Generation of realistic simulated B-mode image texture with a GAN

Nolann Lainé, Guillaume Zahnd, Olivier Bernard, Maciej Orkisz, Herve Liebgott

Generation of realistic simulated B-mode image texture with a GAN

1st Nolann Lainé⋆ 2nd Guillaume Zahnd† 3rd Olivier Bernard⋆ 4th Maciej Orkisz⋆ 5th Hervé Liebgott⋆

⋆*Univ Lyon, Université Lyon 1, INSA-Lyon, CNRS, Inserm, CREATIS UMR 5220, U1294, F-69621, Lyon, France*
†*Institute of Biological and Medical Imaging, Helmholtz Zentrum München, Neuherberg, Germany*

***Abstract*—The intima-media complex of the common carotid artery is considered the sentinel of a silent killer disease called atherosclerosis. Morphological biomarkers such as the intima-media thickness are already exploitable, but dynamic biomarkers, which reflect tissue deformation over the cardiac cycle, remain to be validated. Recent motion estimation methods seek to quantify compression, shear, and elongation coefficients, but their clinical applicability has not yet been well defined, and their actual accuracy is difficult to assess due to the absence of ground truth. This lack of reference also is the main limitation to explore fully supervised deep learning methods that have shown great potential in other applications. With this in mind, we propose a simulation pipeline to produce realistic *in silico* sequences, by combining a physics-based simulator with a post-processing based on a generative adversarial network.**

***Index Terms*—Ultrasound simulation, carotid artery, deep learning, generative adversarial network, domain adaptation.**

I. Introduction

Cardiovascular diseases—particularly atherosclerosis, known as the "silent killer"—are considered the leading cause of death worldwide [7]. We seek to detect early manifestations of the disease by extracting biomarkers from the common carotid artery (CCA), considered the sentinel of atherosclerosis [16]. In particular, we aim to quantify the displacement field of the CCA far wall in ultrasound (US) image sequences and to deduce thereof its kinematic behavior through a cardiac cycle, in order to predict if a patient is ill. Nevertheless, obtaining reliable references to validate motion tracking algorithms remains difficult, as ground truth information is unavailable. Moreover, newly developed algorithms based on fully-supervised deep learning generally perform better than traditional methods in other applications [6], [14], but the paucity of said ground-truth is an obstacle to the generation of an adequate training set. Therefore, we propose a simulation pipeline that aims to produce realistic *in silico* sequences, along with associated deformation parameters that can serve as fully-determined references.

Most methods to simulate US images of arteries [8], [11] consist of physics-based ultrasound simulators such as Field II [4], [5] or SIMUS [3]. These methods take as input *in silico* phantoms, which are typically generated using specific statistical distributions of reflection coefficients. Alternatively, *in silico* phantoms have also been simulated from *in vivo* images [1], [13]. The latter solution is attractive to generate realistic samples because it directly uses a genuine *in vivo* image as the source, but its full potential remains locked by the incomplete knowledge of the modeled tissues and of the image-acquisition system parameters. In this work, we present a simulation pipeline devised to counteract these limitations via the inclusion of a post-processing operation based on a generative adversarial network (GAN, [12]).

II. Initial simulation phase

A classical simulation pipeline takes an *in vivo* clinical static image of the CCA as input and the output is a temporal sequence of synthetic images. We first construct an *in silico* phantom (part II-A), which is fed into a realistic US simulator (in our case, Field II) along with probe and acquisition parameters (part II-B). These steps are detailed below.

A. Phantom construction

In order to calculate the amplitudes of the reflection coefficients of the scatteres in the *in silico* phantom from an *in vivo* B-mode image I, we first inverse the log-compression and introduce variability through Gaussian distribution (see [13]). Reflection coefficients RC_m, $m = 1 \dots M$ thus calculated are then assigned to M points at locations (x_m, y_m) drawn normally at random:

$$RC_m = \left(\frac{I\left(x_m, y_m\right)}{\max(I)} \right)^{\frac{1}{\gamma}} \cdot \mathcal{N}\left(0, 1\right). \tag{1}$$

Here γ is the gamma-compression constant (set to 0.3), and $\mathcal{N}\left(0, 1\right)$ is the normal distribution. The scatterer map is

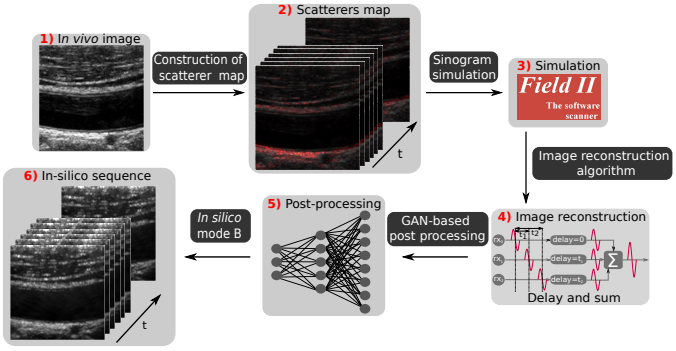

Figure 1. Simulation pipeline with a GAN-based post-processing (step 5).

