Pathology Synthesis of 3D Consistent Cardiac MR Images Using 2D VAEs and GANs

Pathology Synthesis of 3D Consistent Cardiac MR Images Using 2D VAEs and GANs

Sina Amirrajab[1], Cristian Lorenz[2], Juergen Weese[2], Josien Pluim[1], Marcel Breeuwer[1,3]

[1] Biomedical Engineering Department, Eindhoven University of Technology, Eindhoven, The Netherlands
[2] Philips Research Laboratories, Hamburg, Germany
[3] Philips Healthcare, MR R&D - Clinical Science, Best, The Netherlands

Abstract. We propose a method for synthesizing cardiac MR images with plausible heart shapes and realistic appearances for the purpose of generating labeled data for deep-learning (DL) training. It breaks down the image synthesis into label deformation and label-to-image translation tasks. The former is achieved via latent space interpolation in a VAE model, while the latter is accomplished via a conditional GAN model. We devise an approach for label manipulation in the latent space of the trained VAE model, namely pathology synthesis, aiming to synthesize a series of pseudo-pathological synthetic subjects with characteristics of a desired heart disease. Furthermore, we propose to model the relationship between 2D slices in the latent space of the VAE via estimating the correlation coefficient matrix between the latent vectors and utilizing it to correlate elements of randomly drawn samples before decoding to image space. This simple yet effective approach results in generating 3D consistent subjects from 2D slice-by-slice generations. Such an approach could provide a solution to diversify and enrich the available database of cardiac MR images and to pave the way for the development of generalizable DL based image analysis algorithms. The code will be available at https://github.com/sinaamirrajab/CardiacPathologySynthesis.

Keywords: Pathology synthesis, Cardiac MRI, GANs, VAEs, image synthesis.

1 Introduction

Deep generative modeling has gained attention in medical imaging research thanks to its ability to generate highly realistic images that may alleviate medical data scarcity [1]. The most successful family of generative models known as generative adversarial networks (GANs) [2] and Variational Autoencoders (VAEs) [3] are widely used for medical image synthesis and segmentation [4][5]. Many studies have proposed generative models to synthesize realistic and diversified images for brain [6][7] and heart [8][9] among other medical applications [10]. However, the generated data using most mentioned approaches are often unlabeled and therefore not suitable for training a supervised deep learning algorithm, for instance, for medical image segmentation.

1.1 Contributions

We propose to break down the task of cardiac image synthesis into 1) learning the deformation of anatomical content of the ground truth (GT) labels using VAEs and 2) translating GT labels to realistic CMR images using conditional GANs. We devise a strategy, namely **Pathology Synthesis**, to deform labels via interpolation in the latent space of the VAE for the purpose of generating virtual subjects with a target heart disease that affects the heart geometry, e.g. synthesizing a pseudo-pathological subject with thickened myocardium for hypertrophic cardiomyopathy. The synthetic subjects in this study are labeled by design and therefore suitable for medical data augmentation.

Furthermore, we propose a method to generate 3D consistent volumes of synthetic subjects by modeling the correlation between 2D slices in the latent space. The relationship between the slices is captured via estimating the covariance matrix calculated for all latent vectors across all slices. The estimated covariance matrix is used to correlate the elements of a randomly drawn sample. This technique results in a coherent sampling from the latent space and in turn reconstruction of more consistent 3D volume by stacking 2D slices generated from the 2D model.

2 Method

Image synthesis model: The synthesis model architecture includes a ResNet encoder [11] for extracting the style of an input image and a label conditional decoder based on Spatially Adaptive Normalization (SPADE) layer [12]. The model employs SPADE normalization layers throughout the generator architecture to preserve the anatomical content of the GT labels [12][13]. After successful training of the model with pairs of real images and corresponding labels, the generator can translate GT labels to realistic CMR images. To alter the heart anatomy of the synthesized image, we can simply deform the labels. In the previous studies new subjects are synthesized by applying simple transformations such as random elastic deformation, morphological dilation, and erosion on GT labels [13][14]. We utilize the same synthesis network with default training parameters for this study and here we focus on label deformation to generate heart pathology using a VAE model.

