Anatomically Constrained Semi-supervised Learning for Echocardiography Segmentation

Thierry Judge THIERRY.JUDGE@USHERBROOKE.CA
Arnaud Judge ARNAUD.JUDGE@USHERBROOKE.CA
Pierre-Marc Jodoin PIERRE-MARC.JODOIN@USHERBROOKE.CA
Department of Computer Science, University of Sherbrooke, Sherbrooke, QC, Canada

Editors: Under Review for MIDL 2022

Abstract

Deep convolutional neural networks (CNNs) have had great success for medical imaging segmentation. Many methods attained nearly perfect Dice scores, sometimes within inter-expert variability. However, CNNs require large amounts of labeled data and are not immune to producing anatomically implausible results, especially when applied to ultrasound images. In this paper, we propose a method that tackles both of these problems simultaneously. Our method optimizes anatomical segmentation metrics on both labeled and unlabeled data using a training scheme analogous to adversarial training. Our method allows the optimization of several hand-made non-differentiable metrics for any segmentation model and drastically reduces the number of anatomical errors. The code is available at https://github.com/ThierryJudge/anatomically-constrained-ssl.

Keywords: Ultrasound segmentation, Semi-supervised learning, Metric optimization

1. Introduction

Convolutional neural networks (CNNs) are the go-to solution for echocardiography segmentation capable of producing results within inter-expert variability (Leclerc et al., 2019). They however required large amounts of labeled data and are prone to making predictions that are not consistent with the underlying anatomy. These anatomical errors often hinder the credibility that neural networks may have in clinical practice. Recently, (Painchaud et al., 2020) have shown to what extent these errors occur in state-of-the-art segmentation methods. While their post-processing method can scrub out anatomical errors, it nonetheless requires a large number of labeled images and constitutes a non-negligible processing overhead which raises questions regarding its real-life usability.

Techniques for enforcing anatomical constraints are essential, especially when little labeled data is available. In this paper, we propose a novel training method to directly optimize non-differentiable anatomical metrics and leverage them to enforce anatomical prior for both labeled and unlabeled data. We do so by using a classification neural net whose back-propagated gradients emulate those that would have been produced by an anatomical loss. We also show that by its very nature, our method adapts to partly annotated data and show that adding unannotated data further improves results. The system is also trained on a scheme that prevents from falling into class imbalance problems.

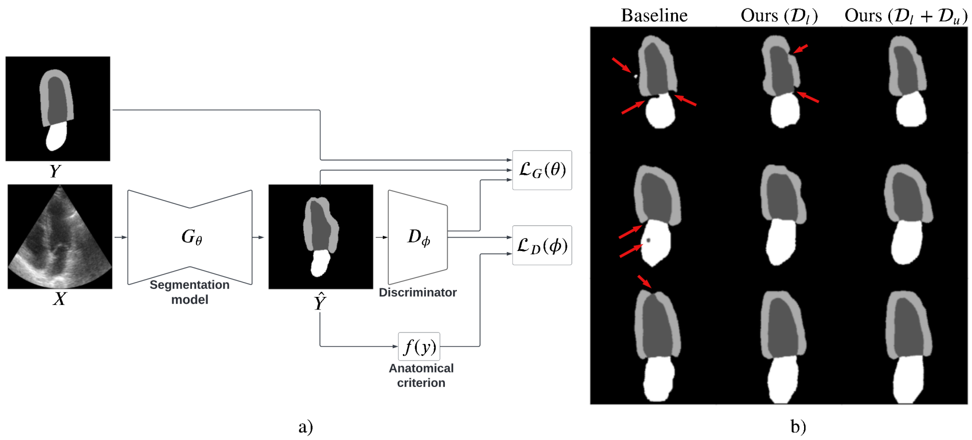

Figure 1: a) Illustration of the proposed method. b) Examples of predictions for different methods. Anatomical errors are indicated with red arrows.

2. Method

Lets consider $\mathcal{D}_l = \{(x_l^{(1)}, y_l^{(1)}), ..., (x_l^{(n)}, y_l^{(n)})\}$, where $x \in \Re^{C \times H \times W}$ is an input image, $y \in \{0,1\}^{K \times H \times W}$ a ground truth segmentation map, and $G_\theta(x) : \Re^{C \times H \times W} \rightarrow \{0,1\}^{K \times H \times W}$ a segmentation network trained with a loss function $\mathcal{L}_{sup}(x, y)$. While segmentation networks generalize well, they offer no anatomical guarantee what so ever.

At the core of our method is an anatomical constraint function based on 12 anatomical criteria (region connections, sizes, concavities, etc.) outlined in (Painchaud et al., 2020). This function $f(y) : \{0,1\}^{K \times H \times W} \rightarrow \{0,1\}$ returns 0 when the segmentation map y is anatomically invalid (i.e. whenever it violates at least one anatomical criterion) and 1 when every criterion is satisfied.