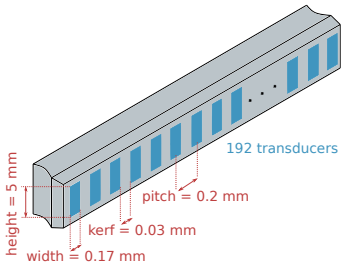

Figure 2. Linear probe used in our simulation.

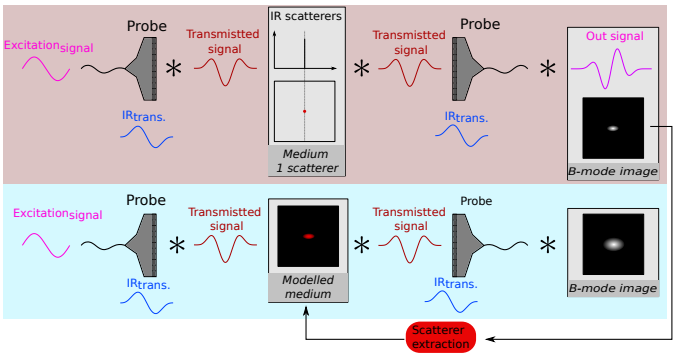

Figure 3. Expansion of the PSF function. Consider the upper part as the *in vivo* system. For ease of explanation, the medium is represented as a single point scatterer. The resulting B-mode image is therefore the PSF. The simulation (bottom) does not take a single point scatterer as above (because we don't know where it really is and what its reflection coefficient is), but a cloud of scatterers with coefficients calculated from the diffuse spot representing the PSF in the B-mode image, which amounts to adding the double convolution of the *in vivo* system to the double convolution simulated.

constructed with a density of 10 per square wavelength, so $M \simeq 300000$ for one image in our case.

B. Modeling of the system

When using clinical images, it is difficult to know all the image-acquisition and reconstruction parameters, particularly for images from open-access multi-center databases such as those used in our work: CUBS 1 [10][1] and CUBS 2 [9][2]. Therefore, we had to design our probe model and specify empirical simulation parameters. We have selected a linear probe *L12-3v* from MUST [3], consistent with our application, on which we just changed the number of elements from 128 to 192 (Fig. 2). The acquisition parameters are summarized in Table I. The *in vivo* images were acquired with scanline-based beamforming. The sonographer therefore chose a depth of focus located around the IMC of the far wall. For each acquisition, the depth varies (in meters) and we should adjust it for each simulation. To ensure the focus at the IMC, we decided to use a synthetic aperture, thus all depths are at focus in both transmission and reception. This choice may lead to a slight lack of realism away from the IMC.

Table I
ACQUISITION PARAMETERS.

Parameters	Settings
Excitation signal ($pulse$)	Sinusoidal (one period)
Probe impulse response (IR)	Hanning window $\times$ $pulse$
Central frequency (f_c)	7.5 MHz
Sampling frequency (f_s)	45 MHz
Insonification mode	Synthetic aperture
F-number	1.5
Apodization in emission	Hanning
Apodization in reception	Hanning

C. Considerations about the current modeling limitations

We noticed four main limitations with the conventionally-used modeling strategy:

1) The compression parameters used to produce the envelope image are unknown. Applying an inverse gamma compression to return to the gray level distribution of the envelope image (Eq. 1) is likely inaccurate.
2) The inverse reconstruction operator used to obtain the envelope image (namely, the delay-and-sum algorithm) only achieves imperfect fidelity when recovering the initial medium. Therefore, the resulting gray level distribution of the pixels is expected to not exactly correspond quantitatively to the underlying distribution of reflection coefficients.
3) The actual spatial coordinates and the spatial density of the scatterers in the *in vivo* medium are unknown. Populating the scatterer distribution via drawing at random from the normal distribution from an empirical range is an approximation of the model.
4) The *in vivo* images are acquired by physical systems that systematically introduce a "double spread" effect: the point spread function (PSF) of the system is defined by the transmission and the reception properties of the system ($PSF := pulse * IR_{trans} * IR_{trans}$). Therefore, the reflection coefficients extracted from *in vivo* images do not correspond to true values. When subsequently applying an ultrasound simulator onto such scatterer distribution, the PSF of the simulator actuates, for a second time, the response of the physical system, which results in exaggeratedly smoothing the simulated signals (Fig. 3). Note that we also do not know with precision the exact excitation signal and impulse response of the transducer used to acquire *in vivo* images.

