Reconstruction of 3D Cardiac MR Images from 2D Slices Using Directional Total Variation

Nicolas Basty[1(✉)], Darryl McClymont[2], Irvin Teh[2,3], Jürgen E. Schneider[2,3], and Vicente Grau[1]

[1] Department of Engineering Science, Institute of Biomedical Engineering,
University of Oxford, Oxford, UK
nicolas.basty@eng.ox.ac.uk
[2] Division of Cardiovascular Medicine, Radcliffe Department of Medicine,
University of Oxford, Oxford, UK
[3] Leeds Institute of Cardiovascular and Metabolic Medicine,
University of Leeds, Leeds, UK

Abstract. Cardiac MRI allows for the acquisition of high resolution images of the heart. Long acquisition times of MRI make it impractical to image the full heart in 3D at high resolution. As a result, multiple 2D images are commonly acquired with a slice thickness greater than the in-plane resolution. One way of achieving isotropic high-resolution images is to apply post-processing techniques such as super-resolution to produce high resolution images from low resolution input. We use short-axis stacks as well as orthogonal long-axis views in a super-resolution framework, constraining the reconstruction using the contrast independent directional total variation algorithm to produce a high resolution 3D reconstruction with isotropic resolution. The 3D reconstruction retains the contrast of the short-axis stack, but incorporates the edge information from both the short-axis and the long-axis stacks. Results show improved reconstructions, with a segmentation voxel misclassification rate of 3.51% as opposed to 4.27% using linear interpolation.

Keywords: 3D image reconstruction · Super-resolution · Cardiac MRI · Regularisation · Directional total variation

1 Introduction

Magnetic Resonance Imaging (MRI) is a non-ionising and non-invasive imaging method exhibiting particularly good soft tissue contrast. It provides structural as well as functional information, and it is taken as gold standard for soft tissue imaging notably for vital organs such as the brain and the heart. In a clinical setting, patient motion from breathing, cardiac motion as well as long acquisition times restrict imaging from being performed at full isotropic 3D high resolution routinely. As a consequence, 2D slices with highly anisotropic voxels are acquired. The standard clinical protocol is to image a set of stacked parallel short-axis (SA) images and a smaller number of long-axis (LA) views orthogonal to the SA stack

covering most of the heart. 2D low resolution (LR) slices are typically acquired with in-plane pixel size of 1–2 mm and slice thickness of 6–10 mm with a small slice gap of 2–4 mm [1]. SA image stacks often fail to capture the apex or the base of the heart appropriately, and data for those essential features is only contained in the LA slices. In practice, one 4-chamber view and one 2-chamber view are acquired, making it very imbalanced between numbers of SA and LA acquisitions. However, this choice of orientation is not made with the aim of reconstructing a 3D volume.

Methods for improving image quality may occur at different points along the image reconstruction pipeline. For example, compressed sensing algorithms work with the acquired k-space data, reconstructing images using sparse modelling. This serves to decrease imaging time whilst still giving adequate image quality [2]. The technique has been successfully applied to dynamic 2D CMR reconstructions [3]. Other algorithms operate after images have been reconstructed, performing de-noising, super-resolution or other post-processing techniques. The proposed algorithm falls into the latter category. Super-resolution in MRI has first been described in [4], in which a reconstruction algorithm is applied on acquisitions with small shifts in the slice selection direction, giving improved resolution and edge definition. In super-resolution MRI reconstruction, the imaging process is generally modelled as follows: A real ground truth object $\mathbf{G}$ is imaged by a process resulting in an image, $\mathbf{X}$. This is modelled by applying a transformation $\mathbf{T}$ and additive gaussian noise n $\mathbf{X} = \mathbf{TG} + n$. The operator $\mathbf{T}$ is defined as a combination of geometric transformations, convolution with a point-spread function, which is often a Gaussian kernel, and downsampling [5]. Having defined an acquisition model, the image reconstruction process can be posed as an ill-conditioned inverse problem. In addition to the data consistency term, different regularisers R have been applied [6,7]. Such regularisers control features of the reconstructions such as the magnitude of edges and degree of smoothness, and allow ill-conditioned problems to be solved as follows:

$$\mathbf{Y} = \min_Y \sum_{i=1}^{N} ||\mathbf{T}_i\mathbf{Y} - \mathbf{x}_i||^2 + \lambda R \qquad (1)$$

Where $\mathbf{Y}$ is the reconstruction, N 2D slices are used, and λ determines the weighting of the regularisation term.

