SMOD - Data Augmentation Based on Statistical Models of Deformation to Enhance Segmentation in 2D Cine Cardiac MRI

Jorge Corral Acero[1(✉)], Ernesto Zacur[1], Hao Xu[1], Rina Ariga[2], Alfonso Bueno-Orovio[3], Pablo Lamata[4], and Vicente Grau[1]

[1] Department of Engineering Science, University of Oxford, Oxford, UK
jor.corral@eng.ox.ac.uk
[2] John Radcliffe Hospital, Oxford, UK
[3] Department of Computer Science, University of Oxford, Oxford, UK
[4] Biomedical Engineering Department, King's College London, London, UK

Abstract. Deep learning has revolutionized medical image analysis in recent years. Nevertheless, technical, ethical and financial constraints along with confidentiality issues still limit data availability, and therefore the performance of these approaches. To overcome such limitations, data augmentation has proven crucial. Here we propose SMOD, a novel augmentation methodology based on Statistical Models of Deformations, to segment 2D cine scans in cardiac MRI. In brief, the shape variability of the training set space is modelled so new images with the appearance of the original ones but unseen shapes within the space of plausible realistic shapes are generated. SMOD is compared to standard augmentation providing quantitative improvement, especially when the training data available is very limited or the structures to segment are complex and highly variable. We finally propose a state-of-art, deep learning 2D cardiac MRI segmenter for normal and hypertrophic cardiomyopathy hearts with an epicardium and endocardium mean Dice score of 0.968 in short and long axis.

Keywords: Data augmentation · Segmentation · Deep learning · Models of deformation · Cardiac MRI

1 Introduction

The expansion of big data has entered the cardiovascular medicine field in recent years [1, 2]. Machine learning approaches and deep learning techniques have been proven successful in medical imaging, and in particular in the cardiovascular field [3, 4]. This has had a strong impact on a wide range of applications, from classification, object detection and registration to phenotype clustering or risk prediction [3].

Among the uses of deep learning techniques in medical imaging, segmentation has been one of the most successful ones [3]. Since the U-net architecture was introduced by Ronneberger et al. in 2015 [5], a number of variations of that structure using

Y. Coudière et al. (Eds.): FIMH 2019, LNCS 11504, pp. 361–369, 2019.
https://doi.org/10.1007/978-3-030-21949-9_39

convolutional neural networks (CNN) have been explored, reporting excellent performance. A good example of this is the segmentation of both left and right ventricles (LV and RV) in cardiac magnetic resonance imaging (MRI) [6–8].

Despite these promising results, limitations and challenges remain [2], in particular concerning the availability, quantity and quality of training data [1]. Technical, ethical or financial constraints may limit data acquisition and confidentiality issues prevent access to data. Furthermore, manual data annotation is a costly and tedious process. Thus, the size of the training datasets in medical imaging is still a limiting factor. This scenario becomes even worse when dealing with specific diseases or applications. Since the performance of deep learning methods depends heavily on the available data, data augmentation and techniques such as transfer learning have become crucial for a better generalization and to avoid overfitting [9–12].

Image cropping, translation, rotation, scaling, flipping and even shearing are routinely applied as standard practice to enlarge datasets [11–14]. Other more complex approaches such as filtering, gamma correction, addition of Gaussian noise or even training a secondary network to perform the augmentation have been considered [9, 10, 12, 13]. Additionally, non-affine transformations, such as random elastic deformations or diffeomorphic transformations, have also been proposed [15, 16].

Statistical shape and deformation models have been used in a variety of medical image analysis tasks, including cardiac image segmentation but not, to the best of our knowledge, in combination with deep learning approaches [17]. We propose a new augmentation methodology based on models of deformation to segment 2D cine cardiac MRI scans and thus overcome the mentioned dataset size limitations. In brief, the shape variability of the training set space is modelled so new images with the gray level appearance of the original ones but unseen shapes within the space of plausible shapes are generated. As a result, the deformations are non-affine, follow a prior and are anatomically meaningful, in contrast with augmentation approaches based on random deformations [15]. We assess the performance of the proposed augmentation methodology versus standard augmentation approaches depending on data availability and on variability of the structure to be segmented. Finally, we build a state-of-art 2D cine MRI cardiac segmenter for both normal and Hypertrophic Cardiomyopathy (HCM) hearts, a condition that presents the largest amount of cardiac anatomical variability and represents the most common cause of sudden cardiac death in the young.

