Ultrasound denoising using the pix2pix GAN

Afonso Raposo[1]
afonso.raposo@tecnico.ulisboa.pt
António Azeitona[1]
antoniorrazeitona@tecnico.ulisboa.pt
Manya Afonso[2]
manya.afonso@wur.nl
J. Miguel Sanches[1]
jmrs@tecnico.ulisboa.pt

[1] Institute for Systems and Robotics (ISR), LARSyS,
Instituto Superior Técnico,
Departamento de Bioengenharia,
Universidade de Lisboa
[2] Wageningen University and Research,
Wageningen, The Netherlands

Abstract

The use of ultrasound (US) as an imaging technique is essential for the diagnosis of atherosclerotic cardiovascular disease (ASCVD), which depends on US images of the carotid artery. However, US images are plagued by a specific type of noise called Speckle noise, which lowers image quality dramatically. As an attempt to improve US image quality, the use of a generative adversarial network (GAN) is explored. The GAN chosen for this is the pix2pix model and the dataset used for training is composed of images containing simple geometric shapes of various scales and their equivalent corrupted with Speckle noise following the Log-Compression model. The results of this GAN are displayed and a noticeable improvement can be verified in the image quality.

1 Introduction

The two main predictors used for the diagnosis and assessment of atherosclerotic cardiovascular disease risk are the carotid intima-media thickness and analysis of the carotid arterial plaque, both of which are obtained by the use of ultrasound (US) imaging [4, 6]. Although there exist some promising studies on the development of a fully automatic segmentation technique, the performance is still far from ideal due to the high content of Speckle noise [5]. The current approach for the denoising of US images is based on the use of non-linear filters such as anisotropic diffusion filters and adaptive median filters [2]. These filters tend to preserve the contours of the structure but over-smooth the remaining areas.

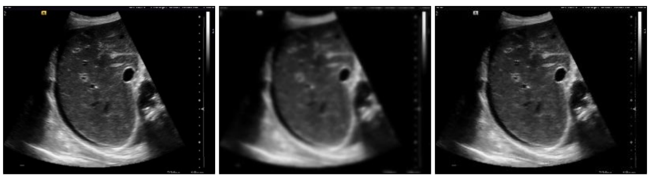

Figure 1: Ultrasound image of a liver (left) and corresponding images resulting from anisotropic diffusion filtering (middle) and adaptive median filtering (right).

The introduction of Generative Adversarial Networks (GANs) as a means to generate images presents a new opportunity for developing novel denoising techniques. The most widespread of these networks is the *pix2pix* [3], which can be trained with pixel-wise paired images in a way that it can receive a certain image as an input and then output a version of that image with different characteristics, image-to-image translation.

2 Problem Formulation

The pix2pix network requires pairs of images to be trained. In this case, these pairs consist of US images with Speckle noise and the same image without Speckle noise.

Speckle noise follows the Rayleigh distribution:

$$\rho(y_i) = \frac{y_i}{\sigma^2} e^{-\frac{y_i^2}{2\sigma^2}} \tag{1}$$

This work was supported by Portuguese funds through FCT (Fundação para a Ciência e Tecnologia) through the projects reference UIDP/50009/2020 and through the reference UID/EEA/50009/2019, LARSyS - FCT Plurianual funding 2020-2023.

Where ρ is the p.d.f., y_i is the intensity value of the ith pixel in the grayscale ultrasound image and σ is a scale factor dependent on the scattering amplitude of the particles in the medium [9].

B-mode US images suffer logarithmic compression after the acquisition of the data, which can be modeled by the equation 2.

$$z_{ij} = \alpha \log(y_{ij} + 1) + \beta \tag{2}$$

Where i and j are the positions of the pixel, z is the pixel after the compression, y is the pixel of the radio frequency (RF) image, and α and β are parameters dependent on the contrast and brightness [7], respectively.

These mathematical models make possible the creation of synthetic pairs of images to train the network, which, after trained, will accept US images and return denoised versions of those images.

3 Methods

The dataset used was composed of synthetic images of several geometric shapes of varying dimensions, intensities, and number. These images were generated with the *draw.random shapes* function from the *skimage* library, using the parameters: shape=(256, 256), allow_overlap=True, min_shapes=128, max_shapes=256, min_size=10, max_size=50. The intensity of the images was then inverted, so the background was black and the pixels with the lowest intensities (<5) were corrected to have an intensity of 5. A total of 2560 images was generated this way, corresponding to the output of the training dataset (denoised US images).

The training input images (noisy US images), were computed based on the images obtained using the method described above. To simulate the US image with Speckle noise, the Rayleigh distribution function (equation 1) was used, taking the original synthetic image as the value for the standard deviation (here denoted by σ). However, the model for logarithmic compression as displayed in figure 2, shows that after the noisy data is obtained (RF image), there are a few steps to reach the final B-mode US image.

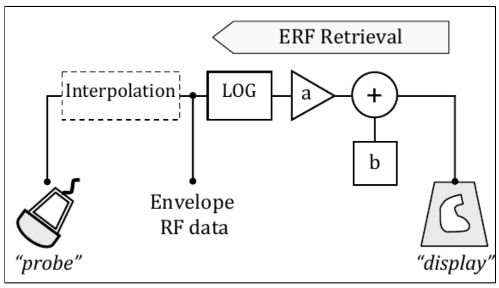

Figure 2: Full diagram of the model for the generic processing operations of an ultrasound system [8].

The first of these steps is an interpolation, which was mimicked by applying a 2D decimation of the images, reducing them to a quarter of their size, followed by applying a linear interpolation, restoring the dimension of the images.