The appropriate analytical solution to construct *in silico* phantoms would be to apply a mathematically-rigorous model-based inverse reconstruction operator, or at minima, to utilize a proper deconvolution algorithm to extract the scatterers distribution from the *in vivo* image, however these approaches are still exploratory in the field of ultrasound. An alternative solution may be the use of an artificial intelligence technique named "domain adaptation". The key idea is to train a neural network for *in silico* to *in vivo* image translation. To this

[1]http://dx.doi.org/10.17632/fpv535fss7.1
[2]http://dx.doi.org/10.17632/m7ndn58sv6.1

end, we suggest to use a Generative Adversarial Network (GAN) [12], which is well suited for domain adaptation. Recently, GANs were used to generate synthetic ultrasound images from masks [2], or to mimic the image texture with a physics-based simulator [15]. Nevertheless, to the best of our knowledge, no GAN-based approach was proposed to generate ultrasound image sequences.

III. GAN-based post processing phase

A. Dataset

We used a subset $(n = 769)$ of *in vivo* images from the CUBS 1 [10] and CUBS 2 [9] databases. As the images also contain text and other graphical inlays, we clipped them out and cropped each sample down to only a region of interest (ROI). Then, the scatterer maps were constructed for each image and simulated using the pipeline presented in Figure,1, bypassing step 5 (post-processing). Thus, the collection of *in vivo* images (R) is matched with a corresponding collection of simulated *in silico* images (S). It is noteworthy that no post-processing is applied during the simulation. The GAN takes as input S and learns how to predict R. These pairs of images (S, R) made up the data set, which was split into training (70 %), validation (20 %), and test (10 %) subsets.

B. Network & training

A GAN consists of two different networks during the training process: a generator and a discriminator. The generator performs the image-domain adaptation while the discriminator helps the generator by assessing if the generator output resembles an image from the target domain (here *in-vivo* image). Learning therefore relies on two different loss functions, one for the discriminator (Eq. 2) and one for the generator (Eq. 3).

$$L_{Discr} = -\frac{1}{2}\left(\log\left(1 - \hat{y}_{\tilde{R}}\right) + \log\left(\hat{y}_R\right)\right), \tag{2}$$

$$L_{Gen} = -\lambda_{discr}\log\left(\hat{y}_{\tilde{R}}\right) + \lambda_{pixel}||R - \tilde{R}||_2, \tag{3}$$

where R is the *in vivo* image, $\tilde{R}$ is the GAN output, $\hat{y}_R$ and $\hat{y}_{\tilde{R}}$ are the probabilities returned by the discriminator that R and $\tilde{R}$ respectively are real images, while λ_{discr} and λ_{pixel} are weighting factors set to 10^{-3} and 1, respectively.
During training, we used a batch size of 6. Concerning the pre-processing, the ROIs were normalized between $[0, 255]$ and resized to 256×512 pixels. We evaluated three different architectures for the generator (namely, U-net, dilated U-net, and ResNet-based super-resolution GAN).

IV. Results

Peak signal-to-noise ratio (PSNR, Eq. 4) was used to measure the similarity between images I_1 and I_2 (S *vs.* R or S *vs.* $\tilde{R}$):

$$PSNR(I_1, I_2)_{(dB)} = 10 \cdot \log_{10}\left(\frac{255^2}{||I_1 - I_2||_2}\right), \tag{4}$$

where 255 is the maximal gray level of the pixels. Quantitative results are reported in Table II, while Figure 4 qualitatively illustrates example results obtained with the best architecture.

