A Multi-view Crossover Attention U-Net Cascade with Fourier Domain Adaptation for Multi-domain Cardiac MRI Segmentation

Marcel Beetz(✉), Jorge Corral Acero, and Vicente Grau

Institute of Biomedical Engineering, Department of Engineering Science, University of Oxford, Oxford OX3 7DQ, UK
marcel.beetz@eng.ox.ac.uk

Abstract. Cardiac image segmentation is a crucial step in clinical practice as it allows for the assessment of cardiac morphology and the quantification of image-based biomarkers. While deep learning methods have recently achieved near human-level performance on large, single-domain cine MRI datasets, their accuracy decreases considerably in more complex multi-domain settings, limiting their clinical applicability. To this end, we propose a novel multi-view crossover cascade approach combined with both shape and appearance augmentations for effective multi-domain cardiac image segmentation. Our cascade consists of two Attention U-Net paths that share information across different views and an intermediate heart location crop to reduce variance and improve label balance. In addition to multiple shape augmentations (scaling, elastic deformations, grid distortions, etc.) and histogram matching, we introduce multi-scale Fourier Domain Adaptation to cardiac image analysis. We evaluate both the crossover cascade and the augmentations on the cine MRI dataset of the M&Ms-2 challenge and outperform a U-Net benchmark by respective Dice score increases of ∼0.02 and ∼0.03.

Keywords: Multi-view cascade · Crossover cascade · Fourier domain adaptation · Attention U-net · Multi-domain cardiac image segmentation · Histogram matching · Deep learning · Cardiac MRI

1 Introduction

Cardiac Magnetic Resonance (CMR) imaging is the gold standard for the accurate non-invasive quantification of cardiac function and structure [20]. Anatomical segmentation is a standard pre-requisite for the calculation of both 2D and 3D image-based biomarkers [1,3,4,18] with proven prognostic value in the management of cardiac diseases [12]. Facilitated by the expansion of big data in the cardiovascular medicine field, deep learning models have marked a watershed moment towards the automation of the analysis [8,19]. Nevertheless, CMR analysis mostly remains manual in clinical practice, with the incurred time burden and associated costs

E. Puyol Antón et al. (Eds.): STACOM 2021, LNCS 13131, pp. 323–334, 2022.
https://doi.org/10.1007/978-3-030-93722-5_35

[19]. The main challenges limiting the automation include: (a) the intrinsic complexity of cardiac dynamics and the anatomical variability of the heart; (b) the heterogeneity in imaging acquisitions, introduced by different scanners, centers, and protocols; and (c) the limited availability of clinical data, constrained by financial, technical, ethical and confidentiality issues [13,19,21]. In order to address these challenges, data augmentation has become standard practice in fully automatic method designs to enlarge datasets and expose the network to higher degrees of variability. This includes simple, routinely-used approaches (cropping, translation, rotation, and flipping) [5,7], deformation-based techniques [11], as well as augmentations addressing imaging heterogeneity [5,23]. The above-mentioned methods aim to overcome the segmentation challenges by exploiting the available training data. Alternatively, a number of architectural variations have been introduced to improve segmentation performance, such as Attention U-Nets [16], cascaded approaches [10,22], domain adaptation methods [5,9], multi-view techniques [6], and ensembles [2,5]. In this work, we explore both data-driven methods and architectural enhancements in the context of the Multi-Disease, Multi-View & Multi-Center RV Segmentation in Cardiac Magnetic Resonance Imaging (MRI) Challenge (M&Ms-2) [15]. Its main goal is the segmentation of the right ventricle (RV), which is especially challenging given its highly complex and variable shape and its sometimes ill-defined borders. Thus, we deploy a novel multi-view crossover cascade pipeline, based on Attention U-Nets, that integrates the information from short and long axis. In addition, we train our approach with augmentations to populate both spaces of anatomical variability and imaging heterogeneity and thereby introduce a multi-scale version of Fourier Domain Adaptation [23] to cardiac image analysis as a new cross-domain data augmentation approach.

2 Methods

We first briefly describe the dataset of the M&Ms-2 challenge before giving an explanation of our proposed approaches.