Label deformation model: We propose a DL based approach using a VAE model to generate plausible anatomical deformations via latent space manipulation to generate subjects with characteristics of heart pathologies. The VAE model consists of an encoder and a decoder network trained on the ground truth label masks and tries to learn underlying geometrical factors of the heart present in the data. The changes in the heart geometry can be associated with a specific type of disease. For instance, thickening and thinning of the left ventricular myocardium can be an indicating factor of hypertrophic cardiomyopathy and dilated cardiomyopathy, respectively. The goal here is to learn the effects of these factors on the heart geometry presented in the GT labels and to explore the latent space of the VAE to generate new labels with plausibly deformed anatomies. Additionally, we model the characteristics of a particular heart disease in the latent space and generate new samples with heart geometries that represent these disease characteristics.

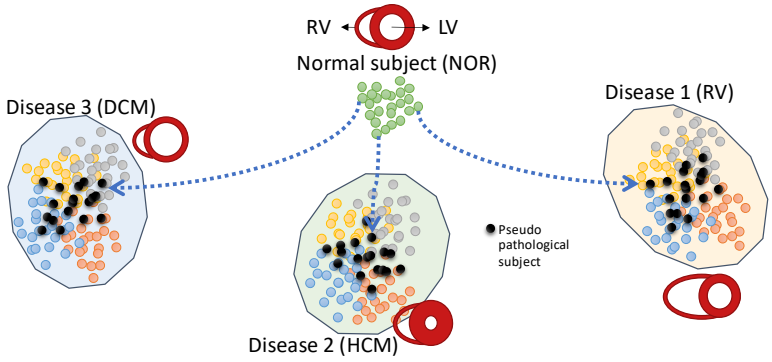

Fig. 1. Strategy to traverse and interpolate in the latent space to perform label deformation using the trained VAE. Each encoded slice of a subject is represented as a dot in the low-dimensional latent space for this schematic view. The statistics of pathological subjects are estimated to draw a sample (pseudo pathological subject) for pathology synthesis. Intermediate latent codes between normal subject (NOR) and the random pseudo-pathological subjects are linearly interpolated (indicated as a dotted blue arrow) to incrementally add pathological features to the heart.

2.1 Pathology Synthesis

Pathology synthesis is designed to generate subjects with informed characteristics of a heart pathology and its effects on the geometry of the heart, given that the pathology is manifested in the ground truth labels. The assumption here is that subjects with a common pathological class have similar heart characteristics and hence they are encoded to the same area in the latent space, as shown in **Fig. 1**.

Suppose we wish to generate subjects with a target pathology, for instance with characteristics of hypertrophic cardiomyopathy (HCM), potentially thickening of the myocardium. Note that we want to preserve the identity of a normal subject (NOR) and only generate disease characteristics such as thickening of the left myocardium for HCM. To this end, assuming that the disease features can be grouped to a neighboring location in the latent space, we encode all subjects with the desired pathology into the latent space and estimate mean, standard deviation, minimum, and maximum across all subjects for all interpolated slices; $(\mu, \sigma, min, max)_{HCM}$. These statistics are calculated on the mean of the posterior distribution which is the output of the encoder. The matrix size for these parameters is $(n_s \times n_z)$, where n_s is the number of interpolated slices (32 in our case) and n_z is the size of the latent vector. Note that we equalize the number of slices for each subject via slice interpolation in the latent space. A sample is drawn from a truncated normal distribution parameterized by these statistics, which we call pseudo-pathology sample; $x_{pHCM} \sim TN(\mu, \sigma, min, max)_{HCM}$. The sample generated with statistics of all HCM subjects should potentially represent heart features of a HCM subject: abnormally thick myocardium. We expect to observe an incremental progression of this anatomical feature on a normal heart by performing linear interpolation between a NOR subject and a pseudo-HCM sample.

To model dependency of variables, the correlation between the dimensions of the latent code for all pathological subjects is measured using Kendall rank correlation coefficient. The uncorrelated generated sample is then transformed in the latent space according to the overall correlation coefficient ($n_z \times n_z$) estimated from the training data to account for relationship between elements of the latent code. The elements of latent vector are correlated using Cholesky matrix decomposition as explained in supplementary material. However, the relationship between different slices of one subject has not yet been modeled. This can lead to generating inconsistent heart geometries in slice direction of one subject as a consequence of slice-by-slice 2D synthesis.

