

Received 10 August 2022, accepted 10 September 2022, date of publication 16 September 2022,  
date of current version 23 September 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3207177

## RESEARCH ARTICLE

# A Data Augmentation Pipeline to Generate Synthetic Labeled Datasets of 3D Echocardiography Images Using a GAN

CRISTIANA TIAGO<sup>1,2</sup>, (Member, IEEE), ANDREW GILBERT<sup>2</sup>, AHMED SALEM BEELA<sup>3</sup>,  
SVEIN ARNE AASE<sup>2</sup>, STEN ROAR SNARE<sup>2</sup>, JURICA ŠPREM<sup>2</sup>, AND KRISTIN MCLEOD<sup>1,2</sup>

<sup>1</sup>Department of Informatics, University of Oslo, 0373 Oslo, Norway

<sup>2</sup>GE Vingmed Ultrasound, GE Healthcare, 3183 Horten, Norway

<sup>3</sup>CARIM School for Cardiovascular Diseases, Maastricht University Medical Center, 6229 ER Maastricht, The Netherlands

Corresponding author: Cristiana Tiago (cristiana.tiago@ge.com)

This work was supported by the European Union's Horizon 2020 Research and Innovation Programme through the Marie-Sklodowska Curie Grant under Agreement 860745.

**ABSTRACT** Due to privacy issues and limited amount of publicly available labeled datasets in the domain of medical imaging, we propose an image generation pipeline to synthesize 3D echocardiographic images with corresponding ground truth labels, to alleviate the need for data collection and for laborious and error-prone human labeling of images for subsequent Deep Learning (DL) tasks. The proposed method utilizes detailed anatomical segmentations of the heart as ground truth label sources. This initial dataset is combined with a second dataset made up of real 3D echocardiographic images to train a Generative Adversarial Network (GAN) to synthesize realistic 3D cardiovascular Ultrasound images paired with ground truth labels. To generate the synthetic 3D dataset, the trained GAN uses high resolution anatomical models from Computed Tomography (CT) as input. A qualitative analysis of the synthesized images showed that the main structures of the heart are well delineated and closely follow the labels obtained from the anatomical models. To assess the usability of these synthetic images for DL tasks, segmentation algorithms were trained to delineate the left ventricle, left atrium, and myocardium. A quantitative analysis of the 3D segmentations given by the models trained with the synthetic images indicated the potential use of this GAN approach to generate 3D synthetic data, use the data to train DL models for different clinical tasks, and therefore tackle the problem of scarcity of 3D labeled echocardiography datasets.

**INDEX TERMS** 3-D image generation, data augmentation, deep learning, echocardiography, generative adversarial networks, segmentation.

## I. INTRODUCTION

Medical imaging plays a crucial role in optimizing treatment pathways. Saving time when it comes to diagnosis and treatment planning enables the clinicians to focus on more complicated cases.

Many modalities are used to image the heart, such as Computed Tomography (CT), Magnetic Resonance (MR), and Ultrasound imaging, enabling several structural and functional parameters related to the organ's performance to be

The associate editor coordinating the review of this manuscript and approving it for publication was Janmenjoy Nayak<sup>1</sup>.

estimated. Such parameters are the basis of clinical guidelines for diagnosis and treatment planning.

Echocardiography is the specific use of ultrasound to image the heart. This imaging modality is widely used given its advantages of being portable, relatively low-cost, and the fact that the use of ionizing radiation is not required.

Deep Learning (DL), and specifically Convolutional Neural Networks (CNNs), have become extensively applied in medical image analysis because they facilitate the automation of many tedious clinical tasks and workflows such as estimation of ejection fraction, for example. These algorithms are capable of approaching human-level performance [1],

thus potentially saving clinicians' time without decreasing the quality of care for patients. In fact, clinicians agree that using DL algorithms in the clinical workflow also improves patient access to disease diagnoses, increasing the final diagnosis confidence levels [2]. DL models can be developed to perform numerous medical tasks such as image classification, segmentation and even region/structure detection [3].

Echocardiography images can be acquired both in 2D and 3D. Time can also be taken into account, generating videos. 3D echocardiography images can be more difficult to assess than 2D images. However, for some specific application cases, 3D image acquisition brings great advantages since it can offer more accurate and reproducible measurements. One such case is ventricle and atrium volumes [4]. Amongst the causes of lack of annotated 3D echocardiography datasets are the higher complexity to acquire 3D echocardiography images and the fact that 3D is still not part of all echocardiography routine exams. Also, even when 3D images are recorded, delineating the structures in them is a challenging, time consuming, and user dependent task. Taken together and adding the fact that privacy regulations to access medical data are becoming stricter, these can explain why there is a lack of publicly available datasets of such type of images. Therefore, having an approach able to address this image scarcity is necessary. This current lack of 3D medical data and the great need of high quality annotated data required by the DL models impacts the development of such algorithms and therefore the scientific and technological development of the 3D medical imaging field. Synthetic generation of labeled 3D echocardiography images is a DL based approach that provides a solution for this problem.

Synthetic data can help in the development of DL models for image analysis [5] and accurate labeling of these images. Furthermore, this approach works as a data augmentation strategy by generating additional data. It is known that creating datasets with a combination of real and synthetic images and use them to train algorithms that tackle medical challenges represents a successful solution to the image scarcity [6] problem. Such type of synthetic images even increase the heterogeneity present on these datasets, facilitating a more efficient performance of the trained models as they are exposed to a larger variety of images.

Generative Adversarial Networks (GANs) are specific DL architectures that create models capable of generating medical images closely resembling real images acquired from patients. These deep generative models rely on a generator and a discriminator. While the straightforward GAN discriminator distinguishes between real and fake, i.e., generated, images, the generator not only attempts to deceive the discriminator but also tries to minimize the difference between the generated image and the ground truth.

The generated synthetic images can even be associated with labels facilitating the acquisition of large labeled datasets, eliminating the need for manual annotation, and therefore the variabilities associated with the observer [7], which largely influences the final output [8]. 3D heart models

are a great source of anatomical labels since they capture accurate information about the organ's structures [9]. Different types of models can be used for this purpose, such as animated models, biophysical models, or even anatomical models obtained from different imaging modalities [10], [11]. Recently, CT models were used as label sources to generate 2D echocardiography [12] and cardiac MR images [13], proving the utility of GANs for this task.

Developing a pipeline to generate synthetic data using GANs to create labeled datasets addresses the immense need for the large volume of data that DL algorithms require during training to perform an image analysis task, eliminates the need to acquire the images from subjects, and saves time of experienced professionals when annotating them, as the anatomical labels can be extracted from anatomical models. Usually, when developing such generative models, imaging artifacts are present and visible on the synthetically generated images. This widely common GAN performance drawback is addressed by applying some image post-processing operations [14] on the synthetically generated 3D echocardiography images.

In practice synthetic images can be used to train DL models because they represent a good data augmentation strategy [15]. For instance, 3D medical image segmentation is the most common example of a medical task to which DL can turn out to be a good application. Labeled datasets made of real images combined with synthetic ones, which even include the respective anatomical labels, become the training dataset for 3D DL models, addressing the problem of sparse 3D medical data availability [16].

## A. STATE OF THE ART

DL has become widely used in medical imaging due to its potential in image segmentation, classification, reconstruction, and synthesis across all imaging modalities. Image synthesis has been a research topic for a few decades now, where some of the more conventional approaches use human-defined rules and assumptions like shape priors, for example [17]. Also, these image synthesis techniques depend on the imaging modality being considered to perform certain tasks. To tackle these shortcomings, CNNs are now becoming a widely used approach for image synthesis across many medical imaging modalities.

Many reasons motivate medical image generation, both 2D and 3D. Generative algorithms can perform domain translation, with a large applicability when converting images from one imaging modality to a different one, as Uzunova *et al.* [18] showed in their work converting 3D MR and CT brain images. GANs can also be used to generate a ground truth for a given input, as these DL models can be trained in a cyclic way, as is the case of the CycleGAN [19], for example. Additionally, generation of synthetic data used for DL algorithms also motivates the application and development of GAN architectures. Several research groups were able to generate medical images using this methodology as a data augmentation tool, even though most of them were developed

under a 2D scenario and focused on a few imaging modalities, mainly MRI and CT. These imaging modalities raise less challenges when compared with Ultrasound due to the nature of the physics behind the acquisition process.

Ultrasound images have an inherent and characteristic speckle pattern and their quality is largely influenced by the scanner, the sonographer, and the patient anatomy. When it comes to generating 3D Ultrasound images a few more challenges arise, with the speckle pattern having to be consistent throughout the whole volume being the main one. The anatomical information present in the generated volume also has to hold this consistency feature.

Huo *et al.* [20] trained a 2D GAN model, SynSegNet, on CT images and unpaired MR labels using a CycleGAN. Similarly, Gilbert *et al.* [12] proposed an approach to synthesize labeled 2D echocardiography images, using anatomical models and a CycleGAN as well. The CycleGAN was proposed by Zhu *et al.* [19] and works under an unpaired scenario: the images from one training domain do not have to be related with the images belonging to the other domain. This GAN learns how to map the images from one to another and vice-versa. The paired version of this GAN is called Pix2pix. Isola *et al.* [21] proposed this image synthesis method which generates images from one domain to the other, and vice-versa, however the images belonging to the training domains are paired.