2.1 Dataset

The M&Ms-2 training dataset consists of both short-axis (SA) and long-axis (LA) cine MRI acquisitions of 40 healthy subjects and 120 patients. It exhibits considerable variation in terms of scanner type, acquisition center and protocol, and contains multiple diseases affecting both ventricles. The pixel sizes vary from $\sim$0.68 to $\sim$1.63 mm for the SA data and from $\sim$0.68 to $\sim$1.72 mm for the LA data while the image resolution ranges from 192 to 512 pixels for SA and from 208 to 512 pixels for LA images. The left ventricular (LV) bloodpool, the LV myocardium, and the right ventricular (RV) bloodpool were manually segmented by clinical experts and serve as the ground truth for evaluation. We randomly split the dataset into train, validation, and test sets of 120, 8, and 32 cases respectively. In addition to the challenge training dataset available for method development, a test dataset with 160 cases is withheld from all challenge participants by the organizers and used only for the final evaluation of all challenge submissions.

2.2 Preprocessing

As preprocessing steps, we first adjust the pixel sizes of all SA images to ∼1.33 mm and all LA images to ∼1.43 mm using linear interpolation for the MR images and nearest neighbor interpolation for the segmentation masks. We then either crop or pad the resulting images to the same resolution of 256×256 pixels for both SA and LA data. Finally, we apply min-max normalization and intensity clipping (10th and 96th percentiles) to each SA and LA image.

2.3 Multi-view Crossover Cascade

A core feature of the M&Ms-2 dataset is the availability of paired SA and LA images. This allows information of the respective other view to be used in the segmentation task which is especially beneficial for determining the often difficult-to-pinpoint RV boundary or when segmenting the basal slices to avoid confusion between the right atrium and ventricles. In order to enable this information sharing between the SA and LA views in a deep learning approach, we propose a novel multi-view crossover cascade pipeline of Attention U-Nets (Fig. 1).

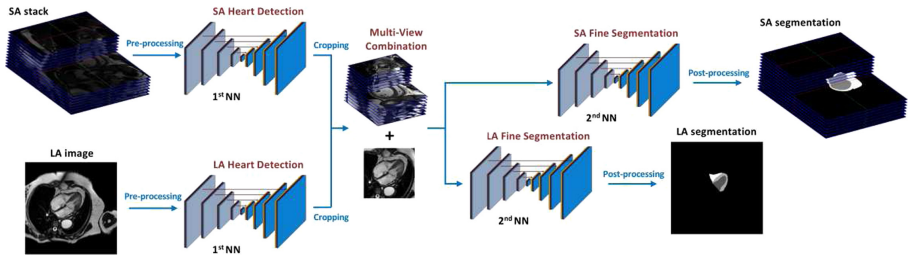

Fig. 1. Overview of the proposed multi-view crossover cascade pipeline. Two separate neural networks first locate the heart in the long-axis and short-axis input images respectively. Their predictions are then used to crop out a rectangular patch around the heart from the input images. The cropped images from both views are then combined into a new crossover volume that is fed into two additional separate neural networks which output the final predictions of the segmentation masks corresponding to the SA and LA input images respectively.

The pipeline consists of four main steps. First, two Attention U-Nets locate the heart in the LA and SA images respectively. Second, the heart location predictions are used to apply a crop of size 128×128 pixels centered at the heart's center of mass to both SA and LA images. Third, the information of both views is combined into a single volume. For LA segmentation, we concatenate the cropped LA image and the three cropped mid-cavity SA slices, as their segmentation quality is usually very high and it thus provides reliable information about the LV and RV boundaries and shapes to the LA segmentation task. The choice of three SA slices reduces the negative effects of potential segmentation errors on

individual SA slices while still maintaining a high probability that ventricular anatomy is present in the given slices. For SA segmentation, we concatenate the respective cropped SA slice with the cropped LA image enabling access of the SA slices to further anatomical information in the basal and apical heart areas. Finally, each of the two volumes is passed through one of two additional Attention U-Nets to produce the final segmentation masks.

2.4 Shape Augmentations

Our first proposed set of augmentations is aimed at representing the considerable spatial variability of anatomical shapes and sizes, both of the heart and its surrounding tissue. In addition to commonly used affine transforms (flipping, rotating, and translating), we therefore use scaling, elastic deformations, grid distortions, and optical distortions as additional augmentation methods. We hypothesize these to be crucial for our highly multi-domain dataset and effective in capturing different disease phenotypes, such as the size variations of the dilated left and right ventricle condition. This is especially important considering the unknown diseases in the external test dataset of the M&Ms-2 challenge.

