Learning cardiac morphology from MR images using a generative adversarial network: a proof of concept study

Davide Piccini[1,2,3], Aurélien Maillot[1,2], John Heerfordt[1,2], Dimitri Van De Ville[4,5], Juerg Schwitter[6], Matthias Stuber[2,7], Jonas Richiardi[1,2], and Tobias Kober[1,2,3]

[1]Advanced Clinical Imaging Technology, Siemens Healthcare, Lausanne, Switzerland, [2]Department of Radiology, Lausanne University Hospital and University of Lausanne, Lausanne, Switzerland, [3]LTS5, École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland, [4]Institute of Bioengineering/Center for Neuroprosthetics, École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland, [5]Department of Radiology and Medical Informatics, University Hospital of Geneva (HUG), Geneva, Switzerland, [6]Division of Cardiology and Cardiac MR Center, University Hospital of Lausanne (CHUV), Lausanne, Switzerland, [7]Center for Biomedical Imaging (CIBM), Lausanne, Switzerland

Synopsis

Learning anatomical characteristics from large databases of radiological data could be leveraged to create realistic representations of a specific subject's anatomy and to provide a personalized clinical assessment by comparison to the acquired data. Here, we extracted 2D patches containing the descending aorta from 297 3D whole-heart MRI acquisitions and trained a Wasserstein generative adversarial network with a gradient penalty term (WGAN-GP). We used the same network to generate realistic versions of the aortic region on masked real images using a loss function that combines a contextual and a perceptual term. Results were qualitatively assessed by an expert reader.

Introduction

Machine learning techniques enable new ways of harvesting information from large patient databases that can be used to extract global features, personalized features of specific subjects' subgroups, or even one single individual. One relevant application consists in learning atlas-like anatomical characteristics from large databases and use this knowledge to provide a personalized "healthy version" of the anatomy of a specific subject. Conceptually, by emphasizing the differences between the expected "learned" subjective anatomy and the real "acquired" anatomy, it is possible to detect morphological anomalies. Here, we test the feasibility of a) learning at least a subset of cardiac anatomical features from a database of cardiac MRI scans, and b) whether we can use this atlas-like knowledge to generate inpainted anatomical details that match the surrounding anatomy.

Methods

Input Data. N=1145 3D whole-heart datasets with high isotropic spatial resolution (0.9 – 1.1mm)[3] were collected between 2015 and 2018 using a prototype respiratory self-navigated sequence[1,2] on a 1.5T clinical MRI scanner (MAGNETOM Aera, Siemens Healthcare, Erlangen, Germany). All datasets were automatically graded for general image quality[3] (Figure 1a) and 297 volumes with grade >3 (out of 4 on a Likert scale) were selected. A semi-automated segmentation algorithm could extract a variable number of 2D patches of 128x128 pixels that contained part of the descending aorta (orange dots, Figure 1b). 80% of the data (237 volumes) were used for training and 20% for testing.

Networks. A generative adversarial network (GAN) architecture[4], consisting of two communicating networks: a critic and a generator, was chosen. While the critic needs to differentiate between real and generated images (e.g. by grading them on a certain distance function), the generator tries to "fool" the critic by producing images that result in critic scores that are similar to those assigned to the training images. Both networks are based on a deep convolutional architecture with one input layer, six hidden layers, and one output layer. The input of the generator was a random 100-dimensional vector ($z \sim \mathrm{N}(0,1)$). The input of the critic consisted of batches with 49 2D images each.

Patch generation. A Wasserstein GAN[5] with a gradient penalty term (WGAN-GP) [6] was used as a loss function (L) as described below:

$$L = \underset{\tilde{x} \sim \mathbb{P}_g}{\mathbb{E}}[D(\tilde{x})] - \underset{x \sim \mathbb{P}_r}{\mathbb{E}}[D(x)]\lambda \cdot \underset{\hat{x} \sim \mathbb{P}_{\hat{x}}}{\mathbb{E}}[(\|\nabla_{\hat{x}} D(\hat{x})\|_2 - 1)^2]$$

$$\hat{x} = t\tilde{x} + (1-t)x, \qquad 0 \leq t \leq 1$$

Where x belongs to the distribution of real images, $\tilde{x} = G(z)$ belongs to the generated images, and $\hat{x}$ is sampled from both distributions. The regularization term penalizes the model if the gradient of the critic moves away from 1. This has been proven to increase the training stability without the need of hyperparameter finetuning.