2.2 Modeling slice relationship

We propose a simple statistical modeling to account for the relationship between slices in the latent space. The goal here is to generate consistent 3D volume of labels by stacking 2D reconstructed slices from the decoder part of the VAE model. The 2D VAE model is trained as normal while we attempt to take advantage of the correlation between slices of a given subject in the latent space and reconstruct a consistent 3D volume during the inference. In pathology synthesis, we want to perform a linear interpolation between a NOR subject (x_{NOR}) and the random pseudo-pathological sample (x_{pHCM}) derived from the previous section in the latent space. Although different slices of the NOR subject are inherently correlated in the latent space, the random sample does not contain any information about the relationship between slices. To model this relationship, we estimate the correlation between slices of the x_{NOR} and construct the associated correlation coefficient matrix ($n_s \times n_s$). Given this matrix, we correlate the slices of the x_{pHCM} using the Cholesky matrix decomposition. The procedure is explained in more detail in the supplementary material.

The interaction between latent vectors as well as the relationship between different slices is modeled to generate more realistic correlated samples in the latent space. We found that both latent correlation matrix ($n_z \times n_z$) and slice correlation matrix ($n_s \times n_s$) are important for consistent synthesis. This simple yet effective approach to sampling would better respect the relationship between features presented in the training data and result in generating 3D consistent subjects, despite utilizing 2D models. A similar idea for modeling the distribution of 3D brain MRI data via estimating the correlation in the latent space of a 2D slice VAE has recently been explored in [15].

2.3 Data and implementation

We utilize ACDC challenge data [16] including normal cases (NOR) and three disease classes such as dilated cardiomyopathy (DCM), hypertrophic cardiomyopathy (HCM) and abnormal right ventricle (RV). All 100 ACDC subjects are resampled to 1.5 x 1.5 mm in-plane resolution and cropped to 128 x 128 pixels around the heart using the provided ground truth labels. Percentile based intensity normalization is applied as a post-processing and the intensity range is mapped to the interval of -1 and 1.

The input of the VAE model is a one-hot encoding version of the label map including three channels for heart classes of right ventricle, left ventricle, myocardium, and one

background class. The encoder part of the model includes four convolutional blocks with three convolutional layers each followed by batch normalization (BN) and LeakyReLU activation function. The encoded features are fed to four sequential fully connected layers to output the parameters of a Gaussian distribution over the latent representation. The decoder part of the model is comprised of four convolutional blocks each with one up sampling layer followed by two convolutional layers with BN and LeakyReLU. The last additional block of the decoder includes one convolutional layer followed by BN and another convolution with four channel outputs and Softmax activation function. The VAE model is trained using a combination of cross-entropy loss as the reconstruction loss and Kullback-Leibler divergence (KLD) with weighting factor of β for regularization of the latent space capacity [17]. We experimentally identify the size of the latent vector ($n_z = 16$) and weight of KLD (β=15) by inspecting the quality of the label reconstruction and the outcome of interpolation.