2 Materials and Methods

2.1 Dataset

We make use of two independent datasets containing cardiac MRIs for the assessment of our proposed augmentation methodology:

STACOM Dataset. This refers to the publicly available MICCAI 2018 LV Full Quantification Challenge STACOM dataset (https://lvquan18.github.io/). It contains 145 cine-MR mid-cavity short axis (SA) cardiac cycle sequences, 20 frames each. They were collected from 3 different hospitals, including diverse pathologies and a wide

range of subject ages. The LV endocardium and epicardium have been manually contoured and revised by two cardiac radiologists in all frames. The original pixel size ranges from 0.68 mm/pixel to 2.08 mm/pixel. Pre-processing to align the images was performed, as described in the challenge.

HCM Dataset. This dataset, published at [18], contains 55 HCM phenotype positive, 18 HCM phenotype negative but genotype positive and 37 normal patients scanned at Oxford John Radcliffe Hospital. Only mid-ventricular SA (to facilitate inter-dataset comparisons), horizontal long axis (HLA) and vertical long axis (VLA) end-diastole instances are considered. The LV endocardium, epicardium and left atrium (LA) along with RV endocardium were contoured in the end-diastole frames and double checked by 2 experts. The pixel size ranges from 1.33 to 2.08 mm/pixel.

2.2 Statistical Models of Deformation

Our registration method consists of an initial rigid registration, minimizing the sum of square distances (SSD) of image intensities, followed by the application of Diffeomorphic Log Demons [19] for non-rigid registration ($\sigma_{fluid} = 2$, $\sigma_{diff} = 1.8$ and $\sigma_i/\sigma_x = 0.82$). We use this method to build a model of deformation, in the following way: (a) we generate an atlas representing the specific dataset population; (b) each image is registered to the atlas to calculate one velocity field per image; (c) velocity fields are combined using Principal Component Analysis (PCA) and the deformation model is then used to generate new samples. Methodological details are provided below.

Generation of a Population-Specific Atlas. (1) The atlas is initialized to a randomly selected image among the training set, which we denote as A_0. (2) Each image from the dataset, I, is rigidly registered to A_0 to obtain the registered set I_{T0}. (3) The average of the images in the I_{T0} set is calculated, obtaining the first iteration of the atlas, A_1. (4) I_{T0} is non-rigidly registered to A_1 and averaged to obtain A_2. Step 4 is iterated until convergence is achieved. (5) The resulting atlas, A, is manually segmented to generate A_s.

Set of Velocity Fields to Describe the Shape Variability. The set of segmentations of the images, S, is non-rigidly registered to A_s, obtaining the set of velocity fields, $\{v_i\}$, to diffeomorphically bring each image to the atlas space. The set $\{v_i\}$ encodes the variability of shape of the set of images, I, with respect to the reference A, and can be exploited to generate new deformations, which result in new images.

Dimensionality Reduction and Image Generation. Each velocity of the set $\{v_i\}$ can be organized as the column of a matrix V, to apply PCA and thus reduce the dimensionality. A mode cut capturing 90% variability is considered and the random velocity fields are sampled as follows:

$$v_g = \bar{V} + U \cdot x \cdot d \tag{1}$$

Where $\bar{V}$ is the mean of the velocity fields; U, the eigenvectors of the principal modes; d, the eigenvalues of the principal modes; and x, an array with random numbers following the distribution $N(0, \sigma)$. We set the value of σ to 1 using visual inspection to achieve a realistic but varied set of shapes.

Each of the images, i, is brought to the atlas space applying v_i and transformed back to the image space applying the inverse of the random velocity field, v_g (different from displacement fields, inverse transformations are obtained just inverting the sign of the velocity field). Thus, a new image with the appearance of image i but a random shape within the space of variability of the original images is obtained (see Fig. 1).