Inpainting approach. Subsequently, an inpainting framework consisting of three steps was developed around the WGAN-GP. First, a mask (M) of 25x25 pixels is placed onto one of the 2D patches from the test set (y) to exclude the aorta. Second, the same mask is applied to a generated image produced by the WGAN-GP network $(G(z))$. Third, the closest latent space representation of the masked real image $(M \circ y)$ is found by applying gradient descent on the masked synthetic image $(M \circ G(z))$. The loss function for the gradient descent operation is an empirical balance between contextual loss (the masked version of the final image needs to resemble its real counterpart) and perceptual loss (the final image needs to look real based on the critic score)[7].

$$L(z) = L_{contextual}(z) + \lambda \cdot L_{perceptual}(z) = \|M \circ G(z) - M \circ y\| - \lambda \cdot D(G(z))$$

To improve convergence towards a global minimum, the initial guess must be already close to the real masked image. This was achieved by choosing among 20 different initialization vectors the image with lowest initial loss function value.

Data Analysis. At this stage, careful visual inspection of the generated patches compared to those used for training, performed by an expert in cardiac MRI (D.P. 10 years of experience), was considered as a qualitative metric for the feasibility of the patch generation. A qualitative stability analysis was performed for the inpainting part by visually assessing the differences in position and perimeter of the generated portion of the inpainted aorta starting from different initialization.

Results

Figure 2 shows two batches of 2D patches side by side, with the real and generated images on the left and right side, respectively. Overall, the appearance of the generated images corresponds well to realistic cardiac anatomies, confirming that the training process was successful. However, at closer inspection, a residual overall "patchiness" as well as some structured noise pattern show that some improvement is needed. The inpainting approach allowed generating visually realistic versions of the masked aorta in a subset of the test set (Figure 3), when comparing to the original image (a), starting from the patch with the lowest initial cost function value (b), and performing gradient descent (Figure 4). Although the inpainted anatomy is variable (e.g. position and diameter of the aorta) and highly dependent on the initial guess (Figure 5), these preliminary results are encouraging.

Discussion and Conclusion

We demonstrated the feasibility of learning the complexity of the heart from a database of MR images using a GAN. This network can also be potentially used to generate realistic anatomical details using inpainting. A systematic optimization and a quantitative comparison between the real images and generated patches as well as between real anatomy and inpainted anatomy is warranted.

Acknowledgements

No acknowledgement found.

References

1. Piccini D, Monney P, Sierro C, et al. Respiratory self-navigated postcontrast whole-heart coronary MR angiography: initial experience in patients. Radiology. 2014;270(2):378-386.

2. Monney P, Piccini D, Rutz T, et al. Single centre experience of the application of self navigated 3D whole heart cardiovascular magnetic resonance for the assessment of cardiac anatomy in congenital heart disease. J. Cardiovasc. Magn. Reson. 2015;17:55.

3. Demesmaeker R, Heerfordt J, Kober T, et al. Deep Learning for Automated Medical Image Quality Assessment: Proof of Concept in Whole- Heart Magnetic Resonance Imaging. Proceedings of the Joint Annual ISMRM-ESMRMB Meeting. 2018.

4. Radford A, Metz L, and Chintala S. Unsupervised representation learning with deep convolutional generative adversarial networks. ArXiv preprint 2015;1511.06434.

5. Arjovsky M, Chintala S, and Bottou L. Wasserstein gan. ArXiv preprint. 2017;1701.07875.

6. Gulrajani I, Ahmed F, Arjovski M, et al. Improved training of wasserstein gans. Advances in Neural Information Processing Systems. 2017;30:5767–5777.

7. Yeh, RA, Chen C, Lim TY, et. Semantic Image Inpainting with Deep Generative Models. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017.