2.5 Domain Adaptation

The general appearance of cine MR images is heavily influenced by its acquisition conditions, such as the type of MRI scanner, acquisition protocol, and center location. To mimic such differences in our training dataset, we propose two types of appearance augmentations: histogram matching and Fourier Domain Adaptation (FDA) [23]. In both cases, a target image is randomly selected for each input image with the aim of transforming the input image in a way that matches the global appearance of the target image without altering its core structural content. Figure 2 shows the effect of each on both a SA (top row) and LA sample (bottom row) image.

Histogram matching achieves this similarity between images by first determining the histogram of both the input and target images, then deriving the respective cumulative probability distributions from each histogram, and finally creating a mapping function from each graylevel value in the input image to the target image. This mapping function can then be applied to the input image to obtain the transformed image. In Fourier Domain Adaptation, both the input and target images are first converted into Fourier space using a standard 2D discrete Fourier transform. A rectangular patch centered at the Fourier image center is then exchanged between the input and target Fourier images before both are converted back to the spatial domain using the inverse Fourier transform. Since the central locations in Fourier space represent low-frequency signals, only global image information is transferred from the target to the input image while its structural integrity remains mostly unchanged. The size of the patch is controlled by a parameter β that is crucial for maintaining a good trade-off between excessive and barely noticeable appearance changes. In this work, we

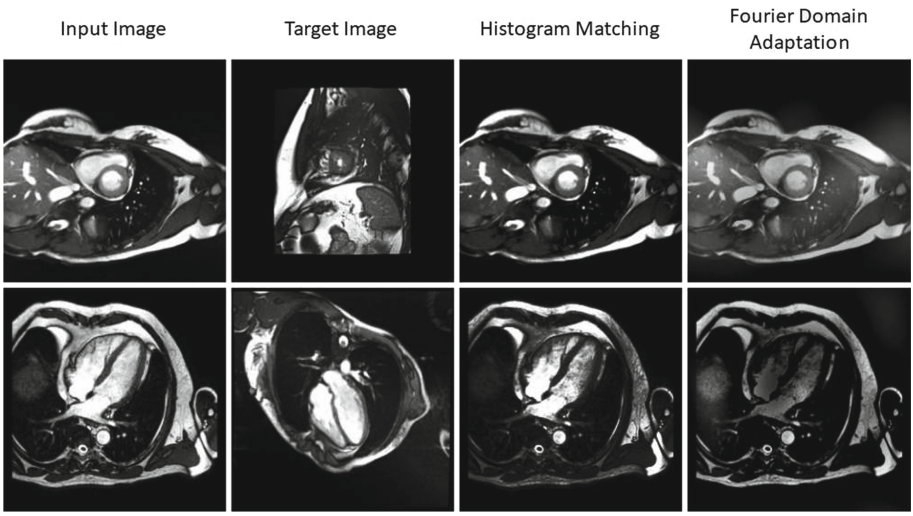

Fig. 2. Effect of histogram matching and Fourier Domain Adaptation on one sample short-axis (top row) and one sample long-axis image (bottom row).

randomly sample the β values from a range between 0.001 and 0.01, thereby extending the original method to a multi-scale FDA setting.

2.6 Postprocessing

As postprocessing steps, we apply hole closing and select only the largest connected RV component in the prediction mask. We also completely remove the right ventricle from slices which do not contain any predicted left ventricle.

2.7 Implementation

We conduct all experiments and method development on a GeForce RTX 2070 Graphics Card with 8 GB memory. All networks are implemented using the PyTorch deep learning framework [17] and trained using the Adam optimizer [14] with a batch size of 4. We stop the training process once network performance on the validation dataset has not improved for 50 epochs and select the best checkpoint as our final model.

3 Experiments

In order to evaluate both the architectural (Sec. 3.1) and data-driven (Sec. 3.2) approaches we select a baseline method as a benchmark for each of the two and compare it to the performance of the proposed changes using the Dice coefficients of three anatomical structures (LV bloodpool, LV myocardium, RV bloodpool)

in a 5-fold cross-validation setup. Furthermore, we report the RV segmentation results of our final challenge submission on the external challenge test dataset in terms of both Dice scores and Hausdorff distances (Sec. 3.3).

