Deep learning-based left heart chambers segmentation and strain analysis from dynamic MRI images

Jonas Leite, Moussa Gueda, Emilie Bollache, Khaoula Bouazizi, Perrine Marsac, Thomas Wallet, Mikaël Prigent, Vincent Nguyen, Jérôme Lamy, Antonio Gallo, et al.

DEEP LEARNING-BASED LEFT HEART CHAMBERS SEGMENTATION AND STRAIN ANALYSIS FROM DYNAMIC MRI IMAGES

Jonas Leite[1], Moussa Gueda[1], Emilie Bollache[1], Khaoula Bouazizi[1,2], Perrine Marsac[1], Thomas Wallet[1], Mikaël Prigent[2], Vincent Nguyen[1], Jérôme Lamy[1,3], Antonio Gallo[1], Elie Mousseaux[3], Alban Redheuil[1,2], Gilles Montalescot[4], Thomas Dietenbeck[1], Nadjia Kachenoura[1]

1. Laboratoire d'Imagerie Biomédicale (LIB), Sorbonne Université, INSERM, CNRS, Paris, France
2. Institute of Cardiometabolism and Nutrition (ICAN), Paris, France
3. European Hospital Georges Pompidou, Paris, France
4. Sorbonne University, ACTION group, Pitié-Salpêtrière Hospital (AP-HP), Paris, France

ABSTRACT

Feature tracking (FT) is increasingly used on dynamic magnetic resonance images for myocardial strain evaluation, but often requires manual initialization of heart chambers, which is tedious and source of variability, especially on the challenging long axis images. Accordingly, we combined a deep learning (DL) approach with FT (DL-FT) to provide fully automated time-resolved left ventricular (LV) and atrial (LA) delineation and strain analysis. This approach was tested on a multi-center and multi-vendor database of 684 healthy controls and patients. DL-initialization achieved Dice scores of 0.89±0.11 for LV endocardium, 0.93±0.07 for LV epicardium and 0.89±0.10 for LA on the testing set of 108 datasets (2-and 4-chambers). LA and LV DL-FT strain peaks were highly associated with expert strains as revealed by correlation coefficients=0.96 for LV and ≥0.70 for LA and mean Bland-Altman biases=0.62% for LV and <1% for LA. Results also revealed stability of our approach over vendors and field strengths.

Index Terms: Deep learning, longitudinal strain, feature tracking, MRI

1. INTRODUCTION

Dynamic cardiac magnetic resonance imaging (cine-MRI) is nowadays the clinical reference for left ventricular (LV) volumes and mass. Such parameters are usually measured from cine-MRI short-axis (SAX) images. Although SAX images provide such valuable clinical measures, their analysis in routine is time consuming since the operator needs to delineate both the endocardium and epicardium on 8 to 12 slices on systolic and diastolic phases. Long-axis (LAX) images, also acquired during standard MRI exam, are often used in routine for the evaluation of left atrial (LA) volumes while manually tracing the LA endocardial borders.

In addition to volumes, recent studies highlighted the clinical value of LV and LA longitudinal strain indices [1], [2] which can be derived from SAX and LAX cine-MRI images using feature tracking (FT) algorithms. However, such additional valuable indices are not included yet in the daily routine because of the lack of extensive validations in populations due to the need for manual initialization of LV and LA contours, which is time consuming and source of variability. To alleviate radiologists from such tedious tasks, several research works including international challenges have been achieved to propose automated analysis of SAX images while using both conventional image processing [3] and deep learning-based [4] algorithms. These achievements have led to a substantial progress in terms of automated processing of SAX images, even in daily routine. However, LAX images could not directly benefit from such progress because of more challenging LV and LA geometry on such views compared to quasi-circular LV shape on SAX views, and the presence of complex neighboring structures such as valves and pulmonary veins.

Accordingly, we designed a fully automated approach to process LAX images, while combining: 1) a dual network based on modified U-Nets [5] for LV and LA contours detection, and 2) FT algorithm initialized by such contours, for longitudinal myocardial strain estimation. Of note, such FT algorithm has been previously validated on both human [6] and animal models [7] and was used here to account for time-continuity of LV and LA myocardial deformation.