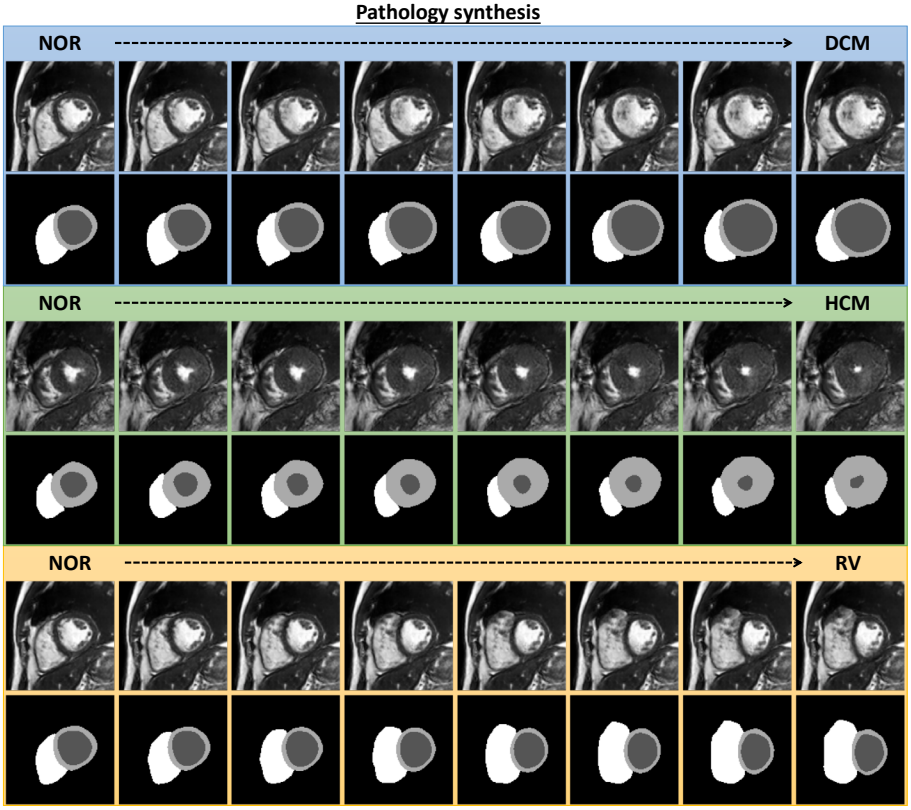

Fig. 2. Pathology synthesis to generate the transition between a normal subject (NOR) to a target pathology such as dilated cardiomyopathy (DCM), hypertrophic cardiomyopathy (HCM) and dilated right ventricle (RV). The effects of a disease on the heart geometry of a subject are respectively left ventricle dilation, myocardial thickening and right ventricle dilation.

3 Results

3.1 Pathology synthesis

The results for pathology synthesis with three target heart diseases namely DCM, HCM, and RV diseases are shown in **Fig. 2**. The characteristics of a particular heart disease are linearly added to the latent code of a normal subject (NOR). The heart shape characteristics of subjects with DCM, dilation of the left ventricle, is progressively appearing on the NOR subject through interpolation from left to right. The same is observed for thickening of the myocardium in the case of NOR to HCM and dilation of the right ventricle for NOR to RV. Note that in pathology synthesis the identity of the NOR subject is not changing while the disease features are manifested on the geometry of the subject's heart and the image appearance stays the same. Interestingly, the detailed structures of the papillary muscles and myocardial trabeculations inside the left and right ventricles are generated despite not being present in the ground truth labels.
We generate synthetic data including five pathological versions of each NOR case, e.g. by interpolating between NOR and HCM subject (synth_HCM). To visualize the anatomical variation of the synthesized data in comparison with the real data, we calculated the ejection fraction (EF) for RV and LV using the ground truth labels ($\mathrm{EF}(\%) = \frac{EDV-ESV}{EDV} * 100$), where EDV and ESV are end-diastolic and end-systolic volumes. As can be seen from **Fig. 3**, there is a considerable overlap between the EF distribution of the synthesized data and the real data.

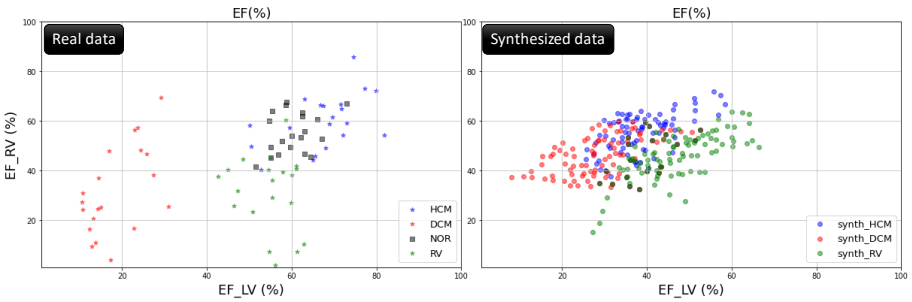

Fig. 3. Distribution of calculated ejection fraction (EF) using the ground truth labels for right-ventricle (EF_RV) and left-ventricle (EF_LV) for the real and synthesized data with pathology.

