Conditional Generative Adversarial Networks for the Prediction of Cardiac Contraction from Individual Frames

Julius Ossenberg-Engels[(✉)] and Vicente Grau

Institute of Biomedical Engineering, University of Oxford, Oxford, UK
julius.ossenberg-engels@dtc.ox.ac.uk

Abstract. Cardiac anatomy and function are interrelated in many ways, and these relations can be affected by multiple pathologies. In particular, this applies to ventricular shape and mechanical deformation. We propose a machine learning approach to capture these interactions by using a conditional Generative Adversarial Network (cGAN) to predict cardiac deformation from individual Cardiac Magnetic Resonance (CMR) frames, learning a deterministic mapping between end-diastolic (ED) to end-systolic (ES) CMR short-axis frames. We validate the predicted images by quantifying the difference with real images using mean squared error (MSE) and structural similarity index (SSIM), as well as the Dice coefficient between their respective endo- and epicardial segmentations, obtained with an additional U-Net. We evaluate the ability of the network to learn "healthy" deformations by training it on ~33,500 image pairs from ~12,000 subjects, and testing on a separate test set of ~4,500 image pairs from the UK Biobank study. Mean MSE, SSIM and Dice scores were 0.0026 ± 0.0013, 0.89 ± 0.032 and 0.89 ± 0.059 respectively. We subsequently re-trained the network on specific patient group data, showing that the network is capable of extracting physiologically meaningful differences between patient populations suggesting promising applications on pathological data.

Keywords: cGANs · Image transformation · Cardiac contraction · UK Biobank

1 Introduction

Cardiovascular diseases account for 31% of annual fatalities worldwide [5], making them the most common cause of death. Reasons for this include the poor understanding of many such diseases, which frequently results in the ineffective treatment of patients and sub-optimal clinical outcomes. In particular, the relationship between disease phenotype and clinical outcome is often poorly understood. This is a particular challenge for pathologies such as hypertrophic cardiomyopathy, which present with high phenotype heterogeneity, with a wide range of structural cardiac abnormalities such as myocardial wall thickening

or aortic outflow obstructions. Structural changes such as these can affect the mechanical function of the heart, i.e. cardiac contraction, in a variety of ways. However, the exact relationships are not well understood. Since it is largely cardiac function that determines clinical outcomes, a better understanding of the relationship between cardiac structure and function could lead to more effective risk estimation, diagnosis and treatment.

Whilst structure-function relationships have been studied, this has mostly occurred in isolated approaches typically examining specific global parameters and clinical biomarkers (e.g. cardiac wall thickness/ventricular volumes vs ejection fraction (EF)). Although such biomarkers are clinically established, there is agreement that they are often too coarse for accurate clinical prognosis.

To better understand these relationships and to subsequently improve outcome predictions and treatments, we believe that a more comprehensive approach is necessary which could be achieved using generative methods. We have developed an approach based on conditional Generative Adversarial Networks (cGANs). cGANs, as shown by Isola et al. [6], can be used for image transformation and are therefore suitable for transforming one frame of the cardiac sequence into another, thereby modelling the cardiac sequence and the functional behaviour of the heart. Here, we are primarily interested in the transformation of end-diastolic (ED) to end-systolic (ES) frames, which represent the two extreme states in the cardiac cycle and contain a large part of the functional information.

It is therefore the hypothesis of this paper that by training a cGAN to transform ED to ES frames we can model healthy cardiac motion. Moreover, we hypothesise that our network is capable of capturing cardiac motion that is specific to individual patient groups. In order to validate these hypotheses we conduct two stages of experiments. Firstly, the network is trained on a data set of healthy volunteers representative of a heterogeneous population as given in the UKBiobank study (https://www.ukbiobank.ac.uk/). The network is evaluated based on how accurately the ES frame predictions match the acquired images for a test set using image and segmentation similarity metrics (MSE, SSIM & Dice). Our assumption is that an accurate prediction of ES frames indicates that the cGAN has captured healthy cardiac motion. Secondly, the network is re-trained on data from specific patient groups and tested on separate test sets from the same and different groups. The assumption is that the cGAN will perform sub-optimally when tested on the latter groups, thereby indicating that it has captured features of the cardiac structure-function relationships that are specific to certain patient groups.