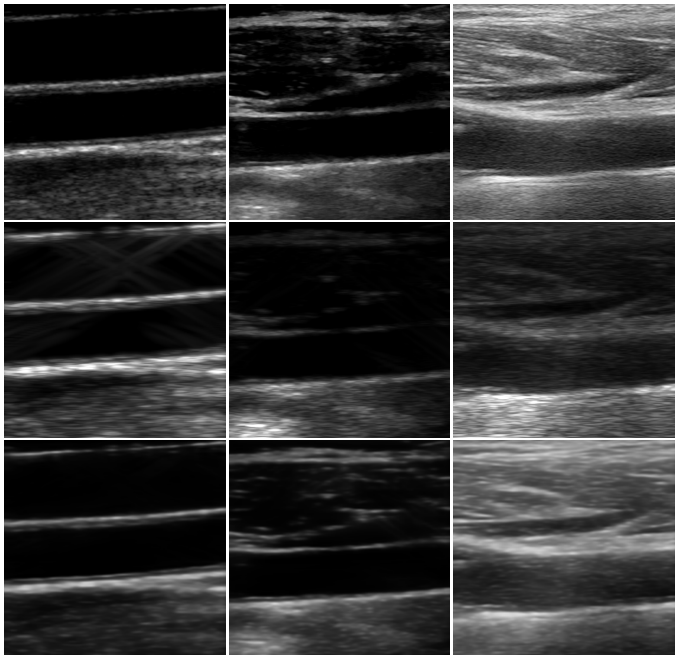

Figure 4. Example images. Top: *in vivo* US common carotid images (R). Middle: corresponding *in silico* images (S), simulated with Field II. Bottom: corresponding *in silico* images ($\tilde{R}$), post-processed by the GAN.

V. Discussion

The networks, and particularly the dilated U-net, help to provide *in silico* images that are more similar to the source clinical images compared to the intermediate simulation result (Fig. 4). In the initial simulated images, we observed that the proximal wall often lacks luminance, which likely means that the simulation fails to reproduce the presence—or lack thereof—of time gain compensation. Conversely, the GAN succeeds in correcting this axial profile, which explains the gain in dB. Nevertheless, current GAN results do not seem to improve the image texture, as the GAN output appears smoother than both simulated and *in vivo* images. Moreover, when applying the GAN to an *in silico* sequence, in which the scatterers obeyed a controlled displacement field, we observed a jitter in the movie appearance.

There are several avenues to explore to improve our method. First, the proposed GAN is not scanner-specific and has learned to average the features from different scanners. Conditional GAN architectures could be a solution to take scanner information into account. Second, temporal information was not explored. A cycle-GAN architecture fed with a 2D + t volume could learn the spatio-temporal filtering and temporal coherence and likely eliminate jitter effects. Last but not least, we noticed a potential issue in the pre-processing step: The ROIs extracted from the *in vivo* images have different initial sizes, but are reshaped to a fixed size. Therefore, the appearance (speckle size and radial-to-axial proportions) of the texture is altered, and the network is forced to operate under

Table II
PSNR ON EACH SUBSET FOR THREE DIFFERENT ARCHITECTURES. THIRD COLUMN: PSNR BETWEEN THE *in silico* IMAGE WITHOUT POST-PROCESSING. FOURTH COLUMN: PSNR BETWEEN IN VIVO IMAGE AND THE IN SILICO IMAGE POST-PROCESSED BY THE GAN.

Model	Subset	PSNR original *vs.* simulated (dB)	PSNR original *vs.* GAN output (dB)
U-net	training	15.5 ± 3.5	24.7 ± 1.6
	validation	15.6 ± 3.4	22.2 ± 2.1
	test	15.5 ± 3.1	22.3 ± 1.8
Dilated U-net	training	15.5 ± 3.5	$\mathbf{25.7 \pm 1.8}$
	validation	15.6 ± 3.4	$\mathbf{23.4 \pm 2.1}$
	test	15.5 ± 3.1	$\mathbf{23.3 \pm 2.0}$
Super-resolution GAN (ResNet-based)	training	15.5 ± 3.5	17.1 ± 2.9
	validation	15.6 ± 3.4	16.4 ± 3.6
	test	15.5 ± 3.1	16.2 ± 3.2

different receptive fields (in meter). Since we aim to mimic the texture, the spatial information captured by the network is essential, and a pre-processing step to transform the database into isotropic and homogeneous data could be beneficial.

VI. ACKNOWLEDGMENTS

This work was partly supported, via NL's doctoral grant, by the LABEX PRIMES (ANR-11-LABX-0063) of Université de Lyon, within the program "Investissements d'Avenir" (ANR-11-IDEX-0007) operated by the French National Research Agency (ANR). The authors have no relevant financial or non-financial interests to disclose.

VII. COMPLIANCE WITH ETHICAL STANDARDS INFORMATION

The data from human subjects used in this work were obtained and treated in line with the principles of the Declaration of Helsinki. Approval was granted by the Ethics Committees of the institutions involved in creating the multicentric database, from which these data were accessed.