3.2 Modeling the slice relationship

Our proposed 2D model synthesizes images slice-by-slice with high visual fidelity and realism. However, the synthetic subject that is composed of stacking multiple 2D slices is not generated coherently by the network when we look at the generated slices from perpendicular directions. The reason is that random samples in the latent space contain no information about the relationship between different slices of one subject, i.e. generated slices are uncorrelated. Synthesis examples with target pathologies and the positive effects of the proposed slice correlation on generating 3D consistent subject

is shown in **Fig. 4** with three-dimensional rendering of the synthesized labels. The irregularities in the slice direction are substantially reduced for the correlated slices for synthesizing different pathological cases. We notice that some real images may originally be hampered by slice misalignment artifact and our correlated sampling cannot reduce this artifact.

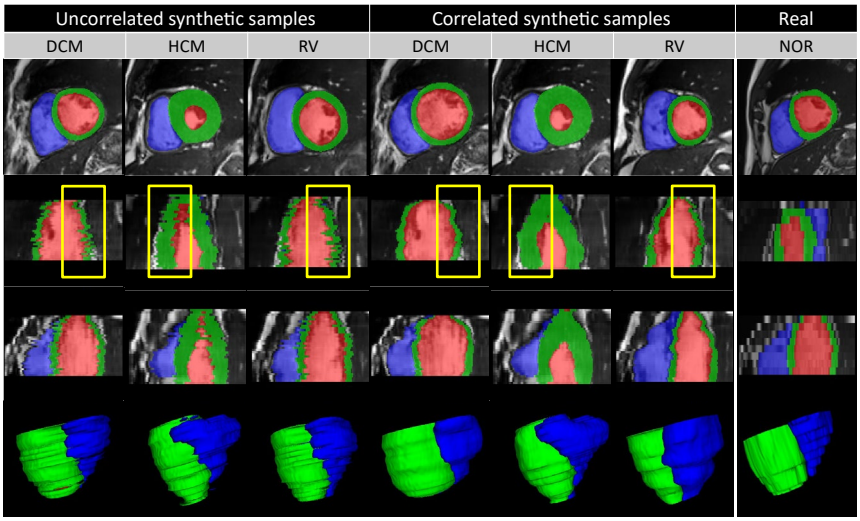

Fig. 4. Three-dimensional rendering of the labels for uncorrelated and correlated synthesis for different cases of pathology synthesis. The first three columns show the uncorrelated slices and its impact on the consistency of the anatomy in the perpendicular views of the short axis slices while the second three columns show the positive effects of correlating samples on reducing the inconsistency and irregularity of the consecutive slices. The last column shows one real example.

4 Discussion and Conclusion

This study investigated an approach for realistic cardiac magnetic resonance image synthesis with target heart pathologies by separating the task into label deformation using a VAE and image generation using a label-conditional GAN. The pathology synthesis was designed to generate subjects with heart characteristics of a particular disease through sampling in the latent space with statistics of a target pathology via performing linear interpolation between a normal subject and pseudo pathological sample in the latent space of the trained VAE.

Furthermore, to tackle one of the important challenges of 3D medical image synthesis, we demonstrated that modeling the correlation between slices in the latent space can be a simple yet effective way to generate consistent 3D subjects from 2D models.

A limitation of our study is the lack of quantitative evaluation of the quality of synthesized images as well as of the 3D consistency of the synthesized subjects. This aspect will be explored in the future work. Visualizations of the synthesized images and the distribution of the heart ejection fractions on the synthesized data nonetheless show

encouraging results. Our approach could provide a solution to diversify and enrich an available database of cardiac MR images and to pave the way for the development of generalizable DL based image analysis algorithms. The methods proposed in this study could be extended for other applications in medical image synthesis such as brain MR image generation and simulation of lesion progression.