The feasibility of such super-resolution methods was shown by Plenge *et al.* [8], in which they compared iterative back-projection, algebraic reconstruction and regularised least squares algorithms on phantoms and in vivo MRI. In [7,9], a Laplacian regulariser is applied to control the high spatial frequencies on reconstruction of small bird and full body mouse MRI images, giving good qualitative results and outperforming standard interpolation techniques. However, as opposed to sparse sampled cardiac MRI, most of the imaging volume is sampled in those studies.

Total variation is another popular regulariser with edge-preserving and convex characteristics, and has especially been used for de-noising [10]. It has also been applied to super-resolution and compressed sensing of MRI data as a regulariser [11]. An added difficulty in cardiac MRI are the differences in intensity and contrast between SA and LA acquisitions occurring because of different imaging protocols and sequence timings. Thus, any reconstruction algorithm aiming to reconstruct images using information from both the SA and LA images must be contrast independent. Directional total variation (dTV) is a recently introduced approach for reconstructing images using a reference image with the same structure but different contrast. In a study by Ehrhardt *et al.* [12] 2D dTV is used to combine information from brain MRI with different T1 and T2 weightings. It uses one image and its structural information as a reference for the dTV to guide the reconstruction of the other. A different regulariser which has been applied to super-resolution of CMR is the Beltrami regularisation [13] in which they solve Eq. (1) using three sets of image stacks covering the whole left ventricle in the SA, horizonal LA and vertical LA orientations. Limitations of this work are that the slice protocol used does not reflect clinical practice as the number of slices is a lot higher, and it does not address differences in contrast between SA and LA. Recent studies such as work by Oktay *et al.* [14] have focused on the use of Convolutional Neural Networks for super-resolution of CMR and shown great promise. However, methods based on machine learning make the assumption that testing data is well represented by the training data, which may not hold in pathological cases.

In this work, we address the problem of reconstructing 3D images from a stack of 2D slices in both SA and LA orientations, in a contrast-independent manner using the directional total variation regulariser. This allows a reconstructed image with the contrast of the short-axis images but with the additional structural information of the LA images.

2   Materials and Methods

2.1   Image Acquisition

Experimental investigations conformed to the UK Home Office guidance on the Operations of Animals (Scientific Procedures) Act 1986 and were approved by the University of Oxford ethical review board. One heart was excised from a female Sprague-Dawley rat during terminal anaesthesia, fixed then embedded in 1% agarose gel, and imaged on a 9.4 T preclinical MRI scanner (Agilent, CA, USA). A single 3D gradient echo image was acquired: FOV $= 25.6 \times 25.6 \times 25.6$ mm, acquisition matrix $= 384 \times 384 \times 384$, TR $= 200$ ms, TE $= 4$ ms, flip angle $= 60°$, scan time $= 8.2$ h. LR 2D slices $\mathbf{X}$: FOV $= 25.6 \times 25.6$ mm, acquisition matrix $= 128 \times 128$, in-plane resolution $= 0.2$ mm, slice thickness $= 1$ mm were synthetically generated from the 3D image $\mathbf{Y}$ using the sampling function $\mathbf{T}$, such that $\mathbf{X}_i = \mathbf{T}_i \mathbf{Y}$. The sampling function differs to the ones generally used in the literature by working in k-space. Instead of averaging points in image space, the Fourier transform of $\mathbf{Y}$ is truncated in k-space, after rotation of the
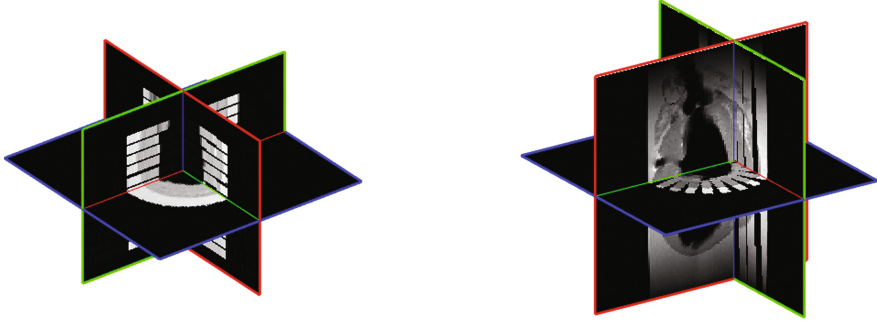
Fig. 1. LHS: cut through 3D view of 11 SA in 3D space. The blue plane is aligned with the SA while the red and green planes are aligned with the LA. RHS: cut through 3D view of 11 LA in 3D space. The LA slices have noticeably different contrast than the SA slices. (Color figure online)