2   Related Work

cGANs form an extension to the original GAN network as proposed by Goodfellow et al. [4] in 2014, which functions on the principle of adversarial learning whereby two networks compete in a minimax two-player game. In [10] this principle has been extended by adding conditional inputs to the GAN structure, allowing for outputs to be generated which are conditioned on input labels.

This has been taken further in a method proposed in [2], which allows for progressively increased information capacity of the latent space and more distinct modes of data variation. Isola et al. [6] have shown that cGANs can be adapted for multi-modal image transformation tasks whereby the input condition is the image itself. This paper is primarily based on their method. Further generative methods include Variational Autoencoders (VAEs) [7] and their conditional extension (cVAE) [12]. These, however, have been shown to struggle with image blurriness, due to injected noise in the re-sampling process and the fact that cVAE loss, in contrast to adversarial GAN loss, does not penalise unrealistic images. GANs/cGANs have been used in medical image analysis for a variety of tasks including reconstruction and registration. Specifically in cardiac imaging, GANs have been used for tasks such as image synthesis, for example CMR image synthesis using a Cycle-GAN based on CT scans [3]. Work on image segmentation in cardiology using GANs include quantification of myocardial infarction by Xu et al. [13]. GANs have also been used in identifying cardiac images with incomplete information using SCGANs by Zhang et al. [14]. Generative models for modelling cardiac motion have been proposed by Krebs et al. [8,9] who have investigated probabilistic approaches of modelling cardiac deformation using generative methods including cVAEs.

3   Methods

3.1   Conditional GANs

The cGAN used in this paper is based on the method proposed by Isola et al. [6] Two networks, namely a generator G and a discriminator D, compete in a minimax two-player game, causing them to learn in an adversarial fashion. The generator, a U-Net (Fig. 1), takes a grayscale ED image as an input and generates a "fake" transformed ES image as its output. The discriminator, which is a PatchGAN with the same structure as the encoder part of G with an additional softmax layer at the end, takes both fake and real ES images and learns to discriminate between them in a supervised manner. The discriminator loss given in Eq. 1 is used to train both D and G using mini-batch gradient descent and the Adam optimizer.

$$G = \arg \min_{G} \max_{D} \mathcal{L}_{cGAN}(G, D) + \lambda \mathcal{L}_{L1}(G), \qquad (1)$$

where

$$\mathcal{L}_{cGAN}(G, D) = \mathbb{E}_{x,y}[\log D(x, y)] + \mathbb{E}_{x,z}[\log(1 - D(x, G(x, z)))] \qquad (2)$$

In addition to the $\mathcal{L}_{cGAN}$, which is a conditioned version of the GAN loss, Eq. 1 also incorporates an $L1$ reconstruction loss which forces the transformed image to be close to the ground truth in an $L1$ sense. Hyperparameters and general methodology are based on the Isola method. However, we have adapted the number of layers and filters in G and D following empirical cross-validation.

We validate the performance of the cGAN at accurately predicting ES frames by computing the Mean Squared Error (MSE) and Structural Similarity Index (SSIM) between ES predictions and ground truth images. These metrics are computed over the entire image.
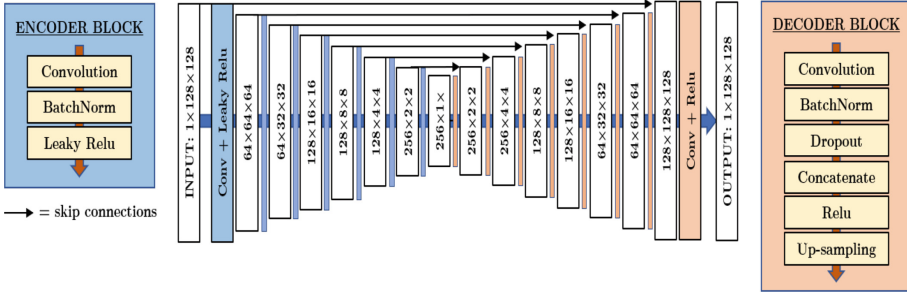
Fig. 1. cGAN U-Net generator architecture; numbers indicate number of filters × image dimensions for each layer. The discriminator architecture is equivalent to the encoding part of the generator, followed by a softmax layer at the end.

