

## Improving the generalization capability of deep learning-based algorithms for ventricular cavity segmentation using simulated CMR images

Y. Al Khalil<sup>1</sup>, S. Amirrajab<sup>1</sup>, C. Lorenz<sup>2</sup>, J. Weese<sup>2</sup>, M. Breeuwer<sup>1,3</sup>

<sup>1</sup>Biomedical Engineering Department, Eindhoven University of Technology, Eindhoven, The Netherlands; <sup>2</sup>Philips Research Laboratories, Hamburg, Germany; <sup>3</sup>MR R&D – Clinical Science, Philips Healthcare, Best, The Netherlands

### Synopsis

The generalization capability of deep learning-based segmentation algorithms across different sites and vendors, as well as MRI data with high variance in contrast, is limited. This affects the usability of such automated segmentation algorithms in clinical settings. The lack of freely accessible medical datasets additionally limits the development of stable models. In this work, we explore the benefits of adding a simulated dataset, containing realistic contrast variance, into the training procedure of the neural network for one of the most clinically important segmentation tasks, the CMR ventricular cavity segmentation.

### Introduction

Cardiovascular magnetic resonance (CMR) imaging is often utilized for clinical evaluation of cardiac function, where an accurate segmentation of multiple cardiac tissues in both end diastolic (ED) and end systolic (ES) phases is one of the primary tasks<sup>1</sup>. Recent advances in medical image processing largely focus on achieving a fully automated segmentation, with deep learning-based methods showing promising results<sup>2, 3</sup>. However, such results are achievable only for a small range of applications, mainly defined by the type of training data used. MRI acquisition can have a substantial impact on tissue segmentation performance, as the abundance of various acquisition protocols and parameters results in a wide range of image appearances and influences image quality. Even state-of-the-art models show degraded accuracy when tested on data obtained from MR imaging sequences or scanners not matching that of the training data<sup>4, 5</sup>. A straightforward solution to this problem consists of training the algorithms with enough data to cover the overall range of variability. This poses a particular challenge in the medical domain, where data is limited and mostly confidential, while obtaining accurate ground truth annotations is costly<sup>6</sup>. To tackle this problem, we propose to incorporate simulated CMR images, obtained from a virtual population of realistic anatomical masks, into a training procedure of a 2D U-net with the aim of segmenting the heart ventricular cavity in short-axis cine CMR images.

### Methods

Short-axis cine MR images were obtained from six different sites, with highly heterogeneous contrasts (as seen in figure 1) due to differences in scanner vendors and models, as well as variable scanner parameters, as summarized in table 1. The segmentation ground truth is provided for both ED and ES phases, containing expert manual segmentations of right ventricular blood pool (RV), left ventricular blood pool (LV) and left ventricular myocardium (LVM). To account for this variance, we propose adding simulated CMR images to the training set, obtained using a human anatomical model for XCAT phantom, where several parameters were varied to provide a realistic variance in the heart's LV function, orientation and position inside the torso. By modifying the sequence parameters to the real ones and scaling the simulated signal intensity, we match the simulated image contrast to their real counterpart. Samples of simulated images are available in figure 2.

The experiments performed in this study are designed to evaluate whether the addition of simulated data into the training set improves the overall segmentation performance (table 2), as well as how it affects the generalization capability of neural networks to data coming from unseen sites (table 3). We adopt a 2D U-Net architecture to perform a multi-structure cardiac segmentation task, following the recommendations of the nnU-Net framework<sup>7, 8</sup>. We choose the 2D U-Net over a 3D U-Net primarily due to the anisotropic nature of the multi-site data. Initially, all data is pre-processed by applying normalization to zero mean and unit standard deviation, as well as by resampling to a fixed voxel spacing of 1.5 mm x 1.5 mm x 1.5 mm. Data augmentation is applied on the fly to increase the variety of the training set and avoid over-fitting, using elastic deformations, random scaling and random rotations.

Submitted to ISMRM Benelux, 24 January 2020, Arnhem, The Netherlands.

During training, a random batch of 32 2D short-axis slices was fed per each iteration into the network. To reduce over-fitting, we use a subset from the training data as a validation set. The network structure is similar to the one proposed in the original paper<sup>9</sup>, with the addition of the batch normalization and leaky ReLU activation functions, as well as dropout regularization (dropout rate of 0.4). We use the sum of cross-entropy and dice loss as a loss function, optimized using the Adam optimizer for stochastic gradient descent with an initial learning rate of 0.001 and a weight decay of  $5 \times 10^{-5}$ .

## Results

Table 2 compares the performance of the networks trained with and without the inclusion of simulated data, where we systematically increase the number of simulated images in the training set. In all cases both the Dice score and intersection over union (IoU) score increase as more simulated data is included. Table 3 shows the segmentation performance of the network on images from different sites. Since the majority of simulated data is designed to match the contrast variance in site F, the improvements in segmentation are most significant for that particular site.