3.1 Architectural Methods

We choose a standard U-Net as a benchmark to assess the effect of three architectural changes, an Attention U-Net [16] and a cascaded Attention U-Net approach, each of which trained separately for SA and LA data, as well as the proposed multi-view crossover Attention U-Net cascade (Sec. 2.3). All methods are trained using the same baseline augmentations (flipping, rotating, translating) to enable a fair comparison. The results on the test dataset for both SA and LA data are depicted in Table 1.

Table 1. Segmentation results of proposed architectures.

Image type	Method	Dice score		
		LV bloodpool	LV myocardium	RV bloodpool
Short-axis	Baseline U-Net	0.88 (± 0.1)	**0.80** (± 0.08)	0.86 (± 0.08)
	Attention U-Net	**0.89** (± 0.08)	**0.80** (± 0.07)	**0.87** (± 0.08)
	Cascade	0.88 (±0.11)	**0.80** (± 0.12)	0.84 (± 0.15)
	Crossover cascade	**0.89** (± 0.14)	**0.80** (± 0.13)	0.85 (± 0.17)
Long-axis	Baseline U-Net	0.89 (± 0.15)	0.77 (± 0.14)	0.83 (± 0.21)
	Attention U-Net	**0.92** (± 0.09)	**0.80** (± 0.12)	0.87 (± 0.15)
	Cascade	0.91 (± 0.11)	0.77 (± 0.13)	0.87 (± 0.14)
	Crossover cascade	**0.92** (± 0.13)	**0.80** (± 0.13)	**0.88** (± 0.15)

All values represent mean (± standard deviation).

We find similar results across all four architectures in the SA dataset, while the standalone Attention U-Net and the crossover cascade achieve the highest scores for the LA data.

When assessing the methods' performance on individual cases and slices, we observe that prediction quality is relatively similar for mid-cavity SA images, but varies to a much larger extent for basal and apical slices. Figure 3 depicts the qualitative prediction results for three sample cases.

For example, the first row in Fig. 3 shows a basal SA slice for which the crossover pipeline was able to correctly detect the absence of the heart while the other approaches erroneously predicted a left ventricular structure.

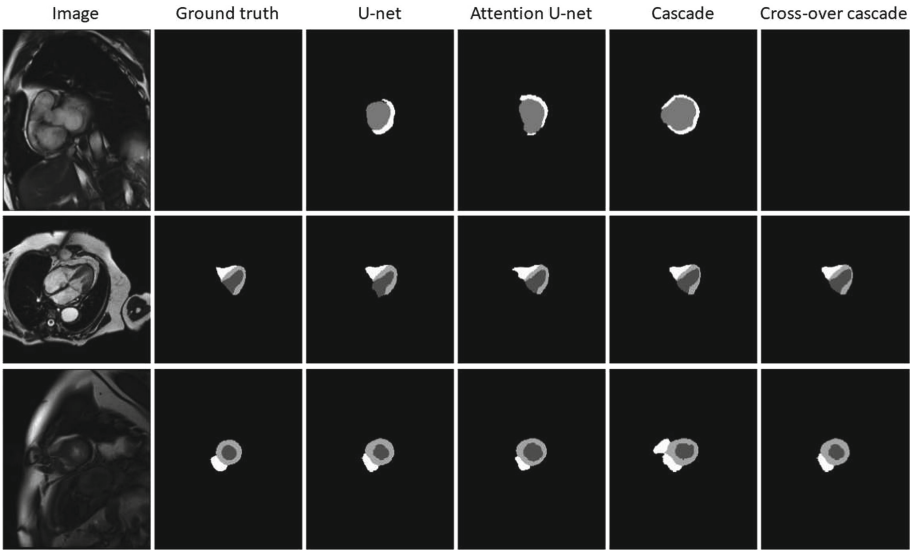

Fig. 3. Qualitative segmentation results of all analyzed architectures on three sample images.

3.2 Data-Driven Methods

In order to evaluate the effects of our proposed data-driven changes, we choose a U-Net architecture with basic augmentations (flipping, rotation, translation) as our baseline. We then compare it to both a U-Net with basic and shape augmentations (Sec. 2.4) and a U-Net with basic and appearance augmentations (Sec. 2.5) for both SA and LA data (Table 2).

Table 2. Segmentation results of baseline, shape, and appearance augmentations.