3.2   Segmentation U-Net

As seen in the workflow diagram of Fig. 2, following the ED to ES image transformation, the ES prediction and ground truth images are fed into the U-Net segmentation network created by Bai et al. [1], who trained it on data from the UKBiobank to segment short-axis frames into a four-label segmentation output, namely left ventricle (LV) cavity, LV myocardium, right ventricle (RV) and background. ED frames are pre-processed in order to fit the orientation and size of the segmentation network input. The binary LV cavity segmentations from predicted and ground truth ES images are compared using Dice coefficients.

4   Experiments

4.1   Healthy Cardiac Motion Prediction

The data for all experiments was obtained from the UK Biobank study. The initial experiment was conducted using the data of a heterogeneous group of healthy volunteers. Short-axis mid-axial cardiac sequences were selected from ~12,000 patients from which ED and ES frames were extracted. The images were subsequently cropped with a 128 × 128 pixel window centred on the LV endocardium, using as guidance rough LV automated segmentations included in UKBiobank. Images intensities were normalised to [0–1]. The image pairs were split into independent training and test sets, with no inter-subject overlap, with ~33,500 ED-ES image pairs for training and a further ~4,500 for testing. No image augmentation was applied, given the size of the training set. The cGAN was trained on a 5 GB 1060 Nvidia GPU for 200 epochs (~8 h training time).
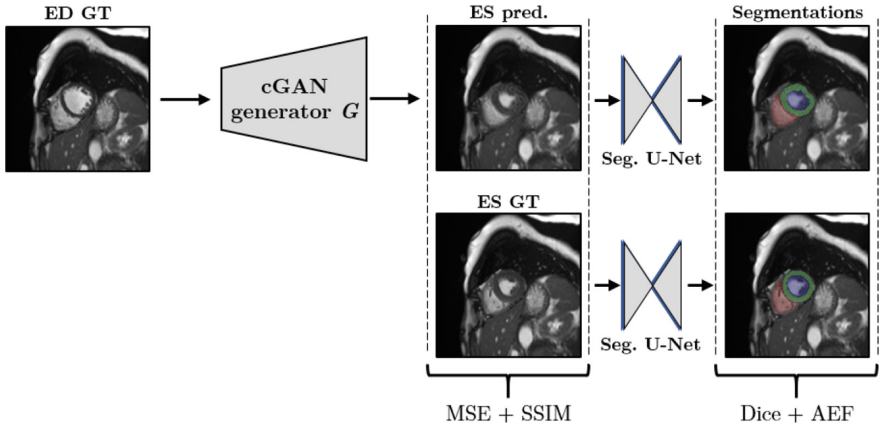
Fig. 2. Workflow diagram: ED ground truth (GT) frames are fed into cGAN generator to produce ES predictions from which MSE and SSIM are computed; predictions and ES GTs are segmented by U-Net from which Dice and AEF are computed.

4.2   Subject Group Re-training

Following the evaluation of the ES frame prediction for a general population of healthy volunteers, the network was re-trained with data from specific subject groups, differentiated by gender, age and Body Mass Index (BMI), also drawn from the UK Biobank study. For gender, the cGAN was separately re-trained on ∼27,500 ED-ES image pairs from female healthy volunteers and tested on separate data sets of female and male healthy volunteers with ∼4,500 image pairs each. For age, the network was re-trained on ∼5,300 ED-ES image pairs from "young" healthy volunteers (ages 40–45) and tested on separate data sets of young and old (ages 65–70) healthy volunteers with ∼700 image pairs each. Lastly, the network was trained on ∼6,000 ED-ES image pairs from volunteers with low BMIs (BMI < 22) and tested on separate sets of subjects with low and high BMIs (BMI > 38) with ∼800 image pairs each.

5   Results

5.1   cGAN ES Frame Predictions

Example results of cGAN ES frame predictions when trained on a general healthy population can be seen in Fig. 4. Mean MSE and SSIM scores between the ES ground truth and ES prediction were 0.0026 ± 0.0013 and 0.89 ± 0.032 respectively. The mean Dice score between the LV myocardial segmentations of ES ground truth and predictions for the general population was 0.89 ± 0.059 (Fig. 3).