As mentioned, 3D echocardiographic data is sparser, but these images can be generated using GANs, and then used to train new algorithms. Both Gilbert *et al.* [12] and Amirrajab *et al.* [22] investigated the potential use of GAN synthesized datasets to train CNNs to segment different cardiac structures on different imaging modalities, but these methods were limited to 2D.

Hu *et al.* [23] attempted to generate 2D fetal Ultrasound scan images at certain 3D spatial locations. They concluded that common GAN training problems such as mode collapse occur. Abbasi-Sureshjani *et al.* [24] developed a method to generate 3D labeled Cardiac MR images relying on CT anatomical models to obtain labels for the synthesized images, using a SPADE GAN [25]. More recently, Cirillo *et al.* [26] adapted the original Pix2pix model to generate 3D brain tumor segmentations.

When dealing with medical images, U-Net [27] is a widely used CNN model to perform image segmentation, for example, since it provides accurate delineation of several structures on these images. More recently, Isensee *et al.* [28] proposed nnU-Net (“no new net”), which automatically adapts to any new datasets and enables accurate segmentations. nnU-Net can be trained on a 3D scenario and optimizes its performance to new unseen datasets and different segmentation tasks, requiring no human intervention.

Existing work to address the challenges of automatic image recognition, segmentation, and tracking in echocardiography has been mostly focused on 2D imaging. In particular, recent work indicates the potential for applying DL approaches to accurately perform measurements in echocardiography

images. Alsharqi *et al.* [29] and Østvik *et al.* [30] used a DL algorithm to segment the myocardium in 2D echocardiographic images, from which the regional motion, and from this the strain, were measured. They showed that motion estimation using CNNs is applicable to echocardiography, even when the networks are trained with synthetic data. This work supports the hypothesis that similar approaches could also work for 3D synthetic data.

A large amount of work has been carried out on medical imaging generation and it still represents a challenge for the research community. To the best of our knowledge, the challenge of synthesizing 3D echocardiography images using GANs did not produce any reproducible results, therefore we propose a framework able to address this need.

## B. CONTRIBUTIONS

We propose an approach for synthesizing 3D echocardiography images paired with corresponding anatomical labels suitable as input for training DL image analysis tasks. Thus, the main contributions of the proposed pipeline beyond the state of the art include:

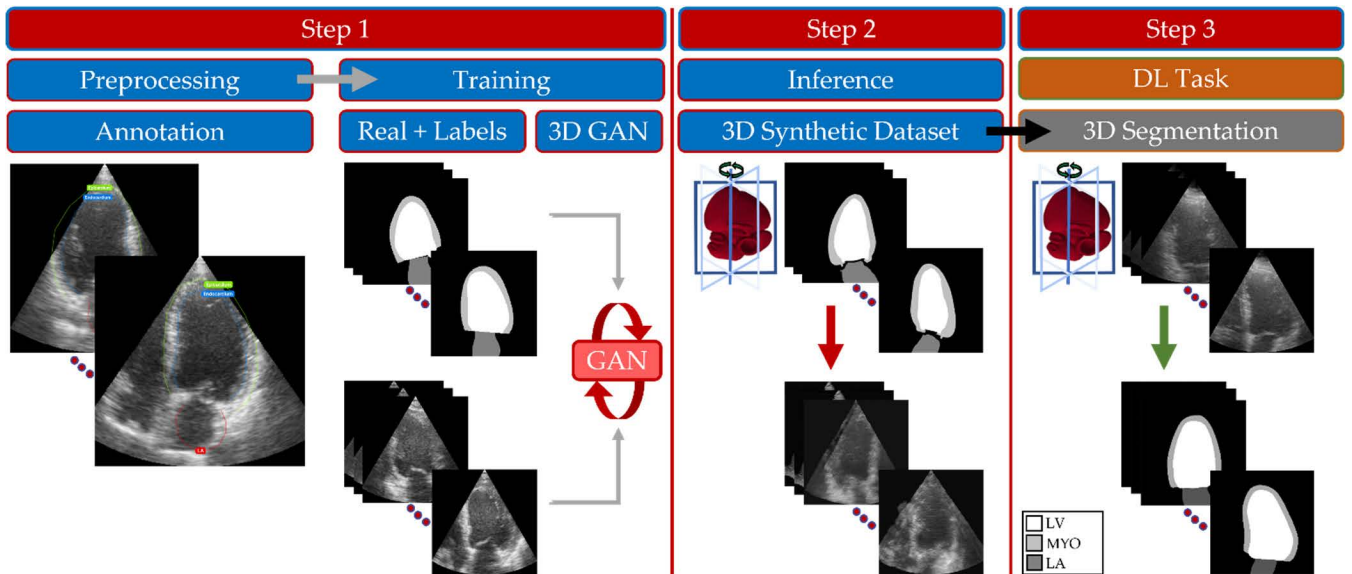
- 1) The extension of Gilbert *et al.* [12] work from 2D to 3D, adapting it from an unpaired to a paired framework (3D Pix2pix) and proposing an automatic pipeline to generate any number of 3D echocardiography images, tackling the lack of public 3D echocardiography datasets and corresponding labels.
- 2) The creation of a blueprint of heart models and post-processing methods for optimal generation of 3D synthetic data, creating a generic data augmentation tool, this way addressing the lack of 3D data generation works in echocardiography, since it significantly varies from 2D.
- 3) The demonstration of the usability of these synthetic datasets for training segmentation models that achieve high performance when applied to real images.

## II. METHODOLOGY

The proposed pipeline is summarized in Fig. 1 and described in the following sections. Section II-A describes the pre-processing stage of annotation of the GAN training images to create anatomical labels for these. The training and inference stages are addressed in Section II-B describing how the GAN model was trained and used to synthesize 3D echocardiography images from CT-based anatomical labels and how different post-processing approaches, as described in Section II-C, were applied to these synthetic images. Next, on Section II-D, details regarding the creation of several synthetic datasets used to train 3D segmentation models are given, followed by Section II-E where the influence of adding real images to the synthetic datasets to train segmentation models is assessed.

### A. DATA COLLECTION

To train the 3D image synthetization model, an annotated dataset was needed since this GAN set up works under a supervised scenario where two sets of images are used for training: a set containing real 3D echocardiography images



**FIGURE 1.** 3D echocardiography image generation pipeline and inference results. Step 1: during the preprocessing stage, a set of 15 3D heart volumes were labeled by a cardiologist and anatomical labels for the LV, LA and MYO were generated. To train the 3D Pix2pix GAN model, the anatomical labels are paired together with the corresponding real 3D images. Step 2: at inference time, the GAN model generates one 3D image. An example obtained during this stage is shown. The proposed method is able to generate physiologically realistic images, giving correct structural features and image details. Step 3: to show the utility of the synthetic datasets, 3D segmentation models were trained using these GAN generated images (black arrow), but other DL tasks can be addressed.

and a second set with the correspondent anatomical labels manually performed by a cardiologist (see Fig.1, training stage).

To create the dataset of real 3D echocardiography images, these were acquired during one time point of the cardiac cycle of normal subjects, end-diastole in this work, when the left ventricle (LV) volume is at its largest value. Using GE Vivid Ultrasound scanners 15 heart volumes were acquired.

The second set of images was made up of the anatomical labels corresponding to each of the 3D real images included in the set previously described. Each anatomical label image contains the label for the LV, left atrium (LA), and the myocardium (MYO).

To annotate the 3D echocardiography images a certified member of the American National Board of Echocardiography cardiologist, with more than 10 years of experience, used the V7 annotation tool [31] and contoured the three aforementioned structures (Fig. 1, preprocessing stage) on each of the volumes. These contours were then post-processed, applying a spline function to the contour points and resampling it, in order to generate gray scale labeled images. All the 3D images present on each training dataset were sized to  $256 \times 256 \times 32$ .

