss. Usually, α coefficients are chosen empirically. In [11], an α starting at 0 with a 0.01 increment by epoch (until reaching 1) is used in most cases and the Dice loss is weighted by $1-\alpha$. This choice is empirical and by definition may not work in all cases. Thus, we designed an adaptive α to solve the compatibility problem between our two losses. It is defined as the ratio between μ_o and μ_d averaged over the last 20 learning iterations, where μ_o and μ_d are the gradients norms associated with $\mathcal{L}_O$ and $\mathcal{L}_D$ respectively.

$$\mathcal{L}_{OD}(X,Y) = \mathcal{L}_O(X,Y) + \alpha\mathcal{L}_D(X,Y) \tag{6}$$

Thus, the distance loss gradient is rescaled to avoid exploding gradient issues, leading to the combined loss $\mathcal{L}_{OD}$ (6). Besides, the distance based loss yielded exploding gradient values when the prediction was too far from the ground truth. Which is specially the case when detection is not achieved yet. Thus, we have chosen to start the training with the Dice loss. Once an average training error of 0.05 (*i.e.* a DSC of 0.95) is reached, the ASD reduction function is integrated to refine the result. This refinement strategy was chosen to shorten the training because when two loss functions are combined, the convergence of the network can be longer. Although the threshold choice is arbitrary, it was found to have little influence on the outcome.

III. Experiments and Results

A. Datasets

Pelvis areas of 26 subjects were imaged with a 1.5T MRI scanner (PHILIPS Gyroscan) using an ultrafast T_2-weighted pulse sequence (TR: 3.6 ms, TE: 1.8 ms, slice thickness: 10 mm, image size: 256 x 256, pixel size : 1.17 x 1.17 mm^2). For each subject, 12 images were acquired during a dynamic acquisition of the sagittal median plane. Among subjects, 16 exhibited pathological deformities of the bladder. To show the interest of our boundary loss function, we have chosen to focus on the bladder which is the organ with visible contours.

B. Implementation

As a preprocessing a N4 bias field correction was applied and images were then normalized to range between 0 and 1. The architecture chosen for the experiment was a 4 layers U-Net [1], which is classically used for medical imaging segmentation tasks. The number of filters per layer for the encoder was chosen as follows: 64/128/256/512, the decoder is symmetric. Within a Pytorch framework, the chosen optimizer was AdamW with a learning rate of 10^{-4}. Batch normalization and spatial dropout layers were added to the network as well as early stopping to prevent from overfitting. The optimization was made with mini-batches of 4 images. Experiments were run on a Linux Xeon Silver Workstation (4214cpu@2.2 GHz - 96Gb) with a Nvidia GeForce RTX 2080 Super GPU, training time was around 2 hours for one network. To evaluate our method, a leave-one-out cross validation was chosen, which means one network per subject of the study has been trained and the whole dataset has been evaluated.

C. Results

To demonstrate the improvements made by our method, the results of three experiments were compared: (1) training with $\mathcal{L}_O$ only, (2) training with $\mathcal{L}_{OD}$ using α starting at 0

and incrementing by 0.01 per epoch (3) training with our proposed method. Figure 2 shows the DSC, HD and ASD by subject. The scores are computed over the entire 12-frame sequence of a subject, allowing us to analyze the network's ability to segment all instants of pelvic dynamics (*i.e.* all possible bladder shapes) of each subject. The HD and ASD are respectively the maximum and mean distance over the 12 images of a sequence. The best scores were achieved with our proposed method with average DSC, HD and ASD of 0.92 ± 0.05, 6.5 ± 6.5 mm, 1.5 ± 1.0 mm. Average DSC, HD and ASD for method (1) and (2) were respectively 0.91 ± 0.06, 8.2 ± 8.4 mm, 2.09 ± 2.88 mm and 0.90 ± 0.05, 11 ± 9.4 mm, 2.15 ± 2.68 mm.

IV. Discussion

The segmentation has been mostly successful with DSC over 0.9 and ASD around $1.5mm$. With the same dataset, our results are better (DSC>0.9 except P51 at 0.90) than those were achieved with an active contour based method in [4]. Comparison of our results with those of the others work cited here is not relevant because of the distinct nature of images.

Two important pieces of information can be extracted from the results. First, the addition of a cost function with an empirical alpha coefficient may not improve the results compared to the classical Dice loss. Indeed, the scores of DSC, HD and ASD are worse and the variability of the scores is larger. Second, the addition of an adaptive alpha coefficient allows to outperform the results obtained with the Dice loss. The HD and ASD values of the network predictions were reduced by 21 and 28%. The dispersion of the scores was also reduced by 23% for the HD and 65% for the ASD. The impact of adding a distance loss function helped greatly the boundaries detection, as illustrated in Fig.1. Our method helped improving the segmentation quality and reduced the inter-subject variability.