3D image. The image is rotated such that the in-plane view corresponds to the orientation of the slice to be synthetised. The LA images were synthetised after applying a histogram shift to the ground truth 3D volume, to ensure the LR SA, and LA images have different contrast. This is visible in Fig. 1.

2.2   Super-Resolution Algorithm

We formulate the problem by simultaneously solving the following

$$\mathbf{Y}_{\mathrm{SA}}^* = \arg\min_{\mathbf{Y}_{\mathrm{SA}}} \sum_{i=1}^{n_{\mathrm{SA}}} ||\mathbf{T}_i \mathbf{Y}_{\mathrm{SA}} - \mathbf{X}_i||_2^2 + \lambda J(\mathbf{Y}_{\mathrm{SA}}, \mathbf{Y}_{\mathrm{LA}}) \tag{2}$$

$$\mathbf{Y}_{\mathrm{LA}}^* = \arg\min_{\mathbf{Y}_{\mathrm{LA}}} \sum_{j=1}^{n_{\mathrm{LA}}} ||\mathbf{T}_j \mathbf{Y}_{\mathrm{LA}} - \mathbf{X}_j||_2^2 + \lambda J(\mathbf{Y}_{\mathrm{LA}}, \mathbf{Y}_{\mathrm{SA}}) \tag{3}$$

In both (2) and (3), the first term in the problem is related to data accuracy, ensuring that the current estimate does not deviate too much from the 2D LR image $\mathbf{X}_i$ which are the ground truth measurements. The second term sets a constraint using the directional total variation of the image. It pushes the first term in J towards being smooth whilst using the structural information of the second term in J as a reference. λ is a weight adjusting the contribution of the directional total variation term. The 3D directional total variation constraint J applied to image $\mathbf{A}$ with reference image $\mathbf{B}$ is defined as follows [12]:

$$J(\mathbf{A}, \mathbf{B}) = \sum_{n=1}^{3} |D_n \nabla \mathbf{A}_n| \tag{4}$$

where matrix field $D_n \in \mathbb{M}^3 = 1 - \xi_n \xi_n^*$ and $\xi_n := \frac{\nabla \mathbf{B}_n}{|\nabla \mathbf{B}_n|_\eta}$. The tuning parameter η relates to the size of the edges in reference image $\mathbf{B}$. Equations (2) and (3) were simultaneously solved using nonlinear conjugate gradient optimisation [15], in which one step towards the minimum in both Eqs. (2) and (3) was taken during each iteration. The image $\mathbf{Y}_{SA}$ was initialised by putting the LR SA slices into their respective orientations in a 3D matrix with isotropic spacing matching the LR in-plane resolution. The gaps between slices are filled by linear interpolation for a fairer comparison than just using nearest neighbor. $\mathbf{Y}_{LA}$ is initialised in similar fashion, by putting the LR LA slices into a 3D matrix and filling the gaps using nearest neighbour interpolation. Voxels where the LA slices overlap are averaged. Nonlinear conjugate gradient method alternates between iterations as to which image to reconstruct of to use as a reference for the dTV. The λ parameter was empirically set to 0.5 and the η parameter was set to 0.05. Data consistency checks SA LR within the SA reconstruction, and LA LR within the LA reconstruction. The process is repeated until convergence. Results are shown on the SA reconstruction as the aim is to increase their through-plane resolution.