## Discussion and Conclusion

The results obtained in this study indicate a promising solution to address the lack of data availability and generalization capability of neural networks in medical imaging segmentation tasks, which affect the use of DL-based methods in clinical settings. This is achieved even without having highly realistic simulations, which we hypothesize is mainly due to the availability of highly accurate “ground-truth” and inclusion of high contrast variance. Future work involves understanding how much realism is needed for networks to achieve a clinically acceptable performance across larger varying datasets and tasks, such as in large multi-center studies. Moreover, the proposed method can serve as a better and more realistic data augmentation strategy compared to existing methods, as well as improve transfer learning and adaptation methods.

## Acknowledgments / Funding Information

This research is a part of the OpenGTN project, supported by the European Union in the Marie Curie Innovative Training Networks (ITN) fellowship program under project No. 764465.

## References:

1. Weinsaft JW, Klem I, Judd RM. MRI for the assessment of myocardial viability. *Magnetic resonance imaging clinics of North America*. 2007 Nov 1;15(4):505-25.
2. Işın A, Direkoğlu C, Şah M. Review of MRI-based brain tumor image segmentation using deep learning methods. *Procedia Computer Science*. 2016 Jan 1;102:317-24.
3. Litjens G, Ciampi F, Wolterink JM, de Vos BD, Leiner T, Teuwen J, Išgum I. State-of-the-Art Deep Learning in Cardiovascular Image Analysis. *JACC: Cardiovascular Imaging*. 2019 Aug 1;12(8):1549-65.
4. Bai W, Sinclair M, Tarroni G, Oktay O, Rajchl M, Vaillant G, Lee AM, Aung N, Lukaschuk E, Sanghvi MM, Zemrak F. Automated cardiovascular magnetic resonance image analysis with fully convolutional networks. *Journal of Cardiovascular Magnetic Resonance*. 2018 Dec;20(1):65.
5. Petitjean C, Dacher JN. A review of segmentation methods in short axis cardiac MR images. *Medical image analysis*. 2011 Apr 1;15(2):169-84.
6. Chen C, Bai W, Davies RH, Bhuvan AN, Manisty C, Moon JC, Aung N, Lee AM, Sanghvi MM, Fung K, Paiva JM. Improving the generalizability of convolutional neural network-based segmentation on CMR images. *arXiv preprint arXiv:1907.01268*. 2019 Jul 2.
7. Isensee F, Petersen J, Klein A, Zimmerer D, Jaeger PF, Kohl S, Wasserthal J, Koehler G, Norajitra T, Wirkert S, Maier-Hein KH. nnu-net: Self-adapting framework for u-net-based medical image segmentation. *arXiv preprint arXiv:1809.10486*. 2018 Sep 27.
8. Isensee F, Petersen J, Kohl SA, Jäger PF, Maier-Hein KH. nnU-Net: Breaking the Spell on Successful Medical Image Segmentation. *arXiv preprint arXiv:1904.08128*. 2019 Apr 17.

9. Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical image computing and computer-assisted intervention 2015 Oct 5 (pp. 234-241). Springer, Cham.

Dataset	Number of subjects	MRI Scanner Attributes		Image Spatial Resolution	
		Manufacturer	Magnetic Field Strength	In-plane Resolution	Slice Thickness
Site A	10	Philips Ingenia	1.5 T	1.8 $mm^2$ / pixel	8 mm
Site B	22	Philips Achieva	1.5 T	1.8 $mm^2$ / pixel	8 mm
Site C	5	Philips Intera	1.5 T	2.1 $mm^2$ /pixel	8 mm
Site D	110	Philips Ingenuity	3.0 T	1.56 $mm^2$ /pixel	8 mm
Site E	12	Philips Ingenia	1.5 T	1.56 $mm^2$ /pixel	8 mm
Site F	24	Siemens Aera	1.5 T	2.1 - 2.78 $mm^2$ /pixel	8 mm

Table 1. General description of available datasets.