2. MATERIALS AND METHODS

2.1. Dataset and ground truth

We studied 684 subjects (61.54% males, 52±12 years), including 94 healthy controls and 590 patients with different cardiac diseases who were included through different protocols approved by local ethics committee (NCT02517944; NCT03715998; NCT02938910). All subjects underwent standard cardiac imaging on MRI

magnets from three different vendors and field strengths (1.5 or 3T). Both SAX and LAX (2-and 4-chamber views) images were acquired with an average pixel size of 1.29±0.34 mm², acquisition matrix of 180-512x200-512 and 25-60 time-phases. Data were analyzed by experienced operators using commercial software (QMass, Medis, Leiden, The Netherlands, version 4.0.24.4) for volumes estimation as well as custom software (CardioTrack, Sorbonne Université,[6]) for LV and LA myocardial strain analysis. Such expert analysis resulted in: 1) LV and LA contours throughout the cardiac cycle on both 2-and 4-chamber LAX views, 2) LV peak global longitudinal strain (LVGLS), 3) LA global longitudinal strains at the reservoir (LARS), conduit (LACS) and booster (LABS) phases. Of note, Among the 1368 two- and four-chamber views, 127 had no expert annotations on both LV and LA, because of the presence of artifacts.

To provide heart structure labels as inputs to our networks, LV and LA blood pools were merged into a single left heart blood pool label (LHBP) to avoid presence of small structures and to ensure spatial continuity through the valvular junction between the LA and LV. Besides, LV myocardium was merged with LV blood pool (LVMBP) to avoid the LV myocardial U-shape with unequal thickness.

Due to the large variability in image size, we first zero-padded all images to 512x512 pixels. Then, we normalized intensities and applied a contrast limiting adaptive histogram equalization (CLAHE) to improve local contrast and reduce noise and boundary artifacts. The dataset was then split into training (490 subjects = 34811 images), validation (134 subjects = 9214 images) and test (60 subjects= 4310 images) sets. Both 2- and 4-chamber views as well as time phases of individual subjects are included in the same set. Besides special care was taken to homogeneously distribute datasets from different vendors, field strength and LV/LA global morphology as derived from ground truth contours through end-diastolic surfaces and ejection fractions.

2.2. Dual-ResUNet architecture

The overall architecture of the Dual-ResUnet can be decomposed into two networks (Fig. 1), which are trained separately. Both networks are designed as a combination of a ResNet34 [8] encoding block and a Unet decoding block to allow training of deep neural network while avoiding vanishing or exploding gradients.

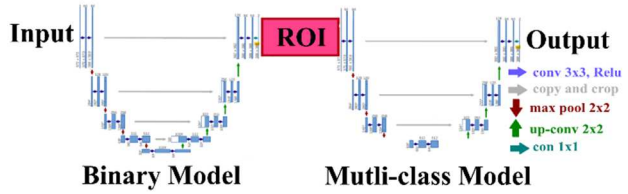

Fig. 1. Dual-ResNet architecture

The first network performs a binary segmentation of the whole left heart, which is then used to center and crop the image around the heart. The aim of such cropping step is to reduce the class imbalance between background and cardiac structures as well as to equally center all our data on the structure of interest. The second network is fed by the cropped centered images and focuses on LHBP and LVMBP prediction. Finally, the segmented images are displayed onto the original field of view (Fig. 2).

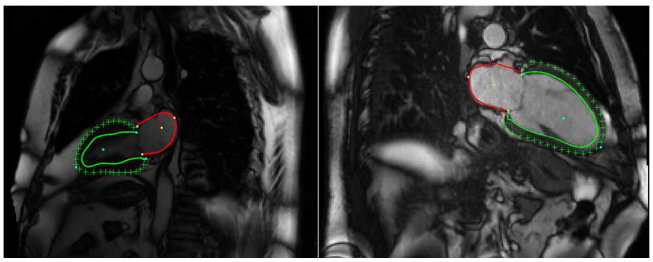

Fig. 2. Examples of predicted LV (green) and LA (red) contours for a healthy subject (left) and a patient (right).

2.3. Loss function and implementation

For both networks, we used the binary cross-entropy (BCE) with logits loss, which combines the BCE loss with a sigmoid layer for further numerical stability than their individual use, since it takes advantage of the log-sum-exp.

