Influence of Loss Function on Left Ventricular Volume and Ejection Fraction Estimation in Deep Neural Networks

Preshen Naidoo[1] 21468831@student.uwl.ac.uk
Eman Alajrami[1] 21452392@student.uwl.ac.uk
Elisabeth Lane[1] elisabeth.lane@uwl.ac.uk
Jevgeni Jevsikov[1] jevgeni.jevsikov@uwl.ac.uk
Matthew J Shun-shin[2] m@shun-shin.com
Darrel P Francis[2] darrel@drfrancis.org
Massoud Zolgharni[1,2] massoud@zolgharni.com
[1] *School of Computing and Engineering, University of West London, London, United Kingdom*
[2] *National Heart and Lung Institute, Imperial College, London, United Kingdom*

Editors: Under Review for MIDL 2022

Abstract

Quantification of the left ventricle shape is crucial in evaluating cardiac function from 2D echocardiographic images. This study investigates the applicability of established loss functions when optimising the U-Net model for 2D echocardiographic left ventricular segmentation. Our results indicate loss functions are a significant component for optimal left ventricle volume measurements when established segmentation metrics could be imperceptible.

Keywords: Echocardiography, Left Ventricle Segmentation, Deep learning

1. Introduction

Accurate left ventricular (LV) volume and ejection fraction (EF) assessment are essential in cardiac diagnostics. Echocardiography is the most frequently used modality in clinical routine to assess EF owing to its widespread availability and high temporal resolution. However, the Simpson's method, widely used and recommended by the American Society of Echocardiography, relies on specialists to delineate the endocardium border of the LV. It is laborious, subjective and prone to poor repeatability. Thus, automated algorithms are desired for accurate, objective, and efficient EF measurements.

Segmentation of the endocardial border is fundamental to automate LV measurements. Numerous echocardiographic segmentation techniques have appeared in the literature. Recently, deep convolutional neural networks show promise due to their speed and accuracy. U-Net is known for producing a higher accuracy in medical image segmentation applications (Ronneberger et al., 2015); many studies use U-net or design their network on it.

The choice of loss function is important as they institute the process of learning an objective function. Therefore, domain specific loss functions to improve results for each dataset is of crucial importance. Previous research has focused on the LV measurements simply as a segmentation task (with its performance metrics such as Dice Coefficient), as opposed to the suitability of various loss functions for the task of LV volume estimation. In this study, therefore, we have investigated the fitness of common segmentation-based loss functions when applied to echocardiographic EF measurements.

© 2021 P. Naidoo, E. Alajrami, E. Lane, J. Jevsikov, M.J. Shun-shin, D.P. Francis & M. Zolgharni.

2. Method and dataset

1224 Videos of the apical 4-chamber echocardiographic view, acquired between 2015 and 2016, were extracted from Imperial College Healthcare NHS Trust's echocardiogram database. The acquisition of the images was performed by experienced echocardiographers and according to standard protocols, using ultrasound equipment from GE and Philips manufacturers. Ethical approval was obtained from the Health Regulatory Agency (Integrated Research Application System identifier 243023). From these videos, a total of 2600 images were sampled from different points in the cardiac cycle. Each image underwent labelling by one individual from a pool of experts using our in-house online labelling platform (https://unityimaging.net). This dataset was used for model developments (i.e., training and validation).

The testing comprised 100 videos, from consecutive studies conducted over 3 working days in 2019, at least 3 years away from the date of collection for the model development dataset. Our previous deep neural network (Lane et al., 2021) was used to identify the end-diastolic and end-systolic frames of the 100 videos. These selected frames were used for the human expert annotations. Each of the 200 resulting images was then labelled by 11 experts, using the platform. The concensus of the expert was finally used as the ground-truth in the testing dataset.