5.2   Subject Group Differences

The analysed subject groups were: female/male, young/old and low/high BMI. For each pair the network has been trained on the former group and tested
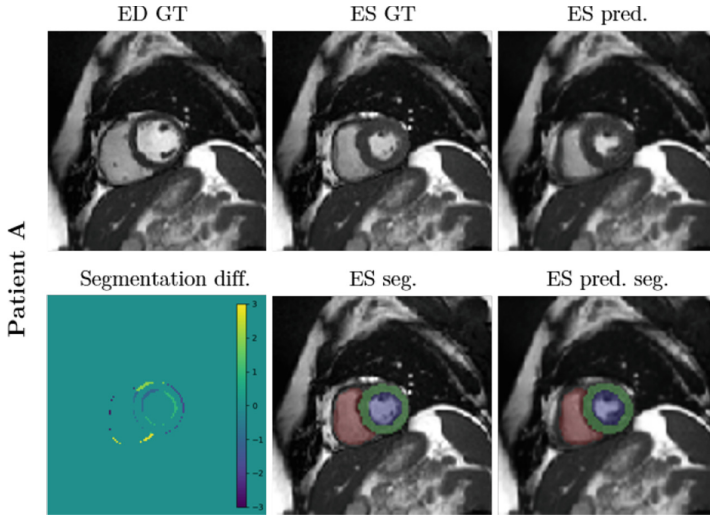
Fig. 3. cGAN prediction of ES frame for Patient A, including segmentations and difference between ES seg and ES pred.seg.; MSE, SSIM and Dice scores of [0.0035, 0.89 & 0.91] respectively.
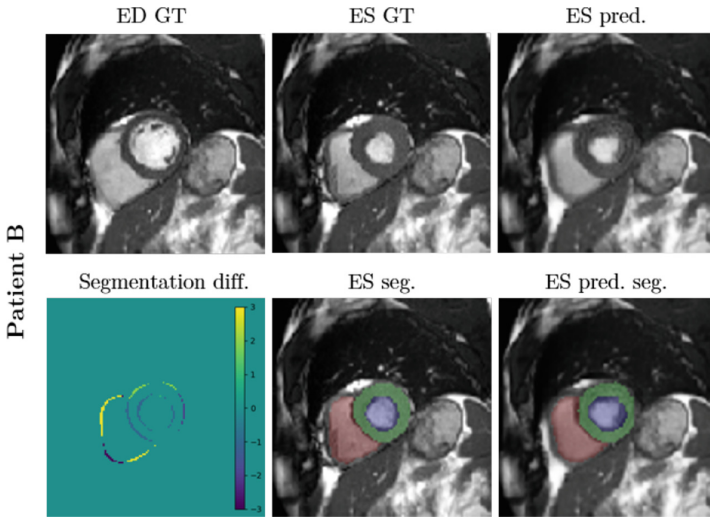

Fig. 4. cGAN prediction of ES frame for Patient B, including segmentations and difference between ES seg and ES pred.seg.; MSE, SSIM and Dice scores of [0.0047, 0.85 & 0.94] respectively.

on separate validation sets from both groups. MSE and SSIM values have been computed using the ES gold standard images and ES predictions for each patient group. Subsequently, all images have been segmented and Dice coefficients of

the segmentations have been computed. Using the LV cavity segmentations, area ejection fractions (AEF) were computed according to $AEF = (EDA - ESA)/EDA$, where EDA is the end-diastolic area and ESA the end-systolic area. The mean percentage differences between the AEF of ground truth images and predictions were computed, as seen in Table 1. The SSIM and Dice scores of the individual patient groups have been plotted on normalised histograms to display frequency distributions, as seen in Fig. 5, which also shows p-values between the distributions calculated using the Kolmogorov-Smirnov test.

Table 1. MSE (in 10^{-3}s), SSIM, Dice scores and AEF differences for patient groups. Note that lower MSE and AEF, and higher SSIM and Dice, indicate better predictions.