Figures

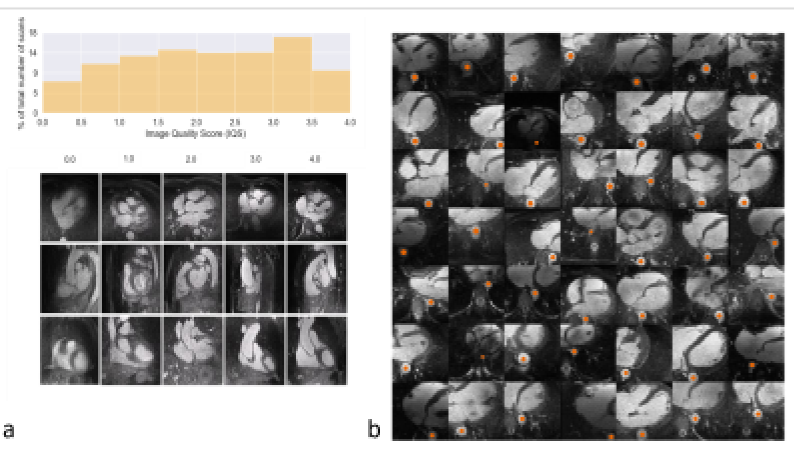

Figure 1: Input data. N=1145 3D whole-heart datasets with high isotropic spatial resolution (0.9 – 1.1mm), collected between 2015 and 2018 using a prototype respiratory self-navigated sequence, were automatically graded for general image quality on a Likert scale from 0 to 4, using a previously published AI-based algorithm (a), 297 volumes with grade >3 were selected for the study and a semi-automated segmentation algorithm was used to extract from all selected 3D volumes a variable number of 2D patches of 128x128 pixels that contained part of the descending aorta (b).

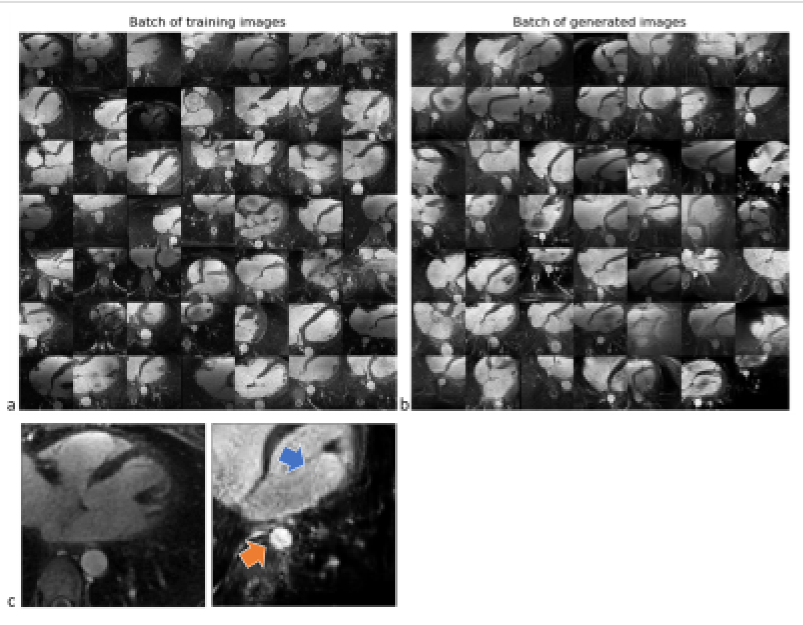

Figure 2: Patch generation. Two batches of 2D images, real images (a) and generated images (b), are shown here side by side. The overall appearance of the generated images corresponds quite well to realistic cardiac anatomies, confirming that the training process of the network was successful. However, at a closer inspection (c), a residual overall patchiness of the generated images (right, orange arrow) as well as a typical structured noise pattern (right, blue arrow) show that improvements is still needed to match the appearance of the real images (left).

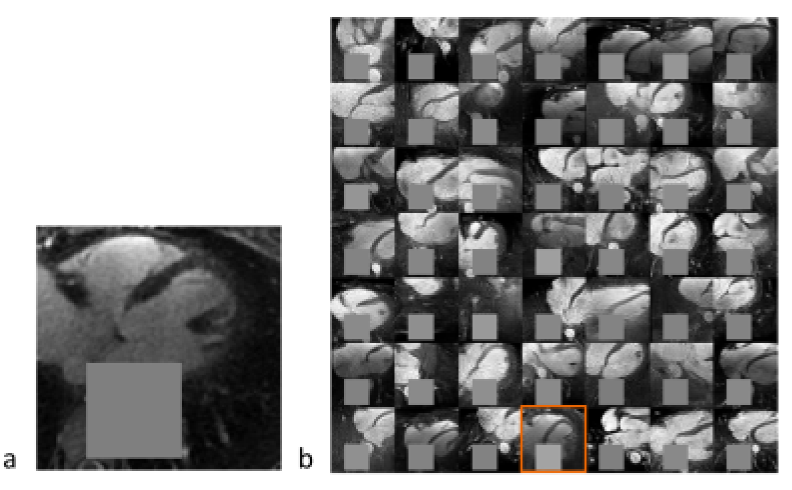

Figure 3: Inpainting initialization. Single batch example in which (a) is a real image multiplied by a binary mask (25x25 pixels) around the aorta, and (b) one of 20 batches of 49 synthetic images multiplied with the same binary mask as the real image. The batch of images is generated from a batch of 49 random, one-hundred-dimensional noise vectors sampled from a normal distribution. The orange framed image is the closest (lowest L1 norm loss) to the real, masked image.