U-Net was implemented in TensorFlow and was trained on an Nvidia RTX3090 GPU. The loss functions chosen for experimentation are common for image segmentation tasks and were selected from three categories:

- Distribution-based loss: *Binary cross entropy (BCE) loss*
- Region-based loss: *Dice loss, Tvsersky loss, Focal Tvsersky loss*
- Compound loss: *BCE+Dice*

After training the model for different loss functions, each model was evaluated using the established evaluation metrics for segmentation tasks, i.e., Dice Coefficient (DSC) and Hausdorff Distance (HD) and domain specific metrics (i.e., volume and EF measurements) by averaging the error across all images in the testing dataset. The volume was computed using Simpson's method. The average Volume Error is calculated using Cartesian pixel coordinates by computing the difference between the volume of the ground-truth and the predicted endocardial borders and can be expressed as:

$$\boldsymbol{V}avg\epsilon = \frac{1}{m}\sum_{i=1}^{m} |\,\boldsymbol{V}gt_i - \boldsymbol{V}pred_i|,$$

EF was estimated by dividing the stroke volume (i.e., the difference between end-diastolic and end-systolic volumes) by the end-diastolic volume:

$$EF(\%) = \left(\frac{EDV - ESV}{EDV}\right) \times 100$$

3. Results and Discussion

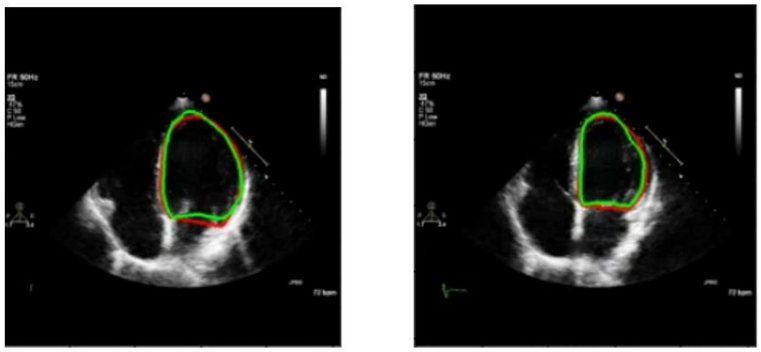
Table 1 highlights the miniscule variation in DSC and HD across all loss functions despite vast differences in Average Volume Error. Hence, excellent scores for DSC and HD do not

Table 1: Error in different metrics computed as average across all images in testing dataset.

Loss Function	Avg EF Error	$\boldsymbol{V}\, avg\epsilon(ml)$	Avg DSC	Avg HD
BinaryCrossEntropy	3.6313	4.7550	0.9278	4.4245
Dice loss	3.6932	4.3181	0.9318	4.3139
BCE and Dice loss	3.4674	6.6417	0.9279	4.2586
Tversky loss	3.7554	3.4348	0.9316	4.3438
Focal Tversky loss	3.4797	3.1088	0.9336	4.2883

imply the most optimal volume computation for EF. For instance, when using 'BCE and Dice loss', a large volume error is observed whereas DSC is insensitive to the shape of the predicted LV border. Interestingly, some large errors in the volume measurements may be cancelled out in the EF measurements. One example of such scenario is shown in Figure 1. Our results demonstrate evaluating model performance based on established metrics alone is insufficient when estimating LV volume. Our future research centres around further investigation of the generalisability of a loss function across many echocardiographic image datasets for training a segmentation model with the objective of computing an improved approximation for EF in clinical practice.

Figure 1: Depicting end-diastolic (left) and end-systolic (right) echo images with ground-truth (red) and predicted endocardium borders (green), with a volume error of 18.24ml and 14.82ml respectively, but with an EF error of 2.29%.



References

Elisabeth S. Lane, Neda Azarmehr, Jevgeni Jevsikov, James P. Howard, Matthew J. Shunshin, Graham D. Cole, Darrel P. Francis, and Massoud Zolgharni. Multibeat echocardiographic phase detection using deep neural networks. *Computers in Biology and Medicine*, 133, 2021. doi: 10.1016/j.compbiomed.2021.104373.

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *MICCAI 2015*, pages 234–241, Cham, 2015. Springer International Publishing. ISBN 978-3-319-24574-4.