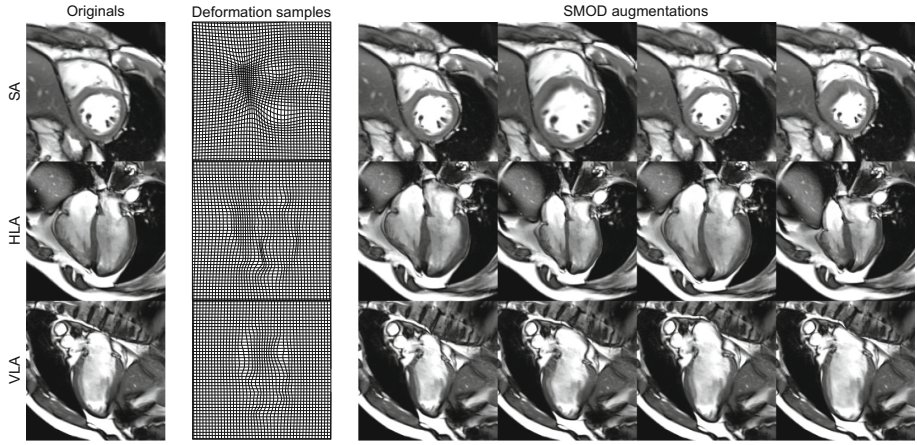

Fig. 1. Left column: Original mid-ventricular SA (top row), HLA (middle row) and VLA (bottom row) end-diastole MRI images. Middle column: Samples of the anatomically meaningful deformations obtained by applying SMOD on originals. Right Column: Augmentation samples from originals based on SMOD.

2.3 Augmentation Strategies

We consider three image augmentation strategies: (a) augmentation based on statistical models of deformations, SMOD, as described above; (b) standard augmentation, STD, based on random rotations (0–360°), flipping and translation (SA: –6 to 6 pixels in x and y, HLA and VLA: –10 to 10 pixels in x and y); and (c) a combination of both, SMOD+. For each cross-validation run, a model of deformations is learnt from scratch based on the particular training set to produce fair results. The augmentation process is "online", meaning that for each epoch the augmented training data fed to the network is different.

2.4 Neural Network

To segment the images, we used a Fully Convolutional Neural Network (FCNN) with skip connections following a U-Net architecture [5], as depicted in Fig. 2. The segmentation output layer has sigmoid activation functions and all the other FCN activations are rectified linear units (ReLU). Two convolutional layers (kernel size 3×3) and two batch normalization (BN) layers form a convolutional block, with the number of filters for each block specified in Fig. 2. Max-pooling and up-sampling layers have a stride of 2 in both dimensions. We used binary cross-entropy as loss function and early

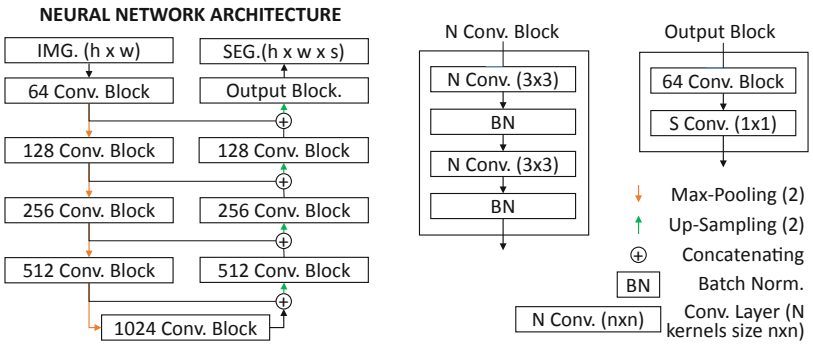

Fig. 2. Graphical overview of the proposed neural network structure.

stopping based on validation loss. The neural networks are implemented in Keras with Tensorflow as backend. The networks are trained with Adadelta optimizer on a NVIDIA GeForce GTX 1080 Ti 10 Gb. The Batch size is 8.

2.5 Assessment

We calculate Dice scores and Hausdorff distances between predicted and original segmentations to assess the performance of the augmentation strategies.

3 Experiments and Results

3.1 STACOM Dataset

We performed three batches of experiments considering different training set sizes to assess the impact of the augmentation strategy depending on data availability. In each batch, the three presented augmentation strategies were tested to segment the LV endocardium and epicardium in SA instances.

In the first batch, a five-fold cross-validation (29 subjects on each fold for testing) is performed. The remaining dataset is randomly split into 100 and 16 subjects, for training and validating data, respectively. In the second and third batches, the split into training/validation/testing data is 10/5/130 and 5/5/135 subjects, respectively. Cross fold validation was not considered in these 2^{nd} and 3^{rd} batches since the testing set, where the assessment is performed, already represents more than the 89% of the whole STACOM dataset in both scenarios. The images are cropped to 80 $\times$ 80 after augmentation and the whole cine sequence is considered in all experiments. That is, a training set of 100 subjects effectively means a training set of 20 $\times$ 100 segmented images.