REFERENCES

[1] M. Alessandrini, M. De Craene, O. Bernard, S. Giffard-Roisin, P. Allain, I. Waechter-Stehle, J. Weese, E. Saloux, H. Delingette, M. Sermesant *et al.*, "A pipeline for the generation of realistic 3D synthetic echocardiographic sequences: Methodology and open-access database," *IEEE Transactions on Medical Imaging*, vol. 34, no. 7, pp. 1436–1451, 2015.

[2] N. J. Cronin, T. Finni, and O. Seynnes, "Using deep learning to generate synthetic B-mode musculoskeletal ultrasound images," *Computer Methods and Programs in Biomedicine*, vol. 196, p. 105583, 2020.

[3] D. Garcia, "Make the most of MUST, an open-source Matlab UltraSound Toolbox," in *2021 IEEE International Ultrasonics Symposium (IUS)*. IEEE, 2021, pp. 1–4.

[4] J. A. Jensen, "Field: A program for simulating ultrasound systems," in *10th Nordicbaltic Conference on Biomedical Imaging, Medical & Biological Engineering & Computing, vol. 4, supplement 1, part 1*, 1996, pp. 351–353.

[5] J. A. Jensen and N. B. Svendsen, "Calculation of pressure fields from arbitrarily shaped, apodized, and excited ultrasound transducers," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 39, no. 2, pp. 262–267, 1992.

[6] S. Jiang, D. Campbell, Y. Lu, H. Li, and R. Hartley, "Learning to estimate hidden motions with global motion aggregation," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 9772–9781.

[7] S. Kaptoge, L. Pennells, D. De Bacquer, M. Cooney, M. Kavousi, G. Stevens *et al.*, "World Health Organization cardiovascular disease risk charts: revised models to estimate risk in 21 global regions," *Lancet Global Health*, vol. 7, no. 10, pp. e1332–e1345, 2019.

[8] M. Makūnaitė, R. Jurkonis, A. Lukoševičius, and M. Baranauskas, "Simulation of ultrasound rf signals backscattered from a 3D model of pulsating artery surrounded by tissue," *Diagnostics*, vol. 12, no. 2, p. 232, 2022.

[9] K. Meiburger, F. Marzola, G. Zahnd, F. Faita, C. P. Loizou, N. Lainé, C. Carvalho *et al.*, "Carotid ultrasound boundary study (CUBS): Technical considerations on an open multi-center analysis of computerized measurement systems for intima-media thickness measurement," *Computers in Biology & Medicine*, 2022.

[10] K. Meiburger, G. Zahnd, F. Faita, C. P. Loizou, C. Carvalho, D. A. Steinman, L. Gibello *et al.*, "Carotid ultrasound boundary study (CUBS): An open multicenter analysis of computerized intima–media thickness measurement systems and their clinical impact," *Ultrasound in Medicine & Biology*, vol. 47, no. 8, pp. 2442–2445, 2021.

[11] K. M. Meiburger, F. Molinari, J. Wong, L. Aguilar, D. Gallo, D. A. Steinman, and U. Morbiducci, "Validation of the carotid intima–media thickness variability: Can manual segmentations be trusted as ground truth?" *Ultrasound in Medicine & Biology*, vol. 42, no. 7, pp. 1598–1611, 2016.

[12] M. Mirza and S. Osindero, "Conditional generative adversarial nets," *arXiv preprint arXiv:1411.1784*, 2014.

[13] Y. Sun, F. Vixege, K. Faraz, S. Mendez, F. Nicoud, D. Garcia, and O. Bernard, "A pipeline for the generation of synthetic cardiac color Doppler," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 69, no. 3, pp. 932–941, 2022.

[14] Z. Teed and J. Deng, "Raft: Recurrent all-pairs field transforms for optical flow," in *European Conference on Computer Vision*. Springer, 2020, pp. 402–419.

[15] F. Tom and D. Sheet, "Simulating patho-realistic ultrasound images using deep generative networks with adversarial learning," in *IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*. IEEE, 2018, pp. 1174–1177.

[16] F. Yousefi Rizi, J. Au, H. Yli-Ollila, S. Golemati, M. Makūnaitė, M. Orkisz *et al.*, "Carotid wall longitudinal motion in ultrasound imaging: An expert consensus review," *Ultrasound in Medicine & Biology*, vol. 46, no. 10, pp. 2605–2624, October 2020.