Image type	Method	Dice score		
		LV bloodpool	LV myocardium	RV bloodpool
Short-axis	Baseline	0.88 (± 0.10)	0.80 (± 0.08)	0.86 (± 0.08)
	+ Shape	**0.90** (± 0.08)	0.81 (± 0.09)	**0.88** (± 0.07)
	+ Appearance	**0.90** (± 0.06)	**0.82** (± 0.06)	**0.88** (± 0.06)
Long-axis	Baseline	0.89 (± 0.15)	0.77 (± 0.14)	0.83 (± 0.21)
	+ Shape	0.90 (± 0.11)	0.80 (± 0.12)	0.86 (± 0.16)
	+ Appearance	**0.93** (± 0.08)	**0.82** (± 0.09)	**0.88** (± 0.13)

All values represent mean (± standard deviation).

Both shape and appearance augmentations improve performance compared to the baseline in both views with differences between methods larger for LA data. Appearance augmentations achieve the highest Dice scores for each cardiac substructure and view.

3.3 Ensemble

Following the experimental results described in Sec. 3.1 and Sec. 3.2, we combine the architectural and data-driven changes that showed improvements over their respective baselines in an ensemble approach as our final submission to the M&Ms-2 challenge. To this end, we train both the crossover cascade and the normal cascade with Attention U-Net architectures and both augmentation types (baseline plus shape and appearance) in a 5-fold cross-validation setting on the training dataset of the challenge and average the outputs of each network to obtain the ensemble predictions. The results of this approach on the external challenge test dataset are depicted in Table 3 as reported by the challenge organizers. Both Dice scores and Hausdorff distances between predicted and ground truth segmentations were chosen as evaluation metrics using a 2D-based formulation for LA images and a 3D-based one for SA data.

Table 3. Results of our ensemble method on the external test dataset of the M&Ms-2 challenge as provided by the challenge organizers. Row 1 and row 2 show mean values for each image type. Row 3 represents the weighted average values across both image types. The respective weight coefficients were selected by the challenge organizers and set to 0.75 for SA data and 0.25 for LA images [15].

Image type	Right ventricle	
	Dice score	Hausdorff distance
Short-axis	0.83	17.62
Long-axis	0.85	10.95
Average	0.84	15.95

We observe an overall better performance of our method on long-axis images than on short-axis images in terms of both metrics. Dice scores for both image types are lower compared to the ones obtained in our experiments on the challenge training dataset.

The challenge organizers also provide our method's segmentation results separated by the pathologies present in the external challenge test dataset (Table 4).

Overall, our method achieves both its best and worst results on pathological imaging data while intermediate scores are recorded for healthy subjects. The highest Dice scores and lowest Hausdorff distances for long-axis images are obtained for tetralogy of fallot and hypertrophic cardiomyopathy patients respectively. Furthermore, we observe the best performance on short-axis images for congenital arrhythmogenesis patients in terms of both evaluation metrics.

Table 4. Results per pathology of our ensemble method on the external test dataset of the M&Ms-2 challenge as provided by the challenge organizers.

Image type	Pathology	Right ventricle	
		Dice	Hausdorff
Short-axis	Tricuspid Regurgitation	0.82 (± 0.13)	19.46 (± 9.99)
	Dilated Right Ventricle	0.86 (± 0.13)	16.90 (± 7.57)
	Interatrial Communication	0.77 (± 0.26)	19.11 (± 10.67)
	Dilated Left Ventricle	0.81 (± 0.18)	19.21 (± 11.07)
	Congenital Arrhythmogenesis	**0.88** (± 0.05)	**13.49** (± 4.49)
	Hypertrophic Cardiomyopathy	0.81 (± 0.20)	18.78 (± 11.37)
	Tetralogy of Fallot	0.85 (± 0.09)	16.75 (± 8.56)
	Normal	0.85 (± 0.11)	15.57 (± 8.95)
Long-axis	Tricuspid Regurgitation	0.84 (± 0.24)	9.61 (± 12.61)
	Dilated Right Ventricle	0.82 (± 0.20)	13.51 (± 14.46)
	Interatrial Communication	0.81 (± 0.28)	11.81 (± 12.15)
	Dilated Left Ventricle	0.88 (± 0.12)	8.02 (± 7.25)
	Congenital Arrhythmogenesis	0.86 (± 0.21)	8.99 (± 9.96)
	Hypertrophic Cardiomyopathy	0.89 (± 0.14)	**7.44** (± 8.79)
	Tetralogy of Fallot	**0.90** (± 0.04)	8.45 (± 4.08)
	Normal	0.84 (± 0.18)	16.52 (± 22.07)

All values represent mean (± standard deviation).