2.3   Validation

The 3D reconstructions will be validated against the HR ground truth image acquired for that purpose, using the Peak-Signal-to-Noise ratio (PSNR) which is widely used in image quality assessment because of its simplicity and clear physical meaning. However, this metric is often criticised for not matching visual quality. In addition, we evaluate the Dice score as well as voxel misclassification for segmentations of the Left Ventricular volume by binarising the images via simple thresholding and give a percentage of misclassified voxels. The contrast between the myocardium and the left ventricle is sufficient that the segmentation result is insensitive to minor changes in the threshold value. The experiment will be run using 11 SA and 11 LA slices covering most of the space, and then with a total number of 12 slices with different combinations of LA and SA acquisitions as to not use more slices than acquired in practice.

3   Results

Table 1 contains qualitative results for reconstructions using 3 different combinations of slices. The first one was chosen to resemble clinical acquisitions with a highly unbalanced number of SA and LA slices. The second one was chosen to balance them by taking an equal number of each orientation, and the third one was done to see if an increased number of LA slices is of benefit.

The result in Fig. 2 shows a cut in LA orientation through the final reconstruction, at an orientation not covered by one of the 11 ground truth LA slices, for a fair comparison. The synthetic slices cover most of the space and do not represent a real clinical scenario. In order for a more realistic approach, we chose

Table 1. Quantitative results: Dice score, Voxel misclassifications and Peak-Signal-to-Noise Ratio in the reconstruction and in the interpolated image.

	9 SA, 3 LA	6 LA, 6 SA	3 SA, 9 LA
Dice score (Reconstruction)	0.9690	0.9762	0.9629
Dice score (Interpolation)	0.9742	0.9710	0.9426
Misclassification (Reconstruction)	4.45%	3.51%	5.37%
Misclassification (Interpolation)	3.70%	4.27%	8.18%
PSNR (Reconstruction)	13.52 dB	14.59 dB	12.69 dB
PSNR (Interpolation)	14.31 dB	13.69 dB	10.08 dB
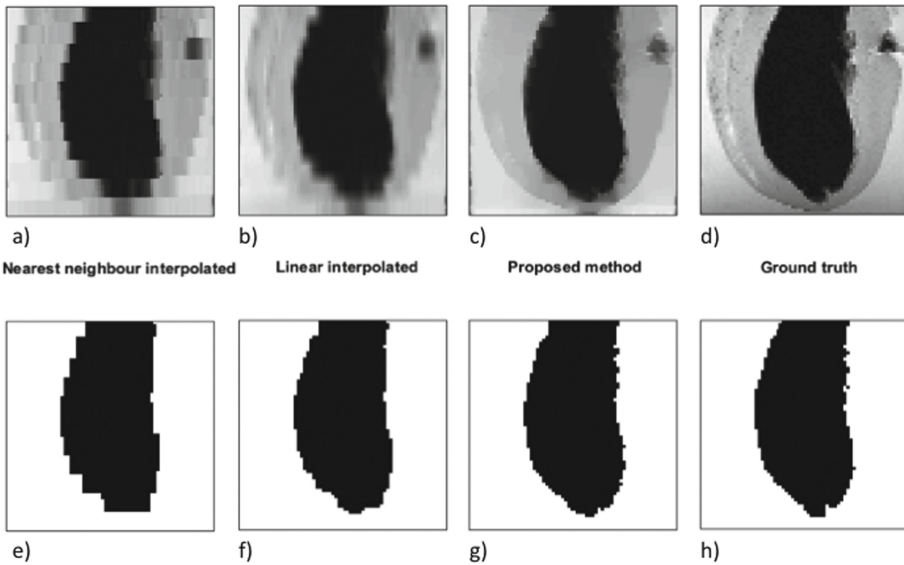
Fig. 2. (a) 11 SA slices in 3D space with nearest neighbour interpolation (b) Initialisation achieved by linear interpolation between 11 SA slices, (c) Reconstruction using the framework aided by 11 SA and 11 LA slices (d) Ground truth (e) Segmentation of nearest neighbour (f) Segmentation of interpolation image, (g) Segmentation of reconstruction (h) Segmentation of ground truth