### B. 3D GAN TRAINING

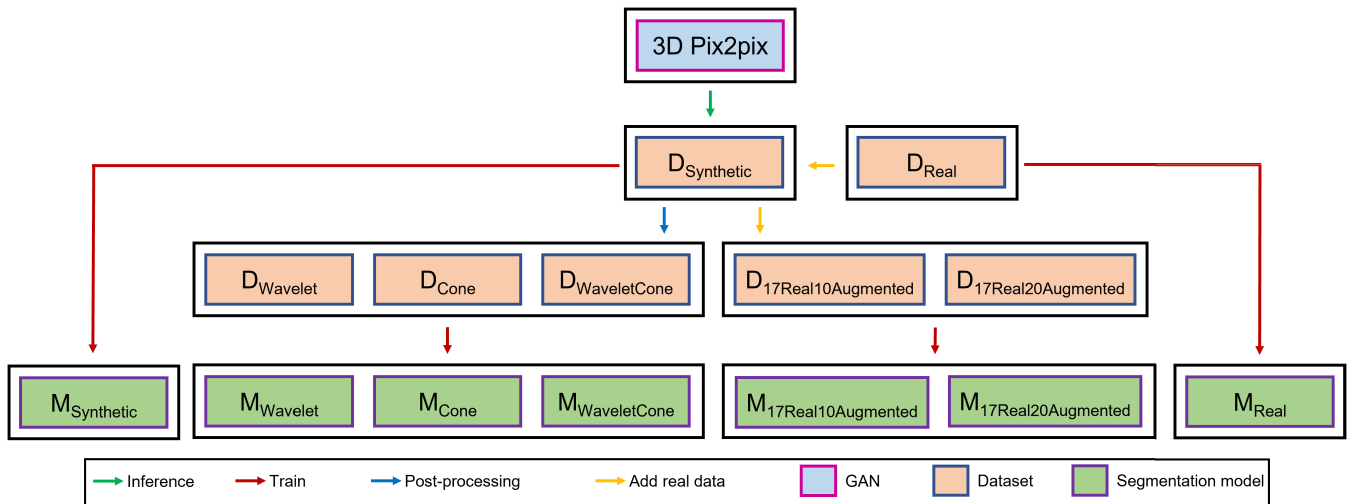
The Pix2pix model was proposed by Isola *et al.* [21] as a solution to image-to-image translation across different imaging domains. This model is capable of generating an output image for each input image by learning a cyclic mapping function across both training domains. The Pix2pix model works as

a conditional paired GAN: given two training domains containing paired images, it learns how to generate new instances of each domain. The loss function was kept the same as presented in the original work – a combination of conditional GAN loss and the L1 distance. This way it is conditioning the GAN performance, assuring the information on the generated output image matches the information provided by the input.

This original work was constructed under a 2D scenario, but in this proposed work an extension to 3D was performed by changing the original architecture of the Pix2pix model.

We considered different architectures for the GAN generator and a 3D U-Net [32] was used to create a 3D version of the Pix2pix model. The discriminator architecture was kept the same, replacing only 2D layers with the correspondent 3D ones. During training of the GAN, data augmentation operations, including blurring and rotation, were performed on the fly, increasing the amount of 3D volumes used without the memory burden of having to save these. The 3D Pix2pix model used here was built using PyTorch [33] and its training was performed over 200 epochs accounting for the images size and computational memory constraints, considering an initial learning rate of 0.0002 and the Adam optimizer.

At inference time, a common problem among image synthesis is the presence of checkerboard artifacts on the generated images. To tackle this problem, which decreases the quality of the synthesized images, we changed the generator architecture as suggested in [34] by replacing the transposed convolutions in the upsampling layers of the 3D U-Net with linear upsampling ones.



**FIGURE 2.** Overview of all the created datasets and trained models in this work. The generative model, 3D Pix2pix, was trained in order to be used to generate synthetic 3D echocardiography datasets. This dataset,  $D_{\text{Synthetic}}$ , was post-processed applying different transformations and 3 other datasets were created –  $D_{\text{Wavelet}}$ ,  $D_{\text{Cone}}$ , and  $D_{\text{WaveletCone}}$ . A fifth dataset completely made of real images,  $D_{\text{Real}}$ , was created and to it, synthetic images from  $D_{\text{Synthetic}}$  were added creating  $D_{17\text{Real}10\text{Augmented}}$  and  $D_{17\text{Real}20\text{Augmented}}$ . All these 7 datasets were used to train 7 3D segmentation models –  $M_{\text{Synthetic}}$ ,  $M_{\text{Wavelet}}$ ,  $M_{\text{Cone}}$ ,  $M_{\text{WaveletCone}}$ ,  $M_{\text{Real}}$ ,  $M_{17\text{Real}10\text{Augmented}}$ , and  $M_{17\text{Real}20\text{Augmented}}$ .

In order to generate synthetic echocardiography images for each of the inference cases, i.e., 3D CT-based heart models, the generator part of the GAN, which translates images from the anatomical labels domain to the echocardiography looking images domain, was used. Anatomical models of the heart [35] obtained from CT were used to create the inference gray scale labeled images, containing anatomical information about the LV, LA, and MYO. The main objective of this work was then accomplished by using the GAN as a data augmentation tool to generate synthetic datasets of 3D echocardiography images of size  $256 \times 256 \times 32$  from these inference images, augmenting the quantity of 3D echocardiographic image data.

### C. SYNTHETIC DATA POST-PROCESSING

During the post-processing stage of the synthetic images generated by the GAN, two different algorithms were experimented. The synthesized images were (a) filtered using the discrete wavelet transform, following Yadav *et al.* [36] work and (b) masked with an Ultrasound cone. The wavelet denoising operation uses wavelets that localize features in the data, preserving important image features while removing unwanted noise, such as checkerboard artifacts. An image mask representing the Ultrasound cone shape was applied to all synthesized images in order to match true Ultrasound data.

### D. 3D SEGMENTATION

The GAN pipeline was able to generate labeled instances of 3D echocardiography images, as the model is capable of performing paired domain translation operations. To investigate the utility of the synthetic images, four 3D segmentation models were trained using the generated synthetic images as training set.

The trained model architecture for the 3D segmentation task was the 3D nnU-Net [28]. This network architecture was proposed as a self-adapting framework for medical image segmentation. This DL model adapts its training scheme, such as the loss function or slight variations on the model architecture, to the dataset being used and to the segmentation task being performed. It automates necessary adaptations to the dataset such as preprocessing, patch and batch size, and inference settings without the need of user intervention.

To train the first of four 3D segmentation models,  $M_{\text{Synthetic}}$ , described in this section, a labeled dataset made of 27 synthetically generated 3D echocardiography images ( $256 \times 256 \times 32$ ),  $D_{\text{Synthetic}}$ , was used. This dataset was obtained from the proposed 3D GAN pipeline at inference time, using anatomical labels from 27 CT 3D anatomical models.

To evaluate the effect of the post-processing operations on the synthesized images, three other datasets were created –  $D_{\text{Wavelet}}$ ,  $D_{\text{Cone}}$ , and  $D_{\text{WaveletCone}}$  – and three additional segmentation models were trained using these –  $M_{\text{Wavelet}}$ ,  $M_{\text{Cone}}$ , and  $M_{\text{WaveletCone}}$ , respectively (Fig. 2).  $D_{\text{Wavelet}}$  was made of the original synthetic images from the  $D_{\text{Synthetic}}$  dataset but where the wavelet denoising post-processing algorithm was applied, and  $D_{\text{Cone}}$ , was composed by the original synthetic images with the cone reshaping post-processing operation. Finally, a fourth dataset where both post-processing transformations – wavelet denoising and cone reshaping – were applied to the original synthetic images,  $D_{\text{WaveletCone}}$ , was created. All four datasets contained 27 3D echocardiography images with corresponding anatomical labels for the LV, LA and MYO.

All four 3D segmentation models,  $M_{\text{Synthetic}}$ ,  $M_{\text{Wavelet}}$ ,  $M_{\text{Cone}}$ , and  $M_{\text{WaveletCone}}$ , using nnU-Net, were trained on a

5-fold cross validation scenario during 800 epochs. The initial learning rate was 0.01 and the segmentation models were also built using PyTorch [33]. The loss function was a combination of dice and cross-entropy losses, as described in the original work by Isensee *et al.* [28].

Dice scores were used to assess the quality of the segmentations. This score measures the overlap between the predicted segmentation and the ground truth label extracted from the CT anatomical models. For each segmented structure the Dice score obtained at validation time is a value between 0 and 1, where the latter represents a perfect overlap between the prediction and the ground truth.

### E. REAL DATA COMBINED WITH SYNTHETIC-DATA AUGMENTATION

In their work, Lustermans *et al.* [16] showed that adding real data to GAN-generated synthetic datasets can help improve DL models train.

To facilitate a clearer analysis of the influence of using synthetic data to train DL models and the utility of this GAN as a data augmentation tool, three other segmentation models were trained on the datasets  $D_{\text{Real}}$ ,  $D_{17\text{Real}10\text{Augmented}}$ , and  $D_{17\text{Real}20\text{Augmented}}$ .  $D_{\text{Real}}$  contained 17 real 3D echocardiography volumes acquired with GE Vivid Ultrasound scanners and labeled by a cardiologist.

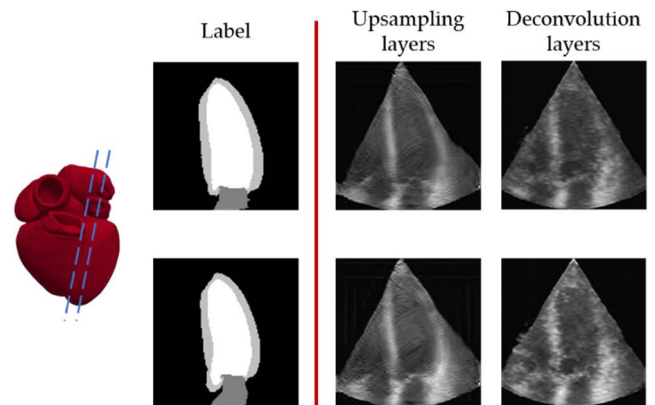
$D_{17\text{Real}10\text{Augmented}}$  and  $D_{17\text{Real}20\text{Augmented}}$  were made up of the same 17 real volumes just described together with 10 and 20 synthetic GAN-generated 3D echocardiography images, respectively. Thus allowing to assess the influence of using such type of images during DL models training (Fig. 2).