The Dice score and Hausdorff distance results of the myocardium segmentations, calculated as the region between epicardium and endocardium, are depicted in Fig. 3. Similar results are obtained when evaluating endocardium and epicardium separately.

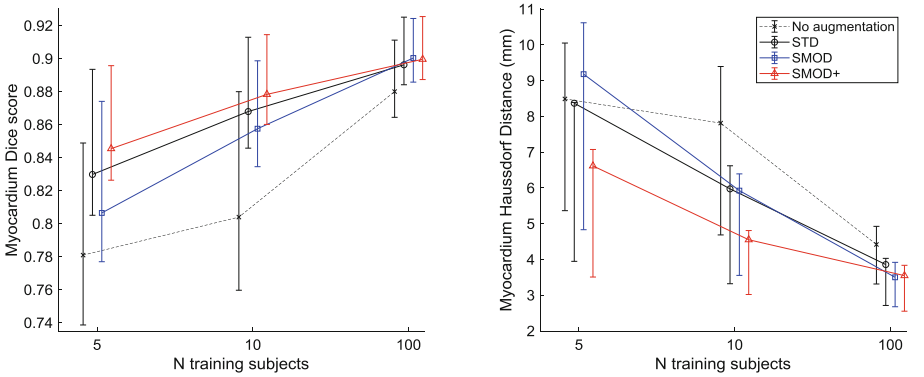

Fig. 3. STACOM dataset results. Mean Dice scores (left) and Hausdorff distances (right) in LV myocardium depending on the number of subjects to train with. Error-bars: 25th and 75th percentiles.

3.2 HCM Dataset

We assess the performance of the augmentation strategies depending on the intrinsic shape variability of the structure to be segmented on the HCM dataset, where not only the LV endocardium and epicardium are segmented but also the LA and the RV endocardium. Experiments to segment each of the three MRI planes were carried out with the three augmentation strategies. Five-fold cross validation is considered in all the experiments. The resulting split, in each of the folds, is: 22 testing, 78 training and 10 validation samples. After augmentation, the images are cropped to 80×80 in SA, and to 224×224 in HLA and VLA.

Figure 4 summarizes the segmentation Dice score results obtained for the different anatomical structures in the different MRI planes training under the three described augmentation strategies. Similar behavior and patterns are observed in Haussdorf distances (not shown). A representative example of the resulting segmentation using SMOD+ is presented in Fig. 5.

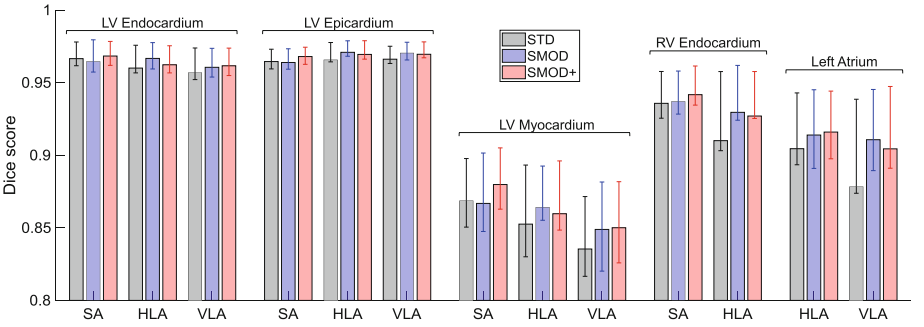

Fig. 4. HCM dataset results. Mean Dice scores and 25th and 75th percentiles grouped by segmented structure and MRI view.