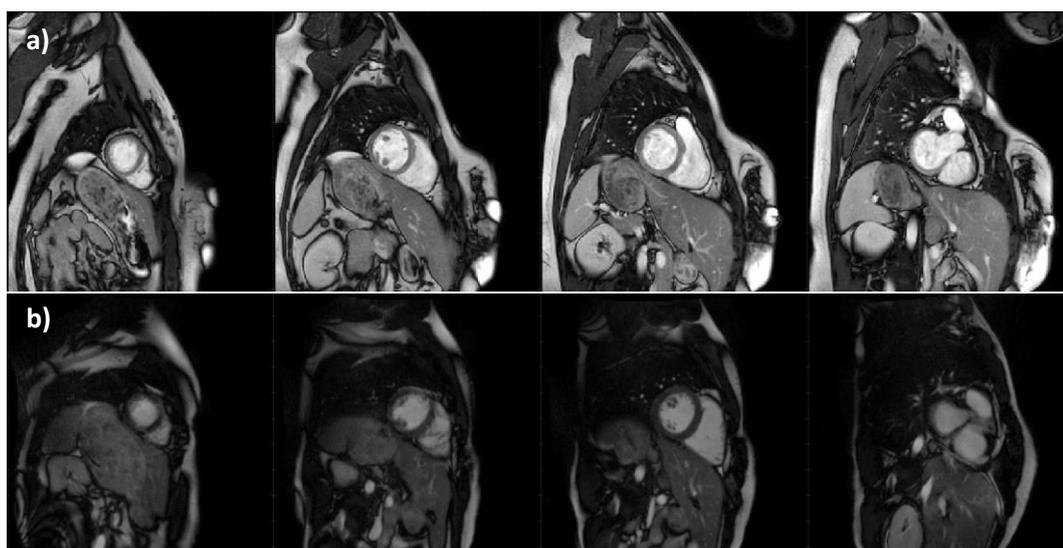


Figure 1. Representative depiction of contrast differences due to variation in scan parameters and vendors between the datasets from sites A to E (a) and site F (b) at end diastolic phase. All data used in this study was obtained with the required approvals and patient consent.

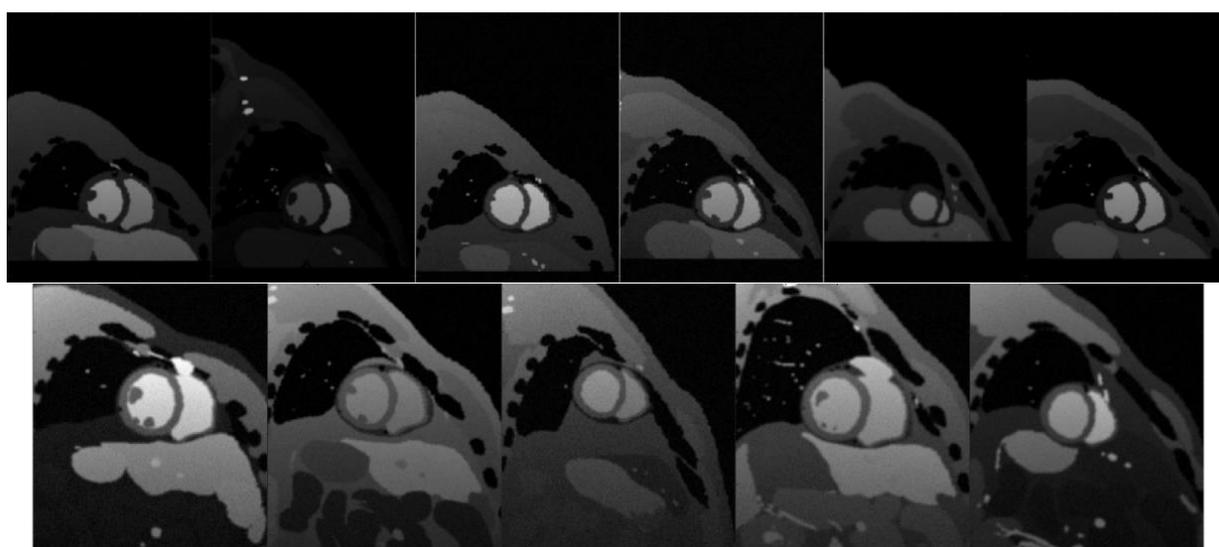


Figure 2. Sample slices from simulated images used in this study for introducing a wider variance in contrast, while providing an accurate ground truth.

#	Train set (# of subjects)		Test set (# of subjects) Site A-F	LV		Myo		RV	
	Site A-F	Simulated Data		Dice	IoU	Dice	Iou	Dice	IoU
1	145	0	38	0.879	0.792	0.754	0.641	0.817	0.708
2	145	12	38	0.881	0.802	0.779	0.659	0.819	0.721
3	145	24	38	0.889	0.812	0.784	0.673	0.825	0.734
4	145	36	38	0.894	0.817	0.789	0.681	0.836	0.739
5	145	48	38	0.901	0.823	0.797	0.689	0.849	0.748
6	145	60	38	<b>0.913</b>	<b>0.834</b>	<b>0.814</b>	<b>0.705</b>	<b>0.856</b>	<b>0.759</b>

Table 2. Segmentation performance with and without the addition of simulated data in the training set. The network was trained for 200 epochs in all cases. All scores listed are the mean of Dice and IoU scores per each test case.

Dataset		Test set	LV		MYO		RV	
Train set			Dice	IoU	Dice	Iou	Dice	IoU
Sites A, B, C, D, E	Site F	0.702	0.549	0.614	0.448	0.595	0.433	
Sites A, B, C, D, E + Simulated Data	Site F	<b>0.734</b>	<b>0.571</b>	<b>0.651</b>	<b>0.493</b>	<b>0.635</b>	<b>0.479</b>	
Sites A, B, C, D, F	Site E	0.887	0.801	0.779	0.641	0.837	0.726	
Sites A, B, C, D, F + Simulated Data	Site E	0.898	0.812	0.791	0.662	0.842	0.731	

Table 3. Segmentation performance with and without the addition of simulated data in the training set on unseen data. The test sets in both cases include the data which we used for simulated data generation and matching. The network was trained for 200 epochs with 30 simulated images added in all cases. All scores listed are the mean of Dice and IoU scores per each test case.