The 3D segmentation models trained on these datasets were  $M_{\text{Real}}$ ,  $M_{17\text{Real}10\text{Augmented}}$ , and  $M_{17\text{Real}20\text{Augmented}}$ , respectively. All models used the nnU-Net architecture implemented with Pytorch. Similar to the ones described on Section II-D, they were trained for 800 epochs on a 5-fold cross validation scenario, with the same learning rate and loss function.

At inference time, a test set including real 3D echocardiography images was segmented by the three aforementioned models. To compare the segmentation results with the ones obtained from a cardiologist, Dice scores and Volume Similarity (VS) were calculated and used as comparison metrics. VS is calculated as the size of the segmented structures and is of high relevance in a 3D scenario since Dice score presents some limitations. Similarly to the Dice score, this evaluation metric takes values between 0 and 1 but is not overlap-based. Instead, it is a volume based parameter where the absolute volume of a region in one segmentation is compared with the corresponding region volume in the other segmentation [37].

## III. RESULTS

This work's results are presented as follows: Section III-A focuses on the GAN training, architectural modifications performed on the 3D Pix2pix model and their influence on the synthesized images. In Section III-B the influence of post-processing the synthetic images is shown. Finally,



**FIGURE 3.** Influence of architectural changes on the GAN generator to remove checkerboard artifacts. At inference time, a 3D anatomical model was used to extract the anatomical labels. The first column shows 2 different slices of this volume at different rotation angles. The middle column shows that synthesizing images using a GAN with upsampling layers smoothens the checkerboard artifacts but introduces blurring, which is not visible on the images when using a GAN with deconvolution layers (right column). Deconvolution layers are preferred to upsampling ones.

Sections III-C and III-D show the segmentation predictions from several models trained on different 3D echocardiography datasets (Fig. 2), as described in Sections II-C and II-D.

### A. GAN ARCHITECTURE AND TRAINING

The chosen GAN architecture influenced the final results. 3D U-Net was chosen as the generator architecture due to its good performance in the medical image domain. The model was trained on a NVIDIA GeForce RTX 2080 Ti GPU and training took five days.

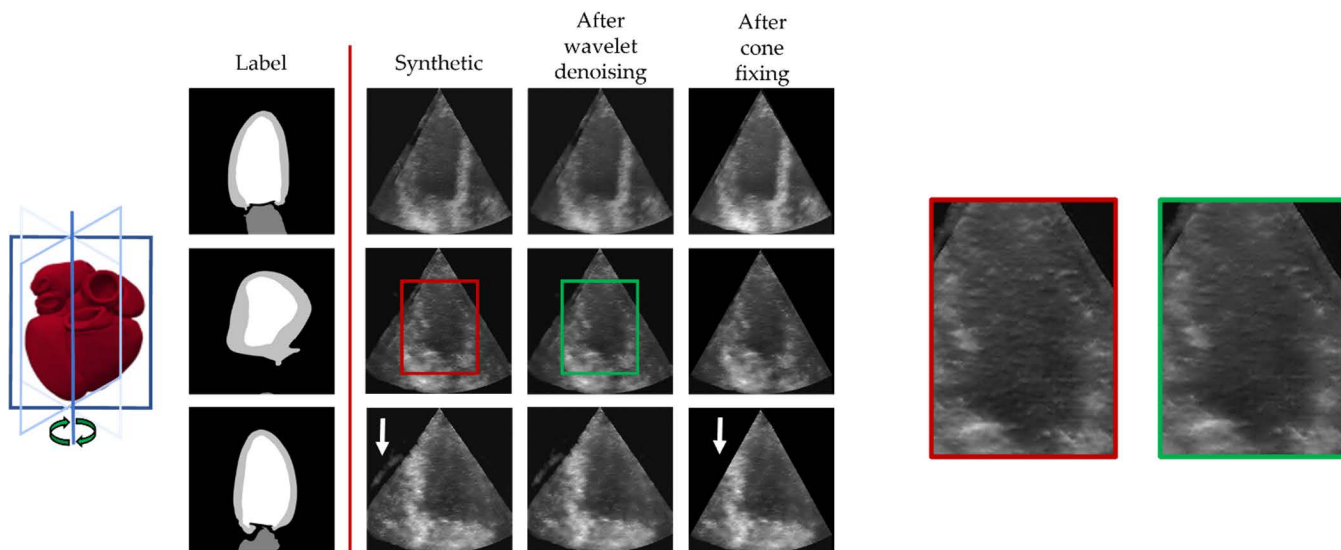
After applying the architectural changes described in Section II-B to remove the checkerboard artifacts, it seemed like these became less visible or even disappeared. However, this correction created some unwanted blurring on the generated images (Fig. 3), therefore the deconvolution layers were used instead of upsampling, and the synthesized images were post-processed to remove the checkerboard artifacts.

### B. SYNTHETIC DATA POST-PROCESSING

After training the 3D GAN model and generating synthetic images corresponding to the input anatomical models, as described in Section II-C, the obtained 3D echocardiography images were post-processed in order to remove the aforementioned checkerboard artifacts.

The cone edges were slightly wavy in some cases and checkerboard artifacts were sometimes present. The post-processing experiment, where different transformations were applied to the synthesized images, showed that applying these can give a more realistic aspect to the GAN-generated images, ensuring that the anatomical information remained intact.

Performing these operations allowed to give a more realistic look to the generated echocardiography images (Fig. 4).



**FIGURE 4.** 3D Pix2pix model inference results and post-processing step. At inference time, the anatomical labels were extracted from a 3D heart model. The first column shows 3 different rotation planes of this volume at different rotation angles. After generating the correspondent synthetic ultrasound image (second column) for this inference case, it was post-processed applying a wavelet denoising transformation to eliminate the checkerboard artifacts (third column) and also a cone reshaping step to smooth the wavy edges of the ultrasound cone (fourth column). Post-processing operations give a more realistic look to the synthesized images as indicated by the enlarged areas framed in red and green (wavelet denoise) and the white arrows (cone reshape).

### C. SEGMENTATION FROM SYNTHETIC DATASETS

Anatomical models were used in order to synthesize 27 3D echocardiography images. These were then used to create the synthetic datasets that were used to train 3D segmentation algorithms, as described in section II-D. Post-processing operations were performed on these images to create the  $D_{Wavelet}$ ,  $D_{Cone}$ , and  $D_{WaveletCone}$  datasets. Table 1 shows the average Dice scores (average  $\pm$  standard deviation) of each segmented structure (LV, LA, and MYO) for each trained model –  $M_{Synthetic}$ ,  $M_{Wavelet}$ ,  $M_{Cone}$ , and  $M_{WaveletCone}$ , obtained from the validation dataset. Training took around five days for each fold using a NVIDIA GeForce RTX 2080 GPU, for all epochs. The complete table with all the Dice scores obtained for each training fold of each model can be found in Appendix – Table 5.

Adding to the Dice scores and to sustain the usability of synthetic images to train segmentation algorithms, Fig. 5 shows the 3D segmentation for an inference 3D echocardiography image acquired from a real subject. Each trained segmentation model was tested on real cases, at inference time.

### D. SEGMENTATION FROM COMBINED DATASETS

In Table 2 one can see the average Dice scores (average  $\pm$  standard deviation), obtained at validation time, of each segmented structure (LV, LA, and MYO) for each trained model on the combined datasets:  $M_{Real}$ ,  $M_{17Real10Augmented}$ , and  $M_{17Real20Augmented}$ . In Appendix – Table 4 the complete table with all the Dice scores for each trained fold of all three models can be found.

**TABLE 1.** Average validation dice scores (average  $\pm$  standard deviation) of each segmented structure (LV, LA, and MYO) for each trained model on completely synthetic datasets –  $M_{Synthetic}$ ,  $M_{Wavelet}$ ,  $M_{Cone}$ , and  $M_{WaveletCone}$ . The best scores are highlighted.

	Models			
	$M_{Synthetic}$	$M_{Wavelet}$	$M_{Cone}$	$M_{WaveletCone}$
LV	0.926 $\pm$ 0.006	<b>0.927</b> $\pm$ <b>0.005</b>	0.926 $\pm$ 0.006	0.924 $\pm$ 0.008
LA	<b>0.818</b> $\pm$ <b>0.011</b>	0.816 $\pm$ 0.010	0.816 $\pm$ 0.021	0.814 $\pm$ 0.016
MYO	<b>0.808</b> $\pm$ <b>0.016</b>	0.808 $\pm$ 0.017	0.803 $\pm$ 0.018	0.801 $\pm$ 0.023

Fig. 6 shows the predicted segmentations given by these trained models, next to the ground truth segmentation provided by a cardiologist. The models were tested on a test set made of 3D echocardiography images from real subjects.

To compare the output segmentation from the DL models, the Dice scores and VS were calculated based on the predicted segmentations and the anatomical labels from a cardiologist and the results are in Table 3.