Subject group	MSE score	SSIM score	Dice score	AEF diff.
A: Female	2.4 ± 1.2	0.90 ± 0.027	0.88 ± 0.058	14.9 ± 12.7
B: Male	3.2 ± 1.6	0.87 ± 0.038	0.87 ± 0.056	17.1 ± 20.3
A: Young	2.8 ± 1.2	0.88 ± 0.033	0.89 ± 0.048	14.8 ± 12.9
B: Old	3.3 ± 1.6	0.88 ± 0.037	0.87 ± 0.072	18.7 ± 19.0
A: Low BMI	2.7 ± 1.4	0.90 ± 0.032	0.88 ± 0.071	15.1 ± 13.7
B: High BMI	3.8 ± 1.8	0.86 ± 0.039	0.85 ± 0.092	22.6 ± 18.5
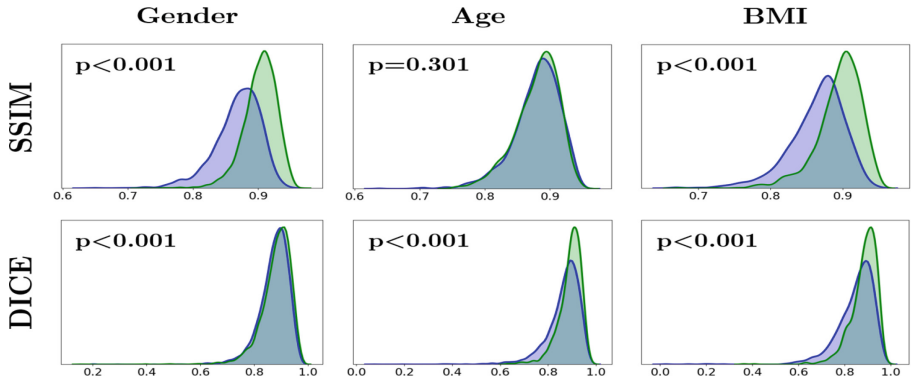
Fig. 5. Subject group histograms; Y-axis = frequency; X-axis = Metric; Group A in green, Group B in blue (Color figure online)

6 Discussion

The prediction of ES frames, shown in Fig. 4, can be evaluated qualitatively and quantitatively. Visual inspection shows that the cGAN is capable of predicting

ES frames with a high degree of accuracy resulting in remarkably realistic predictions. The cGAN can accurately predict LV myocardial wall movements and ES shapes, showing clear delineations of myocardial contours. RV contraction is also captured. It needs to be emphasised, however, that the cGAN does not explicitly capture motion in the sense of predicting motion vector fields, but rather predicts cardiac states as a result of cardiac motion. Quantitative evaluation confirms the quality of cGAN performance with low MSE scores and SSIM scores close to 1. Yet, since these metrics take into account differences across the entire image, local myocardial deformations might be masked by larger feature changes located outside the heart. Hence Dice coefficients based on LV cavity segmentations were used to compare LV shapes. The segmentations showed high agreement between the ground truth and predicted ES frames, which confirms that the cGAN has captured the structure-function interactions in a diverse population of healthy volunteers, thereby validating our first hypothesis.

Nevertheless some limitations can be appreciated, particularly on the RV shape which occasionally lacks crisp delineation of its contours. The edges of the LV myocardium also experience occasional blurriness. The intensities of pixels within the myocardium are sometimes not as homogeneous as in the ground truth. Furthermore, some images display difficulties with capturing papillary muscles or valve movements. Figure 5 and Table 1, together with their p-values, show that when trained on a specific sub-group, the cGAN performs significantly better when tested on separate subjects from that group, as compared to subjects not from that group. Whilst MSE/SSIM/Dice/AEF difference scores are significantly better for trained populations across all tested groups, the level of difference varies between patient group categories. For gender, statistically significant differences can predominately be observed in MSE/SSIM values. This could potentially indicate that image differences other than heart deformation between male and female subjects are responsible for the performance difference. For age groups, the opposite is true. Whilst MSE/SSIM values are more similar, Dice scores differ significantly, indicating significant differences in mechanical function between volunteers of different age groups. In the case of BMI, significant differences can be observed in both Dice and MSE/SSIM scores.