Since the constraint function is non-differentiable, one cannot use it as a loss function. As a workaround, a follow up network $D_\phi(y) : [0,1]^{K \times H \times W} \rightarrow [0,1]$ is used to emulate it. As shown in Figure 1, $D_\phi(y)$ predicts if the input segmentation map y is anatomically valid or not. Since the system is trained end-to-end, the back-propagated gradients force the segmentation network to learn anatomical concepts better adapted to the task at hand than those a network gets to learn when trained solely with a Dice or a cross-entropy loss.

Since $f(y)$ does not rely on groundtruth data, one can use unlabeled sets of data $\mathcal{D}_u = \{x_u^{(1)}, x_u^{(2)}, ..., x_u^{(m)}\}$ to train both networks. This results in the following loss equations:

$$\mathcal{L}_G(\theta) = \sum_{\mathcal{D}_l} \mathcal{L}_{sup}(x_l^{(i)}, y_l^{(i)}) - \lambda_1 \log\left(D_\phi(G_\theta(x_l^{(i)})\right) - \lambda_2 \sum_{\mathcal{D}_u} \log\left(D_\phi(G_\theta(x_u^{(i)})\right) \tag{1}$$

$$\mathcal{L}_D(\phi) = -\sum_{\mathcal{D}_l + \mathcal{D}_u} f(G_\theta(x^{(i)})) \log\left(D_\phi(G_\theta(x^{(i)}))\right) + (1 - f(G_\theta(x^{(i)}))) \log\left(1 - D_\phi(G_\theta(x^{(i)}))\right) \tag{2}$$

with $\lambda_1 = 0.025$ and $\lambda_2 = 0.01$. As the discriminator is trained on samples generated by $G_\theta(x)$, there is a high chance of class imbalance with respect to the valid and in-valid segmentations. To alleviate this problem, we use a replay buffer in which is saved an equal number of valid and in-valid samples at every step. The discriminator is trained with batches made of samples randomly selected from the buffer.

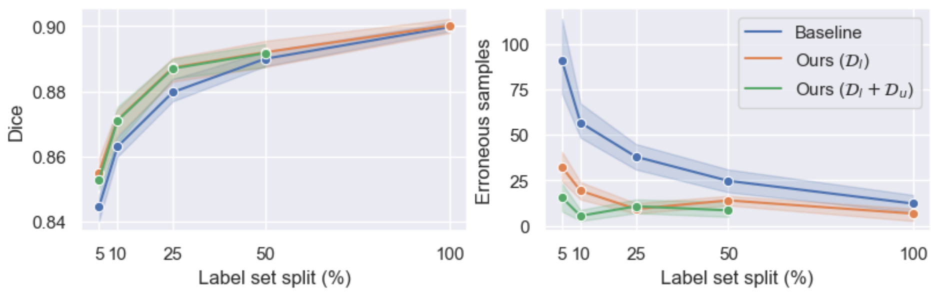

Figure 2: Dice and number of anatomically erroneous samples in the test set (200 samples). Methods trained on datasets containing 5% to 100% of the full training set labels.

3. Results

We trained and tested our method on the CAMUS (Leclerc et al., 2019) dataset which contains cardiac ultrasound images of 500 patients. The left ventricle, myocardium and left atrium are manually labeled for end-diastolic and end-systolic instants for both the 2D four-chamber and two-chamber apical views.

A standard Enet (Paszke et al., 2016) trained with a batch size of 32 was used for all methods. We trained a baseline Enet with a combination of Dice and cross-entropy loss using a constant learning rate of 0.001 with weight decay of 1e-4. We tested our method with labeled data ($\mathcal{D}_l$) and with labeled+unlabeled data ($\mathcal{D}_l + \mathcal{D}_u$). G_θ was initialized with the baseline weights and D_ϕ was pre-trained for 1000 steps to help it better converge. Results in Figure 2 show an increase in Dice score when a small amount of labeled data is used to train the system and, most importantly, an important decrease in the number of anatomical errors. The decrease is even more drastic when unlabeled images are included.

4. Conclusion

In this paper, we proposed a novel method for optimizing a non-differentiable anatomical prior. As the computation of the anatomical prior does not require the ground truth, our method is suitable for semi-supervised learning.

We show that our method drastically reduces the number of anatomical errors in the test set predictions without requiring any post-processing. We also show that our method improves the dice score as a result of the regularization induced by the anatomical constraints.

Acknowledgments

This work was funded by the NSERC Graduate Scholarships – Master's program, the NSERC Discovery Grant and by MITACS Accelerate.

References

Sarah Leclerc et al. Deep learning for segmentation using an open large-scale dataset in 2d echocardiography. *IEEE Transactions on Medical Imaging*, 38(9):2198–2210, 2019.

Nathan Painchaud et al. Cardiac segmentation with strong anatomical guarantees. *IEEE Transactions on Medical Imaging*, 39(11):3703–3713, 2020.

A. Paszke, A. Chaurasia, S. Kim, and E. Culurciello. ENet: A Deep Neural Network Architecture for Real-Time Semantic Segmentation. In *arXiv:1606.02147*, 2016.