## IV. DISCUSSION

In this work we built a pipeline to generate synthetic 3D labeled echocardiography images using a GAN model. These realistic-looking synthetic datasets were used to train 3D DL models to segment the LV, LA, and MYO.

Moreover, combined datasets including synthetic and real 3D images were created, with the VS metric supporting that generated 3D echocardiography images can be used to train

**TABLE 2.** Average validation dice scores (average  $\pm$  standard deviation) of each segmented structure (LV, LA, and MYO) for each trained model on combined datasets –  $M_{\text{Real}}$ ,  $M_{17\text{Real}10\text{Augmented}}$ , and  $M_{17\text{Real}20\text{Augmented}}$ . The best scores are highlighted.

	Models		
	$M_{\text{Real}}$	$M_{17\text{Real}10\text{Augmented}}$	$M_{17\text{Real}20\text{Augmented}}$
LV	<b>0.938 <math>\pm</math> 0.008</b>	0.928 $\pm$ 0.006	0.927 $\pm$ 0.007
LA	<b>0.862 <math>\pm</math> 0.023</b>	0.830 $\pm$ 0.016	0.826 $\pm$ 0.017
MYO	0.724 $\pm$ 0.028	<b>0.767 <math>\pm</math> 0.027</b>	0.763 $\pm$ 0.025

**TABLE 3.** Average test set dice scores (average  $\pm$  standard deviation) of each segmented structure (LV, LA, and MYO) and Volume Similarity of the segmented volume for the  $M_{\text{Real}}$ ,  $M_{17\text{Real}10\text{Augmented}}$ , and  $M_{17\text{Real}20\text{Augmented}}$  models. The best scores are highlighted.

	Models		
	$M_{\text{Real}}$	$M_{17\text{Real}10\text{Augmented}}$	$M_{17\text{Real}20\text{Augmented}}$
	Dice score		
LV	0.924 $\pm$ 0.019	<b>0.929 <math>\pm</math> 0.020</b>	0.922 $\pm$ 0.017
LA	<b>0.876 <math>\pm</math> 0.021</b>	0.874 $\pm$ 0.020	0.867 $\pm$ 0.022
MYO	0.666 $\pm$ 0.041	<b>0.708 <math>\pm</math> 0.053</b>	0.680 $\pm$ 0.063
	Volume Similarity		
Heart Volume	0.831 $\pm$ 0.038	<b>0.844 <math>\pm</math> 0.047</b>	0.836 $\pm$ 0.041

DL models, as data augmentation. Segmentation tasks were considered to exemplify the utility of the synthesized data, however the pipeline is generic and could be applied to generate other imaging data and train any DL tasks with anatomical labels as input, as further discussed in this section.

A brief discussion on future applications and modifications of this approach is also presented.

### A. 3D Pix2pix GAN—QUALITATIVE ANALYSIS

The pipeline synthesizes 3D echocardiographic datasets with corresponding labels delineating different structures in the images.

After training the 3D Pix2pix GAN model, a qualitative analysis of the synthesized images indicated that the main structures of the heart were well delineated in the generated images (Fig. 1, inference stage). Moreover, image details such as the cone, noise, and speckle patterns are also present and are continuous throughout each volume.

### B. POST-PROCESSING AND 3D SEGMENTATION—SYNTHETIC DATASETS

To evaluate the utilization of synthetic images for research purposes and the extent to which the post-processing transformations affected the final results, four segmentation models were trained using four different datasets, as described earlier in Section III-C.

Despite the very small differences in the Dice scores shown in Table 1 and in Appendix – Table 5, the inference

segmentations (Fig. 5) support the idea that the model trained on the dataset whose images were not post-processed,  $M_{\text{Synthetic}}$ , provided the best segmentation prediction.

The results regarding the influence of the post-processing step on the synthetically generated images supported the fact that applying a wavelet denoising transformation or cone reshaping, or even both transformations together, to these, in order to make the synthetic images look even more realistic, does not necessarily lead to better results when segmenting the LV, LA, and MYO (Fig. 5). This result shows some dependence on the DL task being performed. We segmented large volumes of the 3D image, comparing to its whole content. For this reason, the subtle differences in the voxels intensities that create the checkerboard artifacts do not seem to affect the prediction of the segmentation model.

To create the used synthetic datasets, CT acquired 3D anatomical models of the heart were used to extract the anatomical labels and create the input cases to the 3D GAN. The segmentation results and the echocardiography-looking aspect of the synthetic images pointed towards the generalization of this pipeline, as it can synthesize 3D echocardiography images, having as labels source different types of 3D models of the heart. The methodology to generate synthetic datasets can be generalized to other modalities, diseases, organs, as well as structures within the same organ (sub-regions of the heart, for example).

Shin *et al.* [5] and Shorten and Khoshgoftaar [38] showed that GANs can be widely used to perform data augmentation of medical image datasets. The work from these authors, together with the presented results, encourage the main contributions of this work stating that GANs can be used to generate synthetic images with labels, working as a data augmentation strategy, and tackling the concern of scarcity of 3D echocardiography labeled datasets, especially if there are underrepresented data samples within the available real datasets.

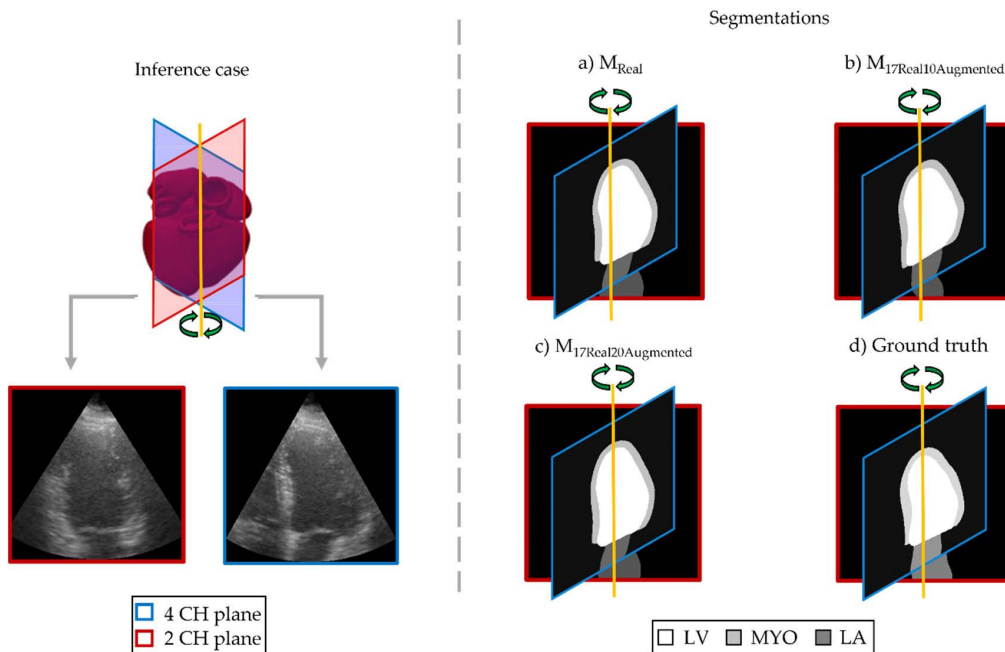
### C. 3D SEGMENTATION—COMBINED DATASETS

Further results on the usage of synthetic datasets were explored and presented in Section III-D. Here, three datasets made of GAN generated and real 3D images were used to train more segmentation models and further evaluate the influence of the presence of real data in these datasets, as demonstrated in [16].

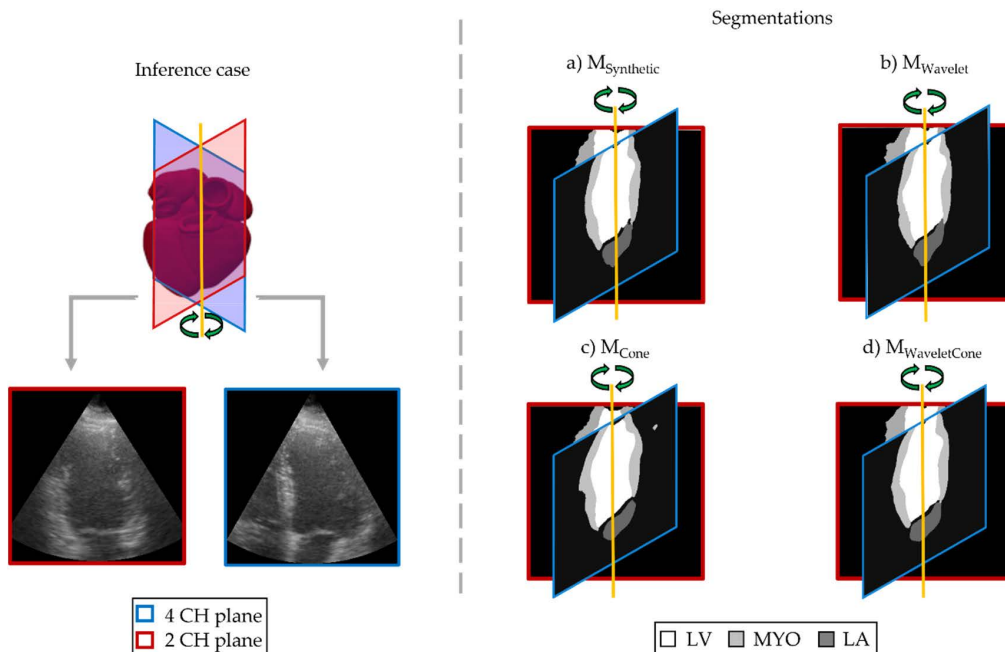
Fig. 6 a), b), and c) showed the anatomical segmentations of the LV, LA, and MYO predicted by the best trained fold of each model –  $M_{\text{Real}}$ ,  $M_{17\text{Real}10\text{Augmented}}$ , and  $M_{17\text{Real}20\text{Augmented}}$ . From the qualitative analysis, the segmentations delineate well the anatomical structures in consideration throughout the whole volume. At the same time, and similarly to what was discussed on Section IV-B, the average Dice scores presented in Table 2 led to the conclusion that having a dataset of real images combined with synthetic ones leads to more accurate final segmentations.