Networks were implemented in PyTorch and initialized via a truncated normal distribution centered on 0. The two networks were trained with the Adam optimizer and a training rate of 0.001 and for 50 and 100 epochs, respectively, with a batch size of 32 and 64, respectively. Our model was trained on a dedicated computing hub: dual Intel® Xeon® Gold 6226R 3.90 GHz (16 heart), 512 GB ddr4 RAM, two NVIDIA A6000 GPU (48 GB of dedicated memory each). Of note, data augmentation (random rotation and inversion) was performed on the training and validation sets to expand the datasets to 52215 training and 19131 validation images.

2.4. Feature tracking and strain estimation

The predicted LHBP and LVMBP labels were converted back into LV and LA labels (LV endocardium: Endo, LV epicardium: Epi and LA blood pool) on time phases corresponding to LV and LA maximal dilation. Such contours were then used to initialize the FT algorithm based on the interface between myocardium and the surrounding structures (blood, epicardial fat). Initialized points were tracked in the adjacent time-phases. For such tracking, a 10x10 mm² region of interest is defined around each point of the initial contour. Then, a map of cross-correlation values between such region and its spatial neighborhood on the next time phase is calculated. This cross-correlation map is weighted with additional maps derived from image properties and constraints related to physiological knowledge regarding contractile function, as described in [6]. Tracking-derived LV and LA contours for all phases of the cardiac cycle were then

used to estimate time-varying strain waveforms by calculating temporal changes in longitudinal L(t) length of the myocardium, relative to and normalized by its initial dimension L(t0), obtained at the beginning of the cardiac cycle (t0): Sl(t)=(L(t)–L(t0))/L(t0). LV endocardial and epicardial strain curves were averaged to capture the whole LV myocardial deformation. Finally, predicted strain curves were used to extract LV and LA strain peaks.

2.5. Prediction evaluation

Predicted LV and LA labels were used to estimate Dice scores (DSC) and Hausdorff distances (HD) against reference labels. Further analysis was performed for out-of-distribution data such as data of Vendor 3 and 3T data, while creating a training set without Vendor 3 and 3T data successively and keeping testing set untouched. For strain evaluation, comparison between reference and predicted strain peaks was performed using linear regression and Bland-Altman analysis resulting in correlation coefficients (R), mean biases (μ) and limits of agreement (LoA).

2.6. Model variability against human operators

To assess strain variability, 27 subjects randomly selected from the test set including 15 healthy subjects and 12 patients were analyzed by three operators with varying level of expertise in cardiac image processing (#1: experimented, #2 and 3: beginner to intermediate). Coefficient of difference (COD) in peak strains between predicted and human observers and inter-observer difference (IOD) were computed. Statistical comparison of these differences was performed using the Williams index (WI) along with its confidence interval (CI) [9]. CI lower bound > 1 indicated that the variability between DL-based strains and human observers was within interobservers variability.

3. RESULTS

The first network for binary segmentation of the whole heart resulted in a DSC of 0.94±0.06. The cropped centered images resulting from this network show a more balanced class distribution, as illustrated in Fig. 3.

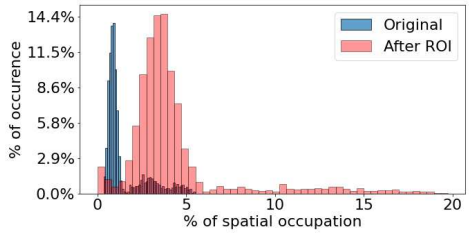

Fig. 3. Spatial whole heart class distribution from original images (blue) and center cropped images (pink)

Multiclass segmentation performances against expert reference are summarized in Table 1, revealing equivalent average performances for the endocardial wall for both LV and LA and slightly higher performances for the less mobile LV epicardium. Overall, 91% of the testing datasets had a mean DSC≥0.85. Stability of our network was evaluated while re-training three times with random weights, resulting in an overall stable mean DSC of 0.90. Most previous studies dedicated to LV and LA AI-based segmentation on LAX images were performed using echocardiography, achieving comparable DSC to those obtained in our study (0.92 for LV and 0.89 for LA [10] or 0.912 for LV Endo and 0.855 for LV Epi [11]). Most MRI studies were focused on LV segmentation on SAX images, while only very few studies analyzed LAX images and such studies mostly reported either LV or LA segmentation. For the LV, DSC of 0.93 [12] was found on a single-center database of 135 patients, while for the LA, a DSC of 0.93 was found [13] on 600 patients collected from a single center. To the best of our knowledge, multi-chamber segmentation from MRI LAX images were proposed in a single study [14], combining CT and MRI of 25 patients, revealing an overall DSC of 0.88 ± 0.03.