Acknowledgments

This research is a part of the OpenGTN project, supported by the European Union in the Marie Curie Innovative Training Networks (ITN) program under project No. 76446

Supplementary Material

Cholesky decomposition and correlated samples

In order to simulate correlated variables with a given covariance matrix (C), Cholesky matrix decomposition is used in this study. The Cholesky matrix decomposition is a factorization of a positive-definite symmetric matrix into a product of a lower and upper triangular matrix, $L\ and\ L^T$, respectively.

$$C = L\,L^T$$

Assuming an uncorrelated random sample X with unit covariance matrix of $\mathbb{E}(XX^T) = I$, a new random vector can be computed as $Y = LX$ that its covariance matrix is derived:

$$\mathbb{E}(YY^T) = \mathbb{E}(LX(LX)^T) = \mathbb{E}(LXX^TL^T) = L\mathbb{E}(XX^T)L^T = LIL^T = LL^T = C$$

Note that the expectation is a linear operator; $\mathbb{E}(cX) = c\mathbb{E}(X)$.

Correlating and generating sample with pathology characteristics

For generating a subject with pathological characteristics, a random sample is drown using a truncated normal distribution parameterized by the statistics of the desired pathology, e.g. mean, standard deviation, minimum, and maximum estimated on all subjects with hypertrophic dilated cardiomyopathy (HCM); namely pseudo-pathological sample x_{pHCM}. These statistics are calculated on the mean of the posterior distribution of features estimated by the encoder part of the VAE. The following steps are followed to correlate the elements of this pseudo-pathological sample cross slice direction and latent dimension:

- Estimate correlation coefficient between latent dimensions across all subjects with desired pathology using Kendall rank correlation coefficient method; $Corr_{zHCM}$ with size $(n_z \times n_z)$ where n_z is the size of the latent vector $(n_z = 16)$
- Calculate the lower triangular matrix L using Cholesky decomposition; L_{zHCM}

- Correlate the latent dimensions of the pseudo pathological sample across the element of latent vector given above formula; $y_{pzHCM} = L_{zHCM} x_{pHCM}$
- Estimate the correlation coefficient between slices of the target normal subject (NOR) we wish to use for interpolation; $Corr_{sNOR}$ with size $(n_s \times n_s)$ where n_s is the number of slices $(n_s = 32)$
- Calculate the lower triangular matrix L using Cholesky decomposition; L_{sNOR}
- Correlate the latent dimensions of the pseudo random sample cross slices given above formula; $z_{pzsHCM} = L_{sNOR} y_{pzHCM}$
- Linearly interpolate between z_{NOR} and z_{pzsHCM} in the latent space
- Reconstruct slices-by-slice the interpolated samples using the decoder part of the 2D VAE
- Compose 3D volume from synthesized 2D slices

The correlation coefficient matrix for all above mentioned steps is shown in **Fig. 5**. Correlating latent dimensions found to be as important as correlating slices of subject for generating coherent slices with smoothly changing features.

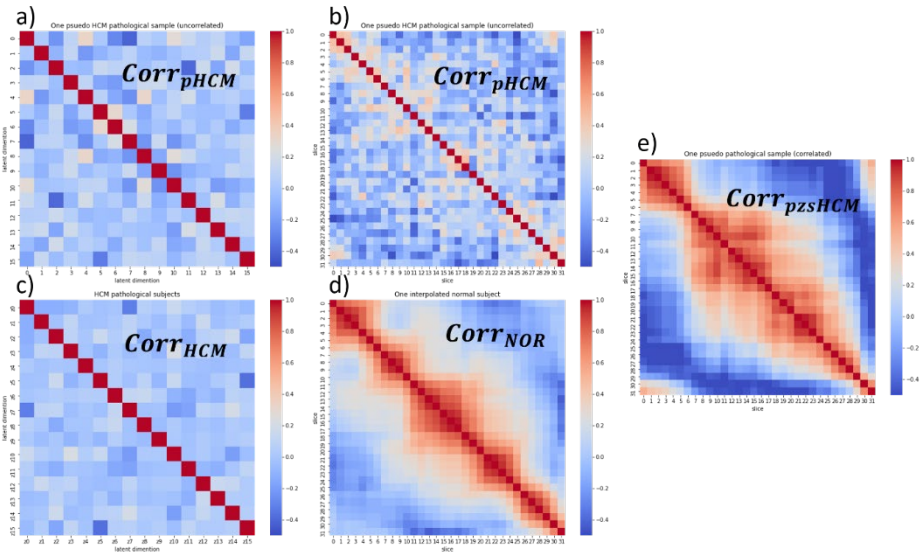

Fig. 5. Correlation coefficient matrix for a) uncorrelated pseudo-HCM sample across latent dimensions and b) across slices, c) all HCM subjects across latent dimensions, d) one normal subject across slices, and e) the correlated pseudo pathological sample calculated using the Cholesky decomposition.