From the obtained results is also possible to assess the influence of using combined datasets with different





**FIGURE 5.** Inference segmentation results from each trained model on synthetic datasets. On the left is shown a schematic representation of the heart and 2 cutting planes correspondent to a real 3D echocardiography image from the test set: the 4-chamber (CH), with blue frame, and the 2-CH, with red frame. On the right, the LV, LA, and MYO segmentation results provided by each of the 4 segmentation models: a)  $M_{Synthetic}$ , b)  $M_{Wavelet}$ , c)  $M_{Cone}$ , and d)  $M_{WaveletCone}$  follow. A qualitative analysis of the segmentation results from each of the models, shows that the one where the training data was not post-processed,  $M_{Synthetic}$ , gives the best output due to a smoother segmentation of the relevant structures.



**FIGURE 6.** Inference segmentation results from the trained models on augmented datasets with synthetic images. On the left is shown a schematic representation of the heart and 2 cutting planes correspondent to a real 3D echocardiography image from the test set: the 4-CH, with blue frame, and the 2-CH, with red frame. On the right, the LV, LA, and MYO segmentation results provided by the following 3 segmentation models: a)  $M_{Real}$ , b)  $M_{17Real10Augmented}$ , and c)  $M_{17Real20Augmented}$ , follow. To allow comparison and measure the Dice score and VS, d) shows the ground truth segmentation performed by a cardiologist. A qualitative analysis of the segmentation results from each of the models, shows that combining synthetic with real data improves the segmentation output due to a more accurate segmentation of the relevant structures.

percentages of synthetic data. Table 2 and Table 3 show that adding synthetic data to the initial real dataset

improves the 3D segmentation of real 3D echocardiography images. They also show that adding larger amounts

of synthetic data does not improve the results to a large extent.

Fig. 6 d) showed the ground truth inference case segmentation performed by a cardiologist. From these ground truth segmentations available for all the cases in the test set, the Dice scores and the VS in Table 3 were calculated.

Given the 3D nature of the task and due to the Dice metric limitations, the VS was additionally calculated and used as comparison metric. In particular,  $M_{17\text{Real}10\text{Augmented}}$  showed to perform better at segmenting when the Dice score was considered as performance metric. On the other hand,  $M_{17\text{Real}20\text{Augmented}}$  performed better in terms of VS metric. These results showed that the models trained on the combined datasets, i.e., with real and synthetic images, provided more accurate segmentation outputs (the 3D volume), relatively to the model trained with only real data,  $M_{\text{Real}}$ . The results support the previous work done by [16], confirming that including synthetic data on datasets made of real data improves and helps the final outcome of the DL models.

Additionally, this result reinforces that the proposed pipeline, relying on a 3D GAN model, can be used as a data augmentation tool. This framework arises as a solution to the lack of publicly available medical labeled datasets.

#### D. FURTHER APPLICATIONS

The presented pipeline has the potential to be further explored. As the demand for medical images is increasing, the proposed approach can be extended to synthesize images from other imaging modalities other than Ultrasound, such as MR or CT. It can also generate images where other organs are represented or even fetuses [13].

Another extension of this work would be to use different types of 3D models from which ground truth anatomical labels could be extracted. Besides anatomical models, animated or biophysical models represent other options that can be considered. The usage of anatomical models of pathological hearts are another possible extension, in order to generate pathological 3D echocardiography scans. Depending on the type of 3D model being considered, different annotations can be extracted, increasing the amount of clinically relevant tasks where these synthetic datasets can be used.

The generated 3D echocardiography images illustrated a heart volume during one time step of the whole cardiac cycle (end-diastole). It would be of great interest to generate 3D images of the heart during other cardiac cycle events and even to generate a beating volume throughout time, as high temporal resolution is one of the main strengths of Ultrasound imaging. On the other hand, a limitation to the Ultrasound images generation is that different scanning probe combinations lead to the acquisition of images with different quality levels. This large variability makes the GAN learning process more complex.

In this work we explored, to an extent, the effects that architectural changes of the GAN model have on the final

synthesized images. We used different architectures for the GAN generator but more 3D CNNs exist and are showing up every day. These can be used to train the generative models, since DL strategies are becoming extremely common to use as medical image synthesis and analysis tools. Once the images were synthesized, we used wavelet denoising and an in-house developed algorithm to fix the Ultrasound cone edges. However, there are other denoising transformations and cone reshaping algorithms that can be experimented to post-process the images.

We trained several DL models to perform 3D segmentation to show that synthesized images can be used as input to train DL models. Nevertheless, the pipeline is generic and could be applied to other DL tasks that automatically assign anatomical labels to images, e.g., structure/feature recognition or automatic structural measurements. Furthermore, the GAN-generated labeled datasets are not only useful as input to train DL models but also could be used to train researchers and clinicians on image analysis.

Finally, during this pipeline development, computational memory constraints were present, mainly due to the large size of 3D volumes, complicating the process of developing a framework adapted to these. Future work will include study strategies to overcome these limitations.

#### V. CONCLUSION

An automatic data augmentation pipeline to create 3D echocardiography images and corresponding anatomical labels using a 3D GAN model was proposed. DL models are becoming widely used in clinical workflows and large volumes of medical data is a fundamental requirement to develop such algorithms with high accuracy. Generating synthetic data that could be used for the purpose of training DL models is of utmost importance since this generative model can become a widely used tool to address the existent lack of publicly available data and increasing challenges with moving data due to privacy regulations. Furthermore, the proposed methodology not only generates synthetic 3D echocardiography images but also associates labels to these synthetic images, eliminating the need for experienced professionals to do so, and without adding potential bias in the labels.

The proposed GAN model shows a generalization component since it can generate synthetic echocardiography images using 3D anatomical models of the heart obtained for imaging modalities other than from Ultrasound.

The obtained results in this work indicate that synthetic datasets made up of GAN-generated 3D echocardiography images, and respective labels, are a good data augmentation resource to train and develop DL models that can be used to perform different medical tasks in the cardiac imaging domain, such as heart segmentation, where real patients' data is analyzed.

#### APPENDIX

See Table 4 and Table 5.

**TABLE 4.** Validation dice scores of each segmented structure (LV, LA, and MYO) for each trained model on combined datasets— $M_{Real}$ ,  $M_{17Real10Augmented}$ , and  $M_{17Real20Augmented}$ . The higher the score, the better the agreement between the model prediction and the ground truth segmentation. The best training fold of each model is highlighted.

	Models														
	$M_{Real}$					$M_{17Real10Augmented}$					$M_{17Real20Augmented}$				
	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
LV	0.933	0.932	<b>0.950</b>	0.930	0.943	0.924	0.929	0.919	<b>0.938</b>	0.930	0.917	0.928	<b>0.935</b>	0.934	0.923
LA	0.837	0.869	0.873	0.837	<b>0.896</b>	0.830	0.841	0.820	0.808	<b>0.853</b>	0.831	<b>0.841</b>	0.838	0.829	0.793
MYO	0.710	0.699	<b>0.766</b>	0.697	0.750	0.745	0.745	0.779	<b>0.815</b>	0.753	0.715	0.771	0.766	0.780	<b>0.785</b>

**TABLE 5.** Validation dice scores of each segmented structure (LV, LA, and MYO) for each trained model on completely synthetic datasets –  $M_{Synthetic}$ ,  $M_{Wavelet}$ ,  $M_{Cone}$ , and  $M_{WaveletCone}$ . The higher the score, the better the agreement between the model prediction and the ground truth segmentation. The best training fold of each model is highlighted.