to use 12 synthetic acquisitions - 6 LA and 6 SA. The more clinically used combination of approximately 10 SA and 2 LA leaves space very under-sampled for through plane detail, especially around the apex and base. Similarly to Fig. 2, the result in Fig. 3 shows a cut in LA orientation through the final reconstruction, at an orientation not covered by one of the 6 ground truth LA slices.
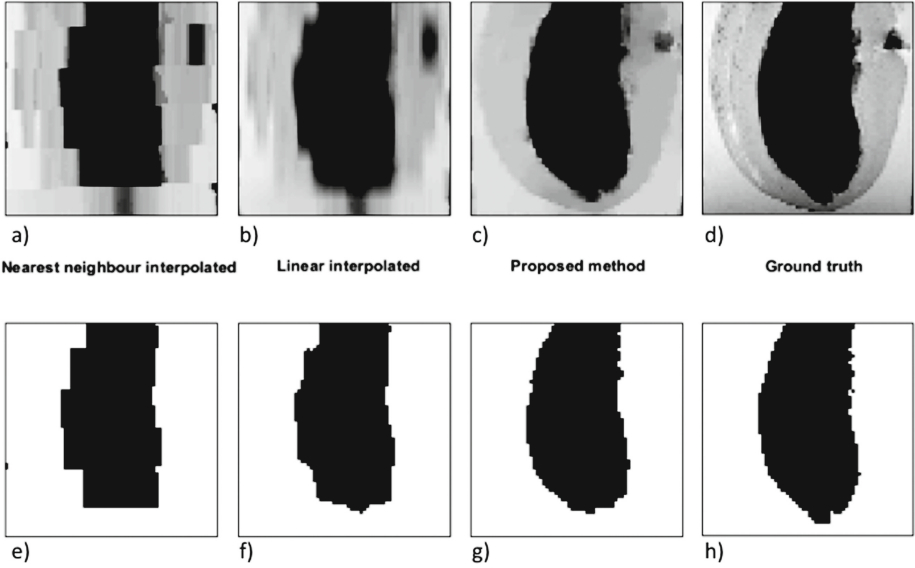
Fig. 3. (a) 6 SA slices in 3D space with nearest neighbour interpolation (b) Initialisation achieved by linear interpolation between 6 SA slices, (c) Reconstruction using the framework aided by 6 SA and 6 LA slices (d) Ground truth (e) Segmentation of nearest neighbour (f) Segmentation of interpolation image, (g) Segmentation of reconstruction, (h) Segmentation of ground truth

4 Discussion and Conclusion

The reconstruction using 9 SA and 3 LA slices does not show improvements with respect to interpolation on any of the metrics that were applied (cf. Table 1), whereas the reconstruction using 3 SA and 9 LA slices does show improvements in the metrics, but starting off with worse quantitative results on the initialisation. This highlights our finding that the slice protocol followed in clinical practice is not ideal for 3D reconstruction, and that increasing the number of LA slices improves the reconstruction. Balancing the number of LA and SA slices shows improvements which outperform the interpolation.

This work has addressed the problem of combining structural information from long-axis images to improve the generation of 3D volumes from short-axis images. Accurate 3D volumes are required for the generation of meshes for mechanical models as well as other applications such as measuring cardiac volumes or estimating ejection fractions.

There are some limitations to this study. The algorithm assumes a pre-processing step of SA-LA registration, and any inaccuracies in this step will be propagated into the image reconstruction. Furthermore, as it is based on total variation, regions outside the sampled planes will typically be as smooth as possible (i.e. the image in-painting is extremely crude). It is therefore crucial that as much of the heart as possible is imaged by at least one plane, which is

not currently done in clinical cardiac MRI. Between the spokes of the LA slices, only the SA is driving the reconstruction and is highly sensitive to initialisation.

Unlike a number of recently proposed methods using convolutional neural networks [14,16], the proposed algorithm does not incorporate any prior information. While these CNN based super-resolution methods have shown excellent performance, the cardiac MRI specific ones assume that short-axis stacks are non-overlapping and parallel [17]. After motion correction, this is rarely the case in clinical acquisitions. It is also unclear how these networks, trained on healthy hearts, will perform on hypertrophic or infarcted hearts. The proposed algorithm does not make any assumptions about the size, orientation, or shape of the heart, or on the slice selection protocol. Thus, it is widely applicable and may be preferable when training data is not available, or when the test data is not well represented by the training data.

Further work will include extending the algorithm to use all frames of cine MRI datasets, rather than operating on a static image. Improved performance is also expected to be achieved by optimising the slice planning, since the slice protocol used in clinical practice is not designed with the aim of 3D reconstruction. At present, standard clinical datasets have too few LA acquisitions, limiting the algorithm's performance.