	LV Epi	LV Endo	LA	Mean
DSC	0.93 ± 0.07	0.89 ± 0.11	0.89 ± 0.10	0.9 ± 0.09
HD (mm)	6.94 ± 5.99	7.24 ± 5.53	7.66 ± 4.62	7.28 ± 5.42

Table 1. Performances of the multiclass segmentation in comparison to the expert reference.

Interestingly, robustness of our multiclass segmentation through MRI field strengths and vendors revealed a slight drop in datasets acquired at 3T against 1.5T as well as in datasets of vendor #3 as compared to #1 and 2 (Table 2). This might be due to the lower representativeness of these two categories in our training set since 3T data represented 5.7% of the training set and vendor 3 data represented 2.7% of the entire training set. Testing on the unseen data from 3T and vendor #3 resulted in similar average DSC (3T from 0.88 ± 0.04 to 0.87 ± 0.06 , vendor #3 from 0.89 ± 0.05 to 0.88 ± 0.04).

	LV Epi	LV Endo	LA	Mean
1.5T	0.93 ± 0.07	0.90 ± 0.11	0.89 ± 0.10	0.90 ± 0.09
3T	0.92 ± 0.02	0.85 ± 0.05	0.86 ± 0.09	0.88 ± 0.04
Vendor1	0.95 ± 0.02	0.94 ± 0.01	0.92 ± 0.01	0.94 ± 0.01
Vendor2	0.93 ± 0.08	0.89 ± 0.12	0.89 ± 0.11	0.90 ± 0.10
Vendor3	0.92 ± 0.05	0.87 ± 0.08	0.88 ± 0.05	0.89 ± 0.05

Table 2. DSC according to MRI field strength and vendor for both the LV and LA.

Table 3 summarizes reference and predicted LV and LA longitudinal strain measures along with Bland-Altman statistics. Predicted strains were associated with reference as revealed by high correlations (r≥0.70) and low mean bias (≤0.97) between the two measures. In line with segmentation performances, stronger associations were found in the LV as compared to the LA.

	GT	Prediction	R	μ (LoA)
LVGLS(%)	-15.9 ± 4.3	-16.5 ± 4.8	0.96	-0.62 (-3.16 ;1.92)
LARS(%)	31.1 ± 12.7	30.1 ± 10.9	0.70	-0.92 (-19.12 ;17.27)
LABS(%)	14.5 ± 7.6	13.5 ± 6.6	0.74	-0.97 (-11.21 ;9.28)
LACS(%)	16.2 ± 8.5	16.2 ± 8.6	0.75	0.11 (-11.73 ;11.96)

Table 3. Comparison between LV and LA predicted and reference (GT) longitudinal strains.

Table 4 summarizes strain variability results. COD was mostly similar to IOD for all strain indices, resulting in upper bounds of the WI 95% CI higher than 1 demonstrating that the agreement between DL-based method and human observers was in the same range as the agreement between operators, with variable level of expertise.

N=27 views	COD	IOD	WI: mean (CI)
LVGLS	0.08 ± 0.01	0.07 ± 0.01	1.68 (1.60 ;1.75)
LARS	0.25 ± 0.00	0.21 ± 0.01	1.64 (1.59 ;1.70)
LABS	0.22 ± 0.02	0.17 ± 0.01	1.54 (1.46 ;1.62)
LACS	0.26 ± 0.01	0.18 ± 0.02	1.35 (1.28 ;1.41)

Table 4. Comparison between the proposed method strain results and inter-observer variability.

4. CONCLUSION

Reliable strain measures were derived from the proposed approach combining a dual network for the LV and LA segmentation from long axis MRI images and a previously validated feature tracking algorithm [7]. The segmentation network was trained and tested on an original multi-center and multi-vendor dataset including healthy controls and patients with either LV or LA alteration. Stability of the results over MRI field strengths and vendors demonstrates the generalizability of the proposed approach. Interestingly strain measurement variability using the proposed method is comparable to variability between operators with varying level of expertise. Further evaluation of the ability of the fully automated strain measures to characterize LV and LA alterations in specific disease conditions is needed to confirm their reliability.