	Models																			
	$M_{Synthetic}$					$M_{Wavelet}$					$M_{Cone}$					$M_{WaveletCone}$				
	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
LV	0.924	0.930	0.924	0.918	<b>0.934</b>	0.927	0.930	0.923	0.919	<b>0.934</b>	0.926	0.930	0.921	0.918	<b>0.933</b>	0.928	0.928	0.914	0.914	<b>0.935</b>
LA	<b>0.833</b>	0.831	0.809	0.810	0.807	0.821	<b>0.831</b>	0.806	0.815	0.805	0.837	<b>0.842</b>	0.805	0.811	0.787	<b>0.834</b>	0.832	0.803	0.793	0.807
MYO	<b>0.824</b>	0.822	0.794	0.784	0.816	<b>0.829</b>	0.822	0.791	0.785	0.814	<b>0.824</b>	0.819	0.783	0.780	0.809	<b>0.828</b>	0.813	0.775	0.773	0.816

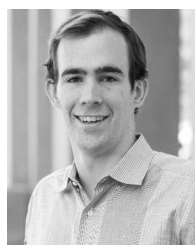
## REFERENCES

- [1] F. M. Asch, N. Poilvert, T. Abraham, M. Jankowski, J. Cleve, M. Adams, N. Romano, H. Hong, V. Mor-Avi, R. P. Martin, and R. M. Lang, "Automated echocardiographic quantification of left ventricular ejection fraction without volume measurements using a machine learning algorithm mimicking a human expert," *Circulat., Cardiovascular Imag.*, vol. 12, no. 9, Sep. 2019, Art. no. e009303, doi: [10.1161/CIRCIMAGING.119.009303](https://doi.org/10.1161/CIRCIMAGING.119.009303).
- [2] J. Scheetz, P. Rothschild, M. McGuinness, X. Hadoux, H. P. Soyer, M. Janda, J. J. Condon, L. Oakden-Rayner, L. J. Palmer, S. Keel, and P. van Wijngaarden, "A survey of clinicians on the use of artificial intelligence in ophthalmology, dermatology, radiology and radiation oncology," *Sci. Rep.*, vol. 11, no. 1, Mar. 2021, Art. no. 1, doi: [10.1038/s41598-021-84698-5](https://doi.org/10.1038/s41598-021-84698-5).
- [3] A. Aljuaid and M. Anwar, "Survey of supervised learning for medical image processing," *Social Netw. Comput. Sci.*, vol. 3, no. 4, p. 292, May 2022, doi: [10.1007/s42979-022-01166-1](https://doi.org/10.1007/s42979-022-01166-1).
- [4] L. P. de Isla, D. V. Balcones, C. Fernández-Golfín, P. Marcos-Alberca, C. Almería, J. L. Rodrigo, C. Macaya, and J. Zamorano, "Three-dimensional-wall motion tracking: A new and faster tool for myocardial strain assessment: Comparison with two-dimensional-wall motion tracking," *J. Amer. Soc. Echocardiogr.*, vol. 22, no. 4, pp. 325–330, Apr. 2009, doi: [10.1016/j.echo.2009.01.001](https://doi.org/10.1016/j.echo.2009.01.001).
- [5] H.-C. Shin, N. A. Tenenholtz, J. K. Rogers, C. G. Schwarz, M. L. Senjem, J. L. Gunter, K. P. Andriole, and M. Michalski, "Medical image synthesis for data augmentation and anonymization using generative adversarial networks," in *Simulation and Synthesis in Medical Imaging*, Cham, Switzerland: Springer, 2018, pp. 1–11, doi: [10.1007/978-3-030-00536-8\\_1](https://doi.org/10.1007/978-3-030-00536-8_1).
- [6] R. J. Chen, M. Y. Lu, T. Y. Chen, D. F. K. Williamson, and F. Mahmood, "Synthetic data in machine learning for medicine and healthcare," *Nature Biomed. Eng.*, vol. 5, no. 6, pp. 493–497, Jun. 2021, doi: [10.1038/s41551-021-00751-8](https://doi.org/10.1038/s41551-021-00751-8).
- [7] M. J. M. Chuquicusma, S. Hussein, J. Burt, and U. Bagci, "How to fool radiologists with generative adversarial networks? A visual Turing test for lung cancer diagnosis," in *Proc. IEEE 15th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2018, pp. 240–244, doi: [10.1109/ISBI.2018.8363564](https://doi.org/10.1109/ISBI.2018.8363564).
- [8] A. Thorstensen, H. Dalen, B. H. Amundsen, S. A. Aase, and A. Stoylen, "Reproducibility in echocardiographic assessment of the left ventricular global and regional function, the HUNT study," *Eur. J. Echocardiogr.*, vol. 11, no. 2, pp. 149–156, Mar. 2010, doi: [10.1093/ejehocardiogr/jep188](https://doi.org/10.1093/ejehocardiogr/jep188).
- [9] I. Banerjee, C. E. Catalano, G. Patané, and M. Spagnuolo, "Semantic annotation of 3D anatomical models to support diagnosis and follow-up analysis of musculoskeletal pathologies," *Int. J. Comput. Assist. Radiol. Surgery*, vol. 11, no. 5, pp. 707–720, May 2016, doi: [10.1007/s11548-015-1327-6](https://doi.org/10.1007/s11548-015-1327-6).
- [10] W. P. Segars, G. Sturgeon, S. Mendonca, J. Grimes, and B. M. W. Tsui, "4D XCAT phantom for multimodality imaging research," *Med. Phys.*, vol. 37, no. 9, pp. 4902–4915, Sep. 2010, doi: [10.1118/1.3480985](https://doi.org/10.1118/1.3480985).
- [11] W. Kainz, E. Neufeld, W. E. Bolch, C. G. Graff, C. H. Kim, N. Kuster, B. Lloyd, T. Morrison, P. Segars, Y. S. Yeom, M. Zankl, X. G. Xu, and B. M. W. Tsui, "Advances in computational human phantoms and their applications in biomedical engineering—A topical review," *IEEE Trans. Radiat. Plasma Med. Sci.*, vol. 3, no. 1, pp. 1–23, Jan. 2019, doi: [10.1109/TRPMS.2018.2883437](https://doi.org/10.1109/TRPMS.2018.2883437).
- [12] A. Gilbert, M. Marciniak, C. Rodero, P. Lamata, E. Samsset, and K. Mcleod, "Generating synthetic labeled data from existing anatomical models: An example with echocardiography segmentation," *IEEE Trans. Med. Imag.*, vol. 40, no. 10, pp. 2783–2794, Oct. 2021, doi: [10.1109/TMI.2021.3051806](https://doi.org/10.1109/TMI.2021.3051806).
- [13] C. W. Roy, D. Marini, W. P. Segars, M. Seed, and C. K. Macgowan, "Fetal XCMR: A numerical phantom for fetal cardiovascular magnetic resonance imaging," *J. Cardiovascular Magn. Reson.*, vol. 21, no. 1, p. 29, May 2019, doi: [10.1186/s12968-019-0539-2](https://doi.org/10.1186/s12968-019-0539-2).
- [14] A. Perperidis, "Postprocessing approaches for the improvement of cardiac ultrasound B-mode images: A review," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 63, no. 3, pp. 470–485, Mar. 2016, doi: [10.1109/TUFFC.2016.2526670](https://doi.org/10.1109/TUFFC.2016.2526670).
- [15] L. Chai, J.-Y. Zhu, E. Shechtman, P. Isola, and R. Zhang, "Ensembling with deep generative views," 2021, *arXiv:2104.14551*.
- [16] D. R. P. R. M. Lusterms, S. Amirrajab, M. Veta, M. Breeuwer, and C. M. Scannell, "Optimized automated cardiac MR scar quantification with GAN-based data augmentation," 2021, *arXiv:2109.12940*.
- [17] J. Pedrosa, S. Queirós, O. Bernard, J. Engvall, T. Edvardsen, E. Nagel, and J. D'hooge, "Fast and fully automatic left ventricular segmentation and tracking in echocardiography using shape-based b-spline explicit active surfaces," *IEEE Trans. Med. Imag.*, vol. 36, no. 11, pp. 2287–2296, Nov. 2017, doi: [10.1109/TMI.2017.2734959](https://doi.org/10.1109/TMI.2017.2734959).
- [18] H. Uzunova, J. Ehrhardt, and H. Handels, "Memory-efficient GAN-based domain translation of high resolution 3D medical images," *Computerized Med. Imag. Graph.*, vol. 86, Dec. 2020, Art. no. 101801, doi: [10.1016/j.compmedimag.2020.101801](https://doi.org/10.1016/j.compmedimag.2020.101801).