References

1. Bogaert, J., Dymarkowski, S., Taylor, A.M.: Clinical Cardiac MRI. Taylor & Francis US (2005)
2. Lustig, M., Donoho, D., Pauly, J.M.: Sparse MRI: the application of compressed sensing for rapid MR imaging. Magn. Reson. Med. **58**(6), 1182–1195 (2007)
3. Caballero, J., Price, A.N., Rueckert, D., Hajnal, J.V.: Dictionary learning and time sparsity for dynamic MR data reconstruction. IEEE Trans. Med. Imaging **33**(4), 979–94 (2014)
4. Greenspan, H., Oz, G., Kiryati, N., Peled, S.: MRI inter-slice reconstruction using super-resolution. Magn. Reson. Imaging **20**(5), 437–446 (2002)
5. Van Reeth, E., Tham, I.: Superresolution in magnetic resonance imaging: a review. Concepts Magn. Reson. Part A **40A**(6), 306–325 (2012)
6. Woo, J., Murano, E., Stone, M., Prince, J.: Reconstruction of high resolution tongue volumes from MRI. IEEE Trans. Biomed. Eng. **59**(12), 1–1 (2012)
7. Poot, D.H.J., Meir, V., Sijbers, J.: General and efficient super-resolution method for multi-slice MRI. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) MICCAI 2010. LNCS, vol. 6361, pp. 615–622. Springer, Heidelberg (2010). doi:10. 1007/978-3-642-15705-9_75

8. Plenge, E., Poot, D.H.J., Bernsen, M., Kotek, G., Houston, G., Wielopolski, P., Van Der Weerd, L., Niessen, W.J., Meijering, E.: Super-resolution methods in MRI: can they improve the trade-off between resolution, signal-to-noise ratio, and acquisition time? Magn. Reson. Med. **68**(6), 1983–1993 (2012)
9. Khmelinskii, A., Plenge, E., Kok, P., Dzyubachyk, O., Poot, D.H.J., Suidgeest, E., Botha, C.P., Niessen, W.J., van der Weerd, L., Meijering, E., et al.: Super-resolution reconstruction of whole-body MRI mouse data: an interactive approach. In: 2012 9th IEEE International Symposium on Biomedical Imaging (ISBI), pp. 1723–1726. IEEE (2012)
10. Rudin, L.I., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. Physica D **60**(1–4), 259–268 (1992)
11. Joshi, S.H., Marquina, A., Osher, S.J., Dinov, I., Van Horn, J.D., Toga, A.W.: MRI resolution enhancement using total variation regularization. In: 2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro, ISBI 2009, pp. 161–164. IEEE (2009)
12. Ehrhardt, M.J., Betcke, M.M.: Multicontrast MRI reconstruction with structure-guided total variation. SIAM J. Imaging Sci. **9**(3), 1084–1106 (2016)
13. Odille, F., Bustin, A., Chen, B., Vuissoz, P.-A., Felblinger, J.: Motion-corrected, super-resolution reconstruction for high-resolution 3D cardiac cine MRI. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 435–442. Springer, Cham (2015). doi:10.1007/978-3-319-24574-4_52
14. Oktay, O., et al.: Multi-input cardiac image super-resolution using convolutional neural networks. In: Ourselin, S., Joskowicz, L., Sabuncu, M.R., Unal, G., Wells, W. (eds.) MICCAI 2016. LNCS, vol. 9902, pp. 246–254. Springer, Cham (2016). doi:10.1007/978-3-319-46726-9_29
15. Shewchuk, J.R.: An introduction to the conjugate gradient method without the agonizing pain (1994)
16. Dong, C., Loy, C.C., He, K., Tang, X.: Learning a deep convolutional network for image super-resolution. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) ECCV 2014. LNCS, vol. 8692, pp. 184–199. Springer, Cham (2014). doi:10.1007/978-3-319-10593-2_13
17. Oktay, O., Ferrante, E., Kamnitsas, K., Heinrich, M., Bai, W., Caballero, J., Guerrero, R., Cook, S., de Marvao, A., O'Regan, D., et al.: Anatomically constrained neural networks (ACNN): application to cardiac image enhancement and segmentation. arXiv preprint arXiv:1705.08302 (2017)