5. ACKNOWLEDGMENT

We acknowledge the H2020 MAESTRIA project (965286) for funding JL and VN, as well as Djibouti Embassy in Paris (654596/AMB/BGDEF/20-23) for funding MG.

6. COMPLIANCE WITH ETHICAL STANDARDS

Data were collected in protocols reviewed and approved by NCT02517944; NCT03715998; NCT02938910.

7. REFERENCES

[1] M. Evin et al., 'Left atrial aging: a cardiac magnetic resonance feature-tracking study', Am. J. Physiol.-Heart Circ. Physiol., vol. 310, no. 5, pp. H542–H549, Mar. 2016.

[2] A. Schuster, K. N. Hor, J. T. Kowallick, P. Beerbaum, and S. Kutty, 'Cardiovascular Magnetic Resonance Myocardial Feature Tracking', Circ. Cardiovasc. Imaging, vol. 9, no. 4, p. e004077, Apr. 2016.

[3] M. Xavier, A. Lalande, P. M. Walker, F. Brunotte, and L. Legrand, 'An Adapted Optical Flow Algorithm for Robust Quantification of Cardiac Wall Motion From Standard Cine-MR Examinations', IEEE Trans. Inf. Technol. Biomed., vol. 16, no. 5, pp. 859–868, Sep. 2012.

[4] L. Xie, Y. Song, and Q. Chen, 'Automatic left ventricle segmentation in short-axis MRI using deep convolutional neural networks and central-line guided level set approach', Comput. Biol. Med., vol. 122, p. 103877, Jul. 2020.

[5] C. Zotti, Z. Luo, A. Lalande, and P.-M. Jodoin, 'Convolutional Neural Network With Shape Prior Applied to Cardiac MRI Segmentation', IEEE J. Biomed. Health Inform., vol. PP, pp. 1–1, Aug. 2018.

[6] J. Lamy et al., 'Scan-rescan reproducibility of ventricular and atrial MRI feature tracking strain', Comput. Biol. Med., vol. 92, pp. 197–203, Jan. 2018.

[7] T. Yoganathan et al., 'Acute stress induces long-term metabolic, functional, and structural remodeling of the heart', Nat. Commun., vol. 14, no. 1, p. 3835, Jun. 2023.

[8] M. Shafiq and Z. Gu, 'Deep Residual Learning for Image Recognition: A Survey', Appl. Sci., vol. 12, no. 18, Art. no. 18, Jan. 2022.

[9] V. Chalana and Y. Kim, 'A methodology for evaluation of boundary detection algorithms on medical images', IEEE Trans. Med. Imaging, vol. 16, no. 5, pp. 642–652, Oct. 1997.

[10] A. Arafati et al., 'Generalizable fully automated multi-label segmentation of four-chamber view echocardiograms based on deep convolutional adversarial networks', J. R. Soc. Interface, vol. 17, no. 169, p. 20200267, Aug. 2020.

[11] T. Kim, M. Hedayat, V. V. Vaitkus, M. Belohlavek, V. Krishnamurthy, and I. Borazjani, 'Automatic segmentation of the left ventricle in echocardiographic images using convolutional neural networks', Quant. Imaging Med. Surg., vol. 11, no. 5, pp. 1763–1781, May 2021.

[12] R. Shahzad, Q. Tao, O. Dzyubachyk, M. Staring, B. P. F. Lelieveldt, and R. J. van der Geest, 'Fully-automatic left ventricular segmentation from long-axis cardiac cine MR scans', Med. Image Anal., vol. 39, pp. 44–55, Jul. 2017.

[13] W. Bai et al., 'Automated cardiovascular magnetic resonance image analysis with fully convolutional networks'. arXiv, May 22, 2018. Accessed: Jan. 03, 2023.

[14] E. D. Morris, A. I. Ghanem, M. Dong, M. V. Pantelic, E. M. Walker, and C. K. Glide-Hurst, 'Cardiac substructure segmentation with deep learning for improved cardiac sparing', Med. Phys., vol. 47, no. 2, pp. 576–586, 2020.