- [19] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," 2017, *arXiv:1703.10593*.
- [20] Y. Huo, Z. Xu, H. Moon, S. Bao, A. Assad, T. K. Moyo, M. R. Savona, R. G. Abramson, and B. A. Landman, "SynSeg-Net: Synthetic segmentation without target modality ground truth," *IEEE Trans. Med. Imag.*, vol. 38, no. 4, pp. 1016–1025, Apr. 2019, doi: [10.1109/TMI.2018.2876633](https://doi.org/10.1109/TMI.2018.2876633).
- [21] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," 2016, *arXiv:1611.07004*.
- [22] S. Amirrajab, S. Abbasi-Sureshjani, Y. A. Khalil, C. Lorenz, J. Weese, J. Pluim, and M. Breeuwer, "XCAT-GAN for synthesizing 3D consistent labeled cardiac MR images on anatomically variable XCAT phantoms," 2020, *arXiv:2007.13408*.
- [23] Y. Hu, E. Gibson, L.-L. Lee, W. Xie, D. C. Barratt, T. Vercauteren, and J. A. Noble, "Freehand ultrasound image simulation with spatially-conditioned generative adversarial networks," in *Molecular Imaging, Reconstruction and Analysis of Moving Body Organs, and Stroke Imaging and Treatment*, vol. 10555, M. J. Cardoso et al., Eds. Cham, Switzerland: Springer, 2017, pp. 105–115, doi: [10.1007/978-3-319-67564-0\\_11](https://doi.org/10.1007/978-3-319-67564-0_11).
- [24] S. Abbasi-Sureshjani, S. Amirrajab, C. Lorenz, J. Weese, J. Pluim, and M. Breeuwer, "4D semantic cardiac magnetic resonance image synthesis on XCAT anatomical model," 2020, *arXiv:2002.07089*.
- [25] T. Park, M.-Y. Liu, T.-C. Wang, and J.-Y. Zhu, "Semantic image synthesis with spatially-adaptive normalization," 2019, *arXiv:1903.07291*.
- [26] M. D. Cirillo, D. Abramian, and A. Eklund, "Vox2Vox: 3D-GAN for brain tumour segmentation," 2020, *arXiv:2003.13653*.
- [27] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," 2015, *arXiv:1505.04597*.
- [28] F. Isensee, P. F. Jaeger, S. A. A. Kohl, J. Petersen, and K. H. Maier-Hein, "nnU-Net: A self-configuring method for deep learning-based biomedical image segmentation," *Nature Methods*, vol. 18, no. 2, pp. 203–211, Feb. 2021, doi: [10.1038/s41592-020-01008-z](https://doi.org/10.1038/s41592-020-01008-z).
- [29] M. Alsharqi, W. J. Woodward, J. A. Mumith, D. C. Markham, R. Upton, and P. Leeson, "Artificial intelligence and echocardiography," *Echo Res. Pract.*, vol. 5, no. 4, pp. R115–R125, Dec. 2018, doi: [10.1530/ERP-18-0056](https://doi.org/10.1530/ERP-18-0056).
- [30] A. Østvik, E. Smistad, T. Espeland, E. A. R. Berg, and L. Lovstakken, "Automatic myocardial strain imaging in echocardiography using deep learning," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Cham, Switzerland: Springer, 2018, pp. 309–316, doi: [10.1007/978-3-030-00889-5\\_35](https://doi.org/10.1007/978-3-030-00889-5_35).
- [31] V7 Ltd. *V7—AI Data Platform for ML Teams*. Accessed: Feb. 25, 2022. [Online]. Available: <https://www.v7labs.com/>
- [32] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3D U-Net: Learning dense volumetric segmentation from sparse annotation," 2016, *arXiv:1606.06650*.
- [33] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, and A. Desmaison, "PyTorch: An imperative style, high-performance deep learning library," 2019, *arXiv:1912.01703*.
- [34] A. Odena, V. Dumoulin, and C. Olah, "Deconvolution and checkerboard artifacts," *Distill*, vol. 1, no. 10, p. e3, Oct. 2016, doi: [10.23915/distill.00003](https://doi.org/10.23915/distill.00003).
- [35] C. Rodero, M. Stocchi, M. Marciniak, S. Longobardi, J. Whitaker, M. D. O'Neill, K. Gillette, C. Augustin, G. Plank, E. J. Vigmond, P. Lamata, and S. A. Niederer, "Linking statistical shape models and simulated function in the healthy adult human heart," *PLOS Comput. Biol.*, vol. 17, no. 4, Apr. 2021, Art. no. e1008851, doi: [10.1371/journal.pcbi.1008851](https://doi.org/10.1371/journal.pcbi.1008851).
- [36] A. K. Yadav, R. Roy, A. P. Kumar, C. S. Kumar, and S. K. Dhakad, "De-noising of ultrasound image using discrete wavelet transform by symlet wavelet and filters," in *Proc. Int. Conf. Adv. Comput., Commun. Informat. (ICACCI)*, Aug. 2015, pp. 1204–1208, doi: [10.1109/ICACCI.2015.7275776](https://doi.org/10.1109/ICACCI.2015.7275776).
- [37] A. A. Taha and A. Hanbury, "Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool," *BMC Med. Imag.*, vol. 15, no. 1, p. 29, Aug. 2015, doi: [10.1186/s12880-015-0068-x](https://doi.org/10.1186/s12880-015-0068-x).
- [38] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *J. Big Data*, vol. 6, no. 1, p. 60, Jul. 2019, doi: [10.1186/s40537-019-0197-0](https://doi.org/10.1186/s40537-019-0197-0).



**CRISTIANA TIAGO** (Member, IEEE) was born in Portugal, in 1995. She graduated in biomedical engineering and biophysics from the University of Lisbon, in 2016. She received the master's degree in biomedical engineering and biophysics from the University of Lisbon, in 2018; with a specialization on clinical engineering and medical instrumentation after spending one year at the Eindhoven University of Technology, closely with Philips Research, working on deep learning algorithms for cardiac magnetic resonance images analysis. She joined GE Healthcare and the MARCIUS Marie Curie Ph.D. Programme, in March 2020, as an ESR 4.



**ANDREW GILBERT** was born in Utah, USA, in 1994. He received the B.S. degree in electrical engineering from the University of Utah, in 2016, the M.S. degree in electrical engineering from Stanford University, in 2018, and the Ph.D. degree from the University of Oslo, working on automating echocardiography analysis through deep learning, in 2021.

He is currently a Data Scientist with the Cardiovascular Ultrasound Research and Development Team, GE Healthcare, working to develop machine learning solutions for robust extraction of anatomical and pathophysiological features from echocardiography images. He has worked on multiple automated measurement and classification systems deployed in GE's line of Vivid ultrasound scanners.



**AHMED SALEM BEELA** was born in Egypt, in 1986. He graduated from the Faculty of Medicine, Suez Canal University, Egypt, in 2008. He received the master's degree in cardiovascular diseases from Suez Canal University, in 2014, and the Diploma of Advanced Master's degree in medical imaging from the University of Leuven, Leuven, Belgium, in 2017.

He was a Clinical Research Fellow with the Medical Imaging Research Center (MIRC), Leuven, Belgium, from 2016 to 2018. He is a Non-Invasive Cardiologist currently enrolled in the MARCIUS Marie Curie Ph.D. Programme focusing on preparing and curate the needed clinical database for training artificial learning algorithms in addition to validating the developed tools.



**SVEIN ARNE AASE** was born in Oslo, Norway, in 1978. He received the M.Sc. degree in computer science from the Norwegian University of Science and Technology, Trondheim, Norway, in 2003, and the Ph.D. degree from the Department of Circulation and Medical Imaging, Faculty of Medicine, Norwegian University of Science and Technology, in 2008.

He spent a period as a Postdoctoral Researcher with the Department of Circulation and Medical Imaging, Norwegian University of Science and Technology, and was employed by GE Vingmed Ultrasound, Horten, Norway, in 2010. Since 2017, his main fields of interests has been deep learning applications in echocardiography.



**STEN ROAR SNARE** was born in Kongsvinger, Norway, in 1981. He received the M.Sc. degree in engineering cybernetics and the Ph.D. degree in medical technology from the Norwegian University of Science and Technology (NTNU).

He worked with the Department of Circulation and Medical Imaging, NTNU, until Fall 2011, when he was hired at GE. His research interests include medical ultrasound technology, in particular image analysis for automated quantification of physiological parameters from ultrasound images.



**JURICA ŠPREM** was born in Zagreb, Croatia, in 1989. He received the B.Sc. degree in computing from the Faculty of Electrical Engineering and Computing, University of Zagreb, Croatia, in 2012, the master's degree in ICT with a focus on signal processing from the University of Zagreb, in 2014, and the Ph.D. degree from UMC Utrecht, in 2019, on the topic of enhanced cardiovascular risk prediction by machine learning.

He joined GE Healthcare, in 2019, as an AI Tech Lead, with a focus on combining AI with cardiac ultrasound where he continues to support research activities.



**KRISTIN MCLEOD** was born in New Zealand, in 1986. She received the B.A. degree in mathematics and psychology and the B.Sc. (Hons.) and M.Sc. degrees in mathematics from Massey University, New Zealand, in 2008, 2009 and 2010, respectively, and the Ph.D. degree in artificial intelligence for medical image analysis from the Université Côte d'Azur, France, in 2013.

She continued her academic career at Simula Research Laboratory, Norway, as a Postdoctoral Research Fellow, from 2013 to 2017, focusing on the development of automatic tools for supporting diagnosis and therapy planning for cardiac diseases through the use of artificial intelligence. In 2017, she joined the GE Vingmed Ultrasound Department, GE Healthcare, Norway, as a Data Scientist, before transitioning to the role of Digital Manager. Through the role of Digital Manager, she has been working at the intersection of research and development, playing the role of the coordinator and principal investigator for national and EU research programs, as well as the strategic lead of the development of AI features and app ecosystem initiatives at GE Vingmed Ultrasound.

...