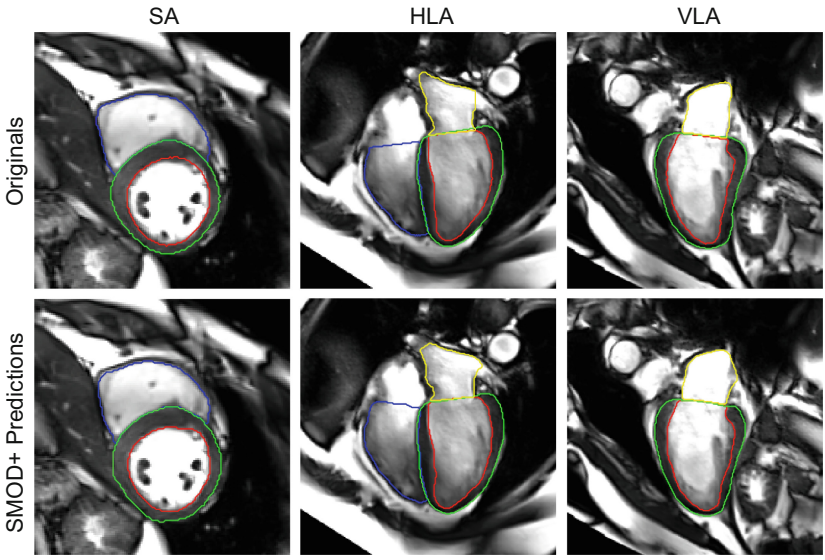

Fig. 5. Top row: LV endocardium (red), LV epicardium (green), LA (yellow) and RV endocardium (blue) ground truth segmentations of mid-ventricular SA (left), HLA (middle) and VLA (right) end-diastole MRI images from HCM dataset. Bottom row: SMOD+ predictions on these unseen instances. (Color figure online)

4 Discussion and Conclusions

Overall, our experiments confirm the importance of data augmentation, as reported in previous literature. Comparing to training with no augmentation, any of the three augmentation strategies significantly improves the results as it can be seen in Fig. 3.

The experiments completed on the STACOM dataset (Fig. 3) show that, while there is little difference among augmentation strategies when training with 100 subjects, SMOD+ outperforms the rest of strategies and becomes more relevant when the number of available subjects to train with decreases. We believe this occurs because the shape of the structures to segment (LV endocardium and epicardium) is reasonably simple and symmetrical, and thus when there are sufficient training samples the space of possible segmentations may be completely described with only rotations, translations and flipping. However, when the number of training subjects drops, standard augmentation is no longer enough to cover the space of variability and the models of deformations, learnt from the few samples, contribute filling the gaps and result in a better training. The observed pattern could be explained because the training network focuses more on local features rather than global ones to perform the segmentation task. The described behavior is confirmed both in Dice scores and Haussdorf distances.

In contrast to the STACOM data results when training with 100 subjects, the experiments on the HCM dataset show a more appreciable improvement in accuracy when using the proposed augmentation method (Fig. 4). SMOD+ reports better results than STD segmenting any structure in any plane, with difference being clearer in LV

myocardium, LA and RV endocardium. We believe this confirms our appreciation as explained above: HCM cases have a larger variation in anatomy due to the disease, and the LA and RV have a less regular shape than the LV. We should also note that, even though the training set size (78 subjects) in the HCM experiments is comparable to STACOM one, only one frame of the cine sequences (end-diastole) is considered in HCM versus 20 instances for the STACOM dataset. Therefore, an alternative explanation for the difference in performance might be the difference in data availability.

A limitation of the work is that only one neural network architecture has been considered. Furthermore, different levels of deformation (σ value, Eq. 1) as well as training in larger and more heterogeneous datasets could be explored. Methods other than PCA, which assumes Gaussianity, could be considered to better describe the distribution of deformations. Moreover, comparison to augmentation based on completely random deformations (no anatomical meaning) could be tested.

We expect the developed augmentation methodology to have more of an impact in 3D since usually fewer samples are available. We also presume that it could be scaled up to other applications and anatomies, particularly those with less regular shapes. Additionally, models of deformation could be learnt from secondary larger datasets and applied to the actual training dataset, enlarging even further the space of plausible shapes and presumably leading to a better generalization. Even though we chose segmentation-to-segmentation registration since less noise is introduced, image-to-image registration could be used to obtain the velocity fields, opening the possibility for learning variability from unsegmented cohorts.

To conclude, (a) a new augmentation methodology based on statistical models of deformation has been developed; (b) it has been proven that it improves the standard augmentation training results when segmenting 2D cardiac MRI slices; (c) the improvement is remarkable especially when the training data available is very limited or the structures to segment are complex and variable; and (d) an accurate state-of-art segmenter for healthy and HCM cine MRI scans has been built.

Acknowledgements. This work was supported by the European Union's Horizon 2020 research and innovation program under the Marie Sklodowska-Curie (g.a. 764738) and by the British Heart Foundation (PG/16/75/32383). Authors are financially supported by a Wellcome Trust Senior Research Fellowship (to PL, 209450/Z/17/Z) and a BHF Intermediate Basic Science Research Fellowship (to ABO, FS/17/22/32644).