4 Discussion and Conclusion

In this work, we have presented a novel multi-view crossover cascade pipeline with both shape and appearance augmentations capable of improving performance for challenging multi-domain cardiac image segmentation.

In our experiments, the crossover cascade slightly outperforms a standard cascade with separate view processing, showing the utility of information sharing across different views. This is especially effective in the LA segmentation task, indicating that the highly reliable mid-cavity SA information is helpful in delineating anatomical LA boundaries. Since the stand-alone Attention U-Net achieved similar results as the crossover Attention U-Net cascade, we conclude that the usage of attention blocks is equally effective for accurately focusing on the correct heart location compared to a cascaded approach in the given dataset. However, we also find noticeable variations in performance between different slices, which shows that cascaded approaches can learn correct mappings in situations where the Attention U-Net does not. In general, we observe a large variability in Dice scores across the 5-fold cross-validation experiments, which could be a reflection of the highly variable dataset (regarding pixel size, image resolution, disease type, etc.) and indicates that our method would benefit from further experiments.

Regarding the data-driven changes, we find positive effects of both shape and appearance augmentations. This demonstrates the special importance of augmentation methods in highly variable multi-domain datasets since they expose the network to a larger variety of inputs during training which in turn strengthens its generalization ability. Furthermore, we observe that interchanging domain information between images as part of the appearance augmentations leads to greater performance increases and smaller standard deviations than the analyzed shape variations which were constrained to individual images in this work. The increased robustness could be explained by the small amounts of additional noise introduced into the images to a varying degree by the FDA approach, forcing the network to be less susceptible to image artifacts.

Our ensemble method performs better on the LA images than on the SA images of the external challenge test dataset in terms of both evaluation metrics, which is in line with our own experimental results. The considerably larger Hausdorff distance indicates that more extreme outliers are present in the RV predictions of the short-axis data. We hypothesize, that this might be caused by the 3D-based metrics used to evaluate the performance on the SA stack where distances between wrong and correct predictions are captured along all three spatial dimensions instead of the 2D-based calculations used for the long-axis data. Since the right ventricle is often only present in some of the slices of the SA stack, small erroneous predictions in slices far away from the true anatomy can lead to large Hausdorff distances despite negligible effects on 3D Dice scores. In addition, the results indicate that the cross-view information exchange in our proposed cascade architecture is more important for LA data since it only consists of one 2D image as opposed to the multiple slices available in the SA view which facilitate the network's segmentation task.

When analyzing the pathology-specific results, we find that our ensemble approach achieves similar levels of performance for the previously unseen disease cases (tricuspid regurgitation and dilated right ventricle) and for the healthy subjects. On the one hand, this shows that the Dice score decrease between the results observed in our own experiments and the ones on the external challenge dataset was likely due to other domain changes unrelated to disease phenotypes. On the other hand, it demonstrates that our method was able to generalize well to new pathologies. In this regard, the access to multi-view information in the crossover pipeline might have played an important role, for example when determining the boundary of the dilated RV. Besides this, the good generalization results could also be attributed to the proposed augmentation strategies since, for example, the scaling and deformation-based shape changes might have mimicked the dilation of the RV while the domain adaption techniques could have exposed the network to a wider variety of acquisition conditions during training.

Acknowledgments. The authors declare that the segmentation methods they developed for participating in the M&Ms-2 challenge did not use any additional datasets besides the ones provided by the organizers. The work of M. Beetz was supported by the Stiftung der Deutschen Wirtschaft (Foundation of German Business). The work of J. Corral-Acero was supported by the EU's Horizon 2020 research and innovation

program under the Marie Sklodowska Curie (g.a. 764738 to JCA). The work of V. Grau was supported by the CompBioMed 2 Centre of Excellence in Computational Biomedicine (European Commission Horizon 2020 research and innovation programme, grant agreement No. 823712).