Despite the progress achieved, the loss function presented in this paper has two main limitations. First, the function is asymmetric as only ground truth distance map was computed. This asymmetry made the estimation less exact, and we can suppose that a better estimation would have improved the accuracy of the result. Second, we experimented that having two losses raises an issue of gradient compatibility. Considering two losses have two different gradient scales, optimizing the first loss won't necessarily optimize the second one. In this study, the issue has been solved with the help of the α coefficient, which rescales the loss gradients. However, more studies on gradient compatibility deserves to be done especially because hybrid losses are developing in every deep learning tasks.

Further researches have to be done around loss function as a function that takes into account all segmentation errors, overlap-, boundary- and shape-based. Studies have been made on metrics to improve the evaluation of segmentation in the medical field. However, these metrics are not always appropriate for deep learning, since the cost function must be differentiable, and need to have a low computational cost to allow an efficient training.

V. Conclusion

This study contribution is twofold. First, a method for segmenting pelvic organs on 2D dynamic MRI in a robust and automatic way has been developed. Second, a loss function combining overlap and boundary aware errors have been introduced to tackle the variety of organ deformations during straining exercise. A rescaling gradient coefficient was introduced to ensure compatibility between loss functions. The results have consistently shown that our contour-aware loss provides more robust results in comparison with the classical Dice loss only. Accurate automatic segmentation of the main organs paves the way towards an automatic gradation of pelvis pathology. Efficiency of the proposed method regarding other architectures and other organs is a work in progress.

References

[1] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.

[2] I. Nygaard, M. D. Barber, K. L. Burgio *et al.*, "Prevalence of symptomatic pelvic floor disorders in us women," *Jama*, vol. 300, no. 11, pp. 1311–1316, 2008.

[3] M. Rahim, M.-E. Bellemare, R. Bulot *et al.*, "A diffeomorphic mapping based characterization of temporal sequences: application to the pelvic organ dynamics assessment," *Journal of mathematical imaging and vision*, vol. 47, no. 1-2, pp. 151–164, 2013.

[4] R. Namías, J.-P. D'Amato, M. del Fresno, M. Vénere, N. Pirró, and M.-E. Bellemare, "Multi-object segmentation framework using deformable models for medical imaging analysis," *Medical & Biological Engineering & Computing*, vol. 54, no. 8, pp. 1181–1192, 2016.

[5] G. Almeida and J. M. R. Tavares, "Deep learning in radiation oncology treatment planning for prostate cancer: A systematic review," *Journal of Medical Systems*, vol. 44, no. 10, pp. 1–15, 2020.

[6] D. Nie, L. Wang, Y. Gao *et al.*, "STRAINet: Spatially varying sTochastic residual AdversarIal networks for MRI pelvic organ segmentation," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 5, pp. 1552–1564, may 2019.

[7] D. D. Pham, G. Dovletov, S. Warwas, S. Landgraeber, M. Jäger, and J. Pauli, "Deep learning with anatomical priors: imitating enhanced autoencoders in latent space for improved pelvic bone segmentation in mri," in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*. IEEE, 2019, pp. 1166–1169.

[8] K. He, X. Cao, Y. Shi, D. Nie, Y. Gao, and D. Shen, "Pelvic organ segmentation using distinctive curve guided fully convolutional networks," *IEEE transactions on medical imaging*, vol. 38, no. 2, pp. 585–595, 2018.

[9] S. Wang, K. He, D. Nie *et al.*, "Ct male pelvic organ segmentation using fully convolutional networks with boundary sensitive representation," *Medical image analysis*, vol. 54, pp. 168–178, 2019.

[10] J. H. Moltz, A. Hänsch, B. Lassen-Schmidt, B. Haas, A. Genghi, J. Schreier, T. Morgas, and J. Klein, "Learning a loss function for segmentation: A feasibility study," in *IEEE 17th International Symposium on Biomedical Imaging*. IEEE, 2020, pp. 357–360.

[11] J. Ma, Z. Wei, Y. Zhang, Y. Wang, R. Lv, C. Zhu, C. Gaoxiang, J. Liu, C. Peng, L. Wang *et al.*, "How distance transform maps boost segmentation cnns: an empirical study," in *Medical Imaging with Deep Learning*. PMLR, 2020, pp. 479–492.

[12] H. Kervadec, J. Bouchtiba, C. Desrosiers *et al.*, "Boundary loss for highly unbalanced segmentation," in *International conference on medical imaging with deep learning*, 2019, pp. 285–296.

[13] D. Karimi and S. E. Salcudean, "Reducing the hausdorff distance in medical image segmentation with convolutional neural networks," *IEEE Transactions on medical imaging*, vol. 39, no. 2, pp. 499–513, 2019.