Reaching intra-observer variability in 2-D echocardiographic image segmentation with a simple U-Net architecture

Hang Jung Ling, Damien Garcia, and Olivier Bernard
CREATIS, CNRS UMR5220, Inserm U1294, University of Lyon, France

Background, Motivation and Objective

Segmentation of cardiac structures in echocardiography, including the endocardial and epicardial contours of the left ventricle (LV), has been the subject of much research for years. To date, only one algorithm, CLAS, managed to achieve performance comparable to intra-observer variability by using a nontrivial loss function. Recently, nnUNet, a powerful segmentation framework, has been introduced. It is capable of adapting to multiple medical datasets without manual tuning. Hence, we carried out an extensive study of nnUNet to investigate whether a simple U-Net architecture could have equivalent segmentation performance to intra-observer variability.

Statement of Contribution/Methods

We proposed four models that were trained on the largest echocardiographic open dataset (CAMUS) by successively adding the core elements of nnUNet (53M parameters) into a simple U-Net (Model #1): deep supervision (Model #2), data augmentation in training (Model #3), data augmentation in inference (DAiI), and patch-wise approach (Model #4). To boost the generalization ability of the models, we used a specific optimization scheme: a small batch size of 2, and a reduced number of iterations of 250 per epoch. The geometrical metrics used were the Hausdorff (HD) and average symmetric surface (ASSD) distances. Correlations (Corr) and mean absolute errors (MAE) of the estimated end-diastole and end-systole volumes, and ejection fraction were computed for clinical evaluation. Model #3 and #4 were evaluated twice, with and without DAiI.

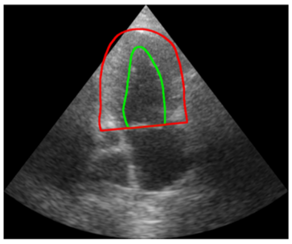
Results/Discussion

Regarding segmentation performance (2^{nd} column of the Table), starting with Model #3 + DAiI, the models outperformed CLAS and intra-observer variability significantly ($p < 0.05$). Consistent with the clinical scores, the estimated volumes were highly correlated with the ground truths ($r > 0.97$), although they were less accurate than CLAS for estimating the ejection fraction due to lack of temporal consistency. Our study shows that the data augmentation in both training and inference, combined with a well-matched optimization scheme, is the key to reaching intra-observer variability. As a result, a simple U-Net architecture (7M parameters) can produce high-quality segmentation and accurate volume estimation.

Methods /no. parameters	$LV_{Endocardial}$ & $LV_{Epicardial}$			*End Diastolic Volume*		*End Systolic Volume*		*Ejection Fraction*	
	HD±σ (mm)	**ASSD±σ (mm)**	**Outliers (%)**	**Corr**	**MAE±σ (mL)**	**Corr**	**MAE±σ (mL)**	**Corr**	**MAE±σ (%)**
Intra-observer	4.7 ±2.0	1.5 ±0.7	-	0.978	6.5 ±4.4	0.981	4.5 ±3.9	0.895	4.7 ±4.1
CLAS	4.8	1.5	-	0.958	-	0.979	-	0.926	-
Model #3 + DAiI /7M	4.5 ±2.1 (**)	1.4 ±0.7 (****)	1.8	0.974	6.8 ±6.1	0.974	5.6 ±4.8	0.863	4.6 ±4.0
Model #4 /30M	4.5 ±1.9 (*)	1.4 ±0.7 (**)	1.9	0.972	6.7 ±5.9	0.972	5.5 ±5.1	0.840	4.9 ±4.3
Model #4 + DAiI /30M	4.4 ±1.9 (***)	1.4 ±0.7 (****)	1.5	0.972	6.6 ±5.7	0.972	5.5 ±4.8	0.843	4.7 ±4.4
nnUNet /53M	4.3 ±1.9 (****)	1.3 ±0.6 (****)	1.5	0.976	6.5 ±5.6	0.976	5.3 ±4.6	0.876	4.4 ±3.6

no. parameters: Number of parameters; ·M: · Million parameters
Statistical test: Left tailed two sample t-test conducted between each model and intra-observer variability
(*): p-value < 0.05; (**): p-value < 0.01; (***): p-value < 0.001; (****): p-value < 0.0001

Ground truth



Model #3 + DAiI