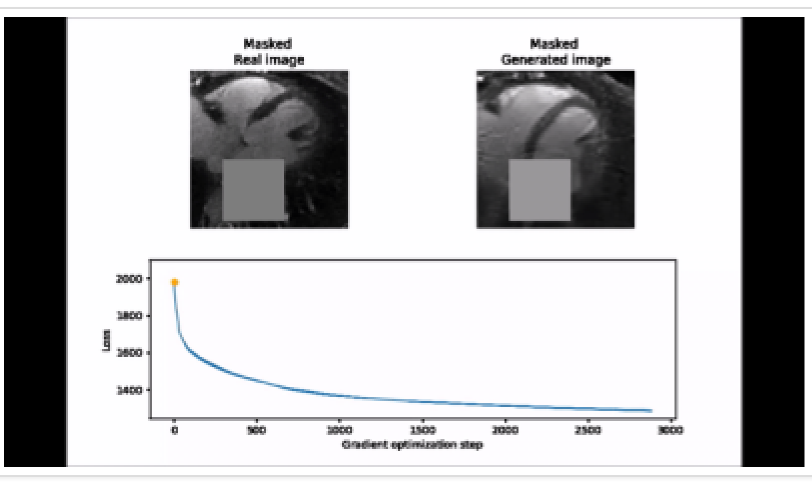

Figure 4: Gradient descent and inpainting. The gif shows how the synthetic image evolution during the gradient descent operation. While the initialization corresponds to the previous figure, in the animation 100 images evenly selected within the 8000 iterations are shown. The final image is visually much closer to the real counterpart.

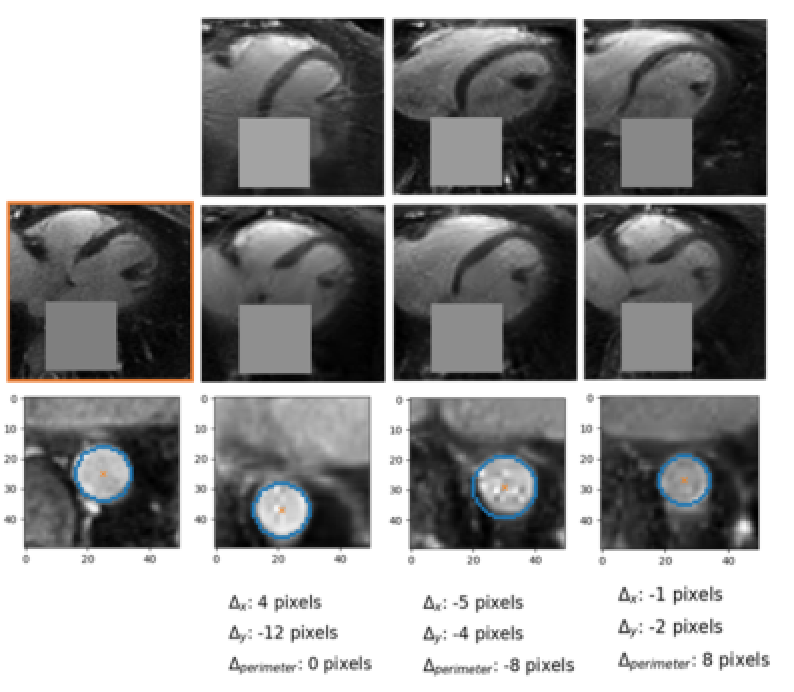

Figure 5: Effect of different initializations on the inpainting algorithm. Top row: 3 different synthetic images used as starting image for inpainting. Middle row: Target real image (framed in orange) and synthetic images after gradient descent optimization. Bottom row: Images within the masked regions showing the different positions and sizes of the aorta (blue) detected automatically using a circular Hough transform.