References

1. Rumsfeld, J.S., Joynt, K.E., Maddox, T.M.: Big data analytics to improve cardiovascular care: promise and challenges. Nat. Rev. Cardiol. **13**(6), 350–359 (2016)
2. Shameer, K., Johnson, K.W., Glicksberg, B.S., Dudley, J.T., Sengupta, P.P.: Machine learning in cardiovascular medicine: are we there yet? Heart **104**(14), 1156–1164 (2018)
3. Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghafoorian, M., et al.: A survey on deep learning in medical image analysis. Med. Image Anal. **42**, 60–88 (2017)
4. Hosny, A., Parmar, C., Quackenbush, J., Schwartz, L.H., Aerts, H.J.W.L.: Artificial intelligence in radiology. Nat. Rev. Cancer **18**(8), 500–510 (2018)

5. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28
6. Oktay, O., Ferrante, E., Kamnitsas, K., Heinrich, M., Wai, B., Caballero, J., et al.: Anatomically Constrained Neural Networks (ACNNs): application to cardiac image enhancement and segmentation. IEEE Trans. Med. Imaging **37**(2), 384–395 (2018)
7. Poudel, R.P.K., Lamata, P., Montana, G.: Recurrent fully convolutional neural networks for multi-slice MRI cardiac segmentation. In: Zuluaga, M.A., Bhatia, K., Kainz, B., Moghari, M.H., Pace, D.F. (eds.) RAMBO/HVSMR -2016. LNCS, vol. 10129, pp. 83–94. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-52280-7_8
8. Vigneault, D.M., Xie, W., Ho, C.Y., Bluemke, D.A., Noble, J.A.: Ω-Net (omega-net): fully automatic, multi-view cardiac mr detection, orientation, and segmentation with deep neural networks. Med. Image Anal. **48**, 95–106 (2018)
9. Hussain, Z., Gimenez, F., Yi, D., Rubin, D.: Differential data augmentation techniques for medical imaging classification tasks. In: Proceedings American Medical Informatics Association Annual Symposium, vol. 2017, pp. 979–984 (2017)
10. Lemley, J., Bazrafkan, S., Corcoran, P.: Smart augmentation learning an optimal data augmentation strategy. IEEE Access **5**, 5858–5869 (2017)
11. Asperti, A., Mastronardo, C.: The effectiveness of data augmentation for detection of gastrointestinal diseases from endoscopical images. In: Proceedings 11th International Joint Conference on Biomedical Engineering Systems and Technologies, pp. 199–205 (2018)
12. Goodfellow, I., Bengio, Y., Courville, A.: Deep Learning, vol. 13, no. 1. MIT Press, Cambridge (2017)
13. Perez, L., Wang, J.: The effectiveness of data augmentation in image classification using deep learning. CoRR, vol. abs/1712.0 (2017)
14. Dong, H., Yang, G., Liu, F., Mo, Y., Guo, Y.: Automatic brain tumor detection and segmentation using U-Net based fully convolutional networks. In: Valdés Hernández, M., González-Castro, V. (eds.) MIUA 2017. CCIS, vol. 723, pp. 506–517. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-60964-5_44
15. Castro, E., Cardoso, J.S., Pereira, J.C.: Elastic deformations for data augmentation in breast cancer mass detection. In: IEEE EMBS International Conference on Biomedical & Health Informatics (BHI), pp. 230–234 (2018)
16. Arteaga, M.O., Sørensen, L., Cardoso, J., Modat, M., et al.: PADDIT: probabilistic augmentation of data using diffeomorphic image transformation, October 2018. arXiv:1810.01928
17. Alba, X., et al.: An algorithm for the segmentation of highly abnormal hearts using a generic stat. shape model. IEEE Trans. Med. Imaging **35**(3), 845–859 (2016)
18. Raman, B., Ariga, R., Spartera, M., Sivalokanathan, S., Chan, K., Dass, S., et al.: Progression of myocardial fibrosis in hypertrophic cardiomyopathy: mechanisms and clinical implications. Eur. Heart J. Cardiovasc. Imaging **20**(2), 157–167 (2019)
19. Lorenzi, M., Ayache, N., Frisoni, G.B., Pennec, X., Alzheimer's Disease Neuroimaging Initiative (ADNI): LCC-Demons: a robust and accurate symmetric diffeomorphic registration algorithm. Neuroimage **81**, 470–483 (2013)