References

[1] X. Yi, E. Walia, and P. Babyn, "Generative adversarial network in medical imaging: A review," *Med. Image Anal.*, vol. 58, p. 101552, 2019.

[2] I. J. Goodfellow *et al.*, "Generative Adversarial Networks," *Commun. ACM*, vol. 63, no. 11, pp. 139–144, 2014.

[3] D. P. Kingma and M. Welling, “Auto-encoding variational bayes,” in *2nd International Conference on Learning Representations, ICLR 2014 - Conference Track Proceedings*.

[4] T. Joyce and S. Kozerke, “3D medical image synthesis by factorised representation and deformable model learning,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 11827 LNCS, pp. 110–119, 2019.

[5] S. Kazeminia *et al.*, “GANs for medical image analysis,” *Artif. Intell. Med.*, vol. 109, p. 101938, 2020.

[6] G. Kwon, C. Han, and D. shik Kim, “Generation of 3D Brain MRI Using Auto-Encoding Generative Adversarial Networks,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 11766 LNCS, pp. 118–126, 2019.

[7] S. U. H. Dar, M. Yurt, L. Karacan, A. Erdem, E. Erdem, and T. Cukur, “Image Synthesis in Multi-Contrast MRI with Conditional Generative Adversarial Networks,” *IEEE Trans. Med. Imaging*, vol. 38, no. 10, pp. 2375–2388, 2019.

[8] A. Chartsias, T. Joyce, R. Dharmakumar, and S. A. Tsaftaris, “Adversarial Image Synthesis for Unpaired Multi-modal Cardiac Data,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 10557 LNCS, pp. 3–13, 2017.

[9] M. Rezaei, “Generative adversarial network for cardiovascular imaging,” *Mach. Learn. Cardiovasc. Med.*, pp. 95–121, 2021.

[10] K. Armanious *et al.*, “MedGAN: Medical Image Translation using GANs,” *Comput. Med. Imaging Graph.*, vol. 79, 2018.

[11] K. He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition,” in *In: Proc. IEEE Comput. Soc. Conf. Comput. Vis. Patt. Recognit.,* pp. 770–778, 2016.

[12] T. Park, M. Y. Liu, T. C. Wang, and J. Y. Zhu, “Semantic image synthesis with spatially-adaptive normalization,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit*., pp. 2332–2341, 2019.

[13] D. R. P. R. M. Lustermans, S. Amirrajab, M. Veta, M. Breeuwer, and C. M. Scannell, “Optimized Automated Cardiac MR Scar Quantification with GAN-Based Data Augmentation,” *Submitt. peer Rev.*, 2021.

[14] Y. Al Khalil, S. Amirrajab, J. Pluim, and M. Breeuwer, “Late Fusion U-Net with GAN-Based Augmentation for Generalizable Cardiac MRI Segmentation,” in *In International Workshop on Statistical Atlases and Computational Models of the Heart. Springer, Cham.*, pp. 360–373, 2021

[15] A. Volokitin *et al.*, “Modelling the Distribution of 3D Brain MRI Using a 2D Slice VAE,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 12267 LNCS, pp. 657–666, 2020.

[16] O. Bernard *et al.*, “Deep Learning Techniques for Automatic MRI Cardiac Multi-Structures Segmentation and Diagnosis: Is the Problem Solved?,” *IEEE Trans. Med. Imaging*, vol. 37, no. 11, pp. 2514–2525, 2018.

[17] I. Higgins *et al.*, “beta-VAE: Learning Basic Visual Concepts with a Constrained Variational Framework,” in *ICLR conference*, 2016.