The difference in the metrics clearly suggests that the network has learned what "normal deformation" during the cardiac cycle looks like, because when subjected to a different population, where a physiological difference might be expected [11], the ES predictions are not as accurate. Whilst this performance may be expected, it has to be emphasised that both test groups were drawn from the same study and only varied in the test variable (age, gender and BMI respectively) whilst the distribution of the all other variables was consistent across the groups. Hence a difference in performance could potentially not be present and there requires experimental evaluation. The positive outcome of this experiment validates the hypothesis that cGANs can be used to learn structure-function interactions unique to specific subject groups, and suggests that when using this method for pathological data, we might be able to detect differences due to pathological cardiac function. Furthermore, by analysing the regions with

greatest differences between prediction and ground truth, further anatomical and physiological understanding of structure-function interactions and clinical outcomes might be derived. The technique could potentially be improved by training the cGAN directly on segmented ED-ES image pairs. Advantages of this approach include the focus on the relevant area around the heart and the reduction of ambiguity in fuzzy areas, however, disadvantages are that the segmentations themselves have to be very precise to begin with and that organ texture for example in the form of intensity variations within the myocardium are not taken into account.

References

1. Bai, W., et al.: Automated cardiovascular magnetic resonance image analysis with fully convolutional networks. J. Cardiovasc. Magn. Reson. **20**(1), 65 (2018)
2. Burgess, C.P., et al.: Understanding disentangling in β-vae. arXiv preprint. arXiv:1804.03599 (2018)
3. Chartsias, A., Joyce, T., Dharmakumar, R., Tsaftaris, S.A.: Adversarial image synthesis for unpaired multi-modal cardiac data. In: Tsaftaris, S.A., Gooya, A., Frangi, A.F., Prince, J.L. (eds.) SASHIMI 2017. LNCS, vol. 10557, pp. 3–13. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-68127-6_1
4. Goodfellow, I., et L.: Generative adversarial nets. In: Advances in Neural Information Processing Systems, vol. 27, pp. 2672–2680. Curran Associates, Inc. (2014)
5. Institute for Health Metrics and Evaluation (IHME): Global burden of disease collaborative network. Global Burden of Disease Study 2017 (GBD 2017) Results (2018)
6. Isola, P., Zhu, J.Y., Zhou, T., Efros, A.A.: Image-to-image translation with conditional adversarial networks. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (July 2017)
7. Kingma, D.P., Welling, M.: Auto-encoding variational bayes. arXiv preprint. arXiv:1312.6114 (2013)
8. Krebs, J., Mansi, T., Ayache, N., Delingette, H.: Probabilistic motion modeling from medical image sequences: application to cardiac cine-MRI. arXiv preprint. arXiv:1907.13524 (2019)
9. Krebs, J., Mansi, T., Mailhé, B., Ayache, N., Delingette, H.: Unsupervised probabilistic deformation modeling for robust diffeomorphic registration. In: Stoyanov, D., et al. (eds.) DLMIA/ML-CDS -2018. LNCS, vol. 11045, pp. 101–109. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00889-5_12
10. Mirza, M., Osindero, S.: Conditional generative adversarial nets. arXiv preprint. arXiv:1411.1784 (2014)
11. 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. Magn. Reson. **19**(1), 18 (2017)

12. Sohn, K., Lee, H., Yan, X.: Learning structured output representation using deep conditional generative models. In: Advances in Neural Information Processing Systems, pp. 3483–3491 (2015)
13. Xu, C., Xu, L., Brahm, G., Zhang, H., Li, S.: MuTGAN: simultaneous segmentation and quantification of myocardial infarction without contrast agents via joint adversarial learning. In: Frangi, A.F., Schnabel, J.A., Davatzikos, C., Alberola-López, C., Fichtinger, G. (eds.) MICCAI 2018. LNCS, vol. 11071, pp. 525–534. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00934-2_59
14. Zhang, L., Gooya, A., Frangi, A.F.: Semi-supervised assessment of incomplete LV coverage in cardiac MRI using generative adversarial nets. In: Tsaftaris, S.A., Gooya, A., Frangi, A.F., Prince, J.L. (eds.) SASHIMI 2017. LNCS, vol. 10557, pp. 61–68. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-68127-6_7