References

1. Alfakih, K., Plein, S., Thiele, H., Jones, T., Ridgway, J.P., Sivananthan, M.U.: Normal human left and right ventricular dimensions for MRI as assessed by turbo gradient echo and steady-state free precession imaging sequences. J. Magn. Reson. Imaging Official J. Int. Soc. Magn. Reson. Med. **17**(3), 323–329 (2003)
2. Audelan, B., Hamzaoui, D., Montagne, S., Renard-Penna, R., Delingette, H.: Robust fusion of probability maps. In: Martel, A.L., et al. (eds.) MICCAI 2020. LNCS, vol. 12264, pp. 259–268. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-59719-1_26
3. Bai, W., et al.: A bi-ventricular cardiac atlas built from 1000+ high resolution mr images of healthy subjects and an analysis of shape and motion. Med. Image Anal. **26**(1), 133–145 (2015)
4. Beetz, M., Banerjee, A., Grau, V.: Biventricular surface reconstruction from cine MRI contours using point completion networks. In: 2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI), pp. 105–109 (2021)
5. Campello, V.M., et al.: Multi-centre, multi-vendor and multi-disease cardiac segmentation: The M&Ms challenge. IEEE Trans. Med. Imaging **40**, 3543–3554 (2021)
6. Chen, C., Biffi, C., Tarroni, G., Petersen, S., Bai, W., Rueckert, D.: Learning shape priors for robust cardiac mr segmentation from multi-view images. In: Shen, D., et al. (eds.) MICCAI 2019. LNCS, vol. 11765, pp. 523–531. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-32245-8_58
7. Chen, C., et al.: Deep learning for cardiac image segmentation: a review. Front. Cardiovasc. Med. **7**, 25 (2020)
8. Corral-Acero, J., et al.: The digital twin to enable the vision of precision cardiology. Eur. Heart J. **41**(48), 4556–4564 (2020)
9. Corral Acero, J., Sundaresan, V., Dinsdale, N., Grau, V., Jenkinson, M.: A 2-step deep learning method with domain adaptation for multi-centre, multi-vendor and multi-disease cardiac magnetic resonance segmentation. In: Puyol Anton, E., et al. (eds.) STACOM 2020. LNCS, vol. 12592, pp. 196–207. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-68107-4_20
10. Corral Acero, J., et al.: Left ventricle quantification with cardiac MRI: deep learning meets statistical models of deformation. In: Pop, M., et al. (eds.) STACOM 2019. LNCS, vol. 12009, pp. 384–394. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-39074-7_40
11. Corral Acero, J., et al.: SMOD - data augmentation based on statistical models of deformation to enhance segmentation in 2d cine cardiac MRI. In: Coudière, Y., Ozenne, V., Vigmond, E., Zemzemi, N. (eds.) FIMH 2019. LNCS, vol. 11504, pp. 361–369. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-21949-9_39
12. Dall'Armellina, E.: From recognized to novel quantitative CMR biomarkers of lv recovery: a paradigm shift in acute myocardial infarction imaging (2017)
13. Dey, D., et al.: Artificial intelligence in cardiovascular imaging: jacc state-of-the-art review. J. Am. Coll. Cardiol. **73**(11), 1317–1335 (2019)
14. Kingma, D.P., Ba, J.: Adam: a method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)

15. Martin-Isla, C.: Multi-disease, multi-view & multi-center right ventricular segmentation in cardiac MRI (M&Ms-2) (2021). https://www.ub.edu/mnms-2/
16. Oktay, O., et al.: Attention u-net: learning where to look for the pancreas. arXiv preprint arXiv:1804.03999 (2018)
17. Paszke, A., et al.: Pytorch: an imperative style, high-performance deep learning library. Adv. Neural Inf. Process. Syst. **32**, 8026–8037 (2019)
18. Petersen, S.E., et al.: Reference ranges for cardiac structure and function using cardiovascular magnetic resonance (cmr) in caucasians from the uk biobank population cohort. J. Cardiovasc. Mag. Reson. **19**(1), 18 (2017)
19. 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)
20. Stokes, M.B., Roberts-Thomson, R.: The role of cardiac imaging in clinical practice. Aust. Prescriber **40**(4), 151 (2017)
21. Tao, Q., et al.: Deep learning-based method for fully automatic quantification of left ventricle function from cine mr images: a multivendor, multicenter study. Radiology **290**(1), 81–88 (2019)
22. 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)
23. Yang, Y., Soatto, S.: FDA: fourier domain adaptation for semantic segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 4085–4095 (2020)