The other step is the logarithmic compression expressed by equation 2. This process depends on two parameters, α and β, that are usually not provided by US equipment manufacturers and, therefore, are unknown. As a way to increase the versatility and robustness of the network, these parameters were randomly selected for each image from a set of intervals: $[10, 50]$ for α and $[-50, 50]$ for β.

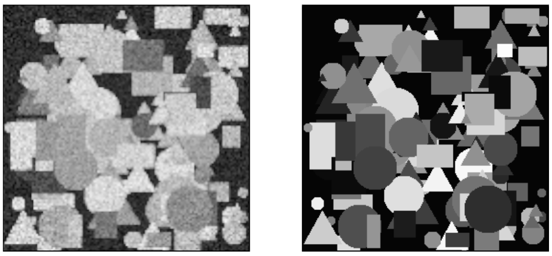

Figure 3: Example of the image pairs resulting from the method used: the input image (left) and the target (right).

The training of the network consisted of using 2048 image pairs for training and 512 image pairs for validation, for a total of 200 epochs.

4 Results and Discussion

After the network was trained, some images from the validation set were fed to it so that the image generated could be compared to the target output, as shown in figure 4.

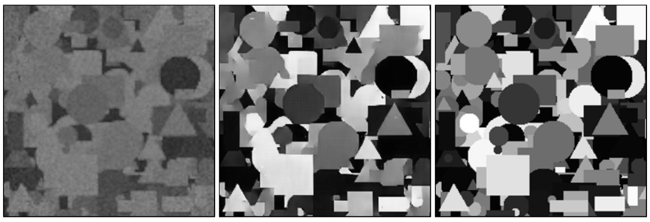

Figure 4: Example of a simulated noisy US image from the validation set (left), the output given by the trained network (middle) and the original synthetic image (right).

Although some distortion is present, the structural information recovered seems to be more than satisfactory and the values of the Structural Similarity Index (SSIM, higher is better) and Peak Signal-to-Noise Ratio (PSNR, higher is better) as quality assessment metrics were calculated [1] and the values are shown in table 1.

Table 1: Values of peak signal-to-noise ratio and structural similarity index corresponding to different denoising techniques

Method	PSNR	SSIM
Anisotropic Filter	11.906	0.461
Adaptive Median	12.053	0.452
Ours	21.085	0.789

The drastic difference in the values can be explained easily when comparing the images (figure 6), seeing as the classical filters do not improve the intensity values of the image in the same way that the network was able to.

Nonetheless, both the SSIM and PSNR, testify to the potential of the network when compared to classical methods of filtering, showing almost double the score.

The finished network was also used to denoise real US images of the carotid artery, resulting in the images shown in figure 5.

In this case, quality assessment measures cannot be performed because, being a real US image, there is no ground truth image (a completely clean image) available. Even so, visual comparison with the aforementioned classical methods is possible (figure 6).

As it can be seen, the image resulting from the GAN highlights the more prominent structures, preserves most contours without retaining noise while, admittedly, losing some of the "realness" of the image as a default US image since the model was trained with geometric synthetic images.

5 Conclusion

The use of the pix2pix network as a tool to denoise and enhance the image quality of US images is an appealing prospect and, as shown in this work, reveals itself promising in this area. In this work, synthetic images were

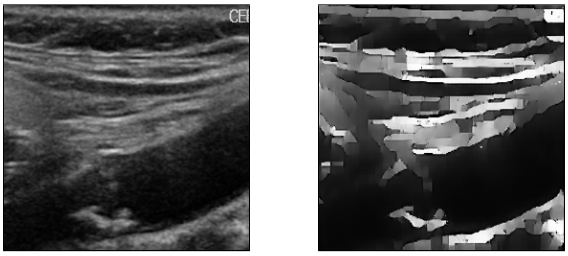

Figure 5: Example of a real carotid US image from (left) and the output given by the trained network (right).

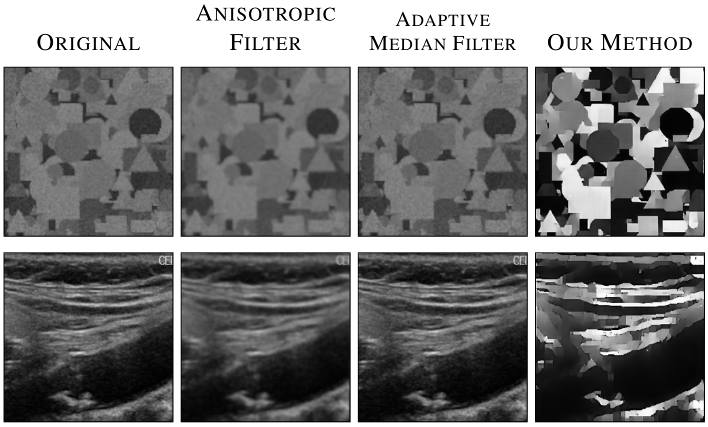

Figure 6: Comparing various ultrasound denoising algorithms to our method.

used to train the network and, although the influence of the geometric shapes is clear on the resulting carotid US image, the increased sharpness and preservation of contours holds great value for both physicians and automatic segmentation algorithms. There is room for improvement still, specifically in the areas of the architecture of the network, adjustments to the network's loss function, and refinement of the datasets used for training, making the use of GANs as a denoising tool an enticing avenue for further study.

References

[1] Li Sze Chow and Raveendran Paramesran. Review of medical image quality assessment. *Biomedical Signal Processing and Control*, 27:145–154, May 2016. doi: 10.1016/j.bspc.2016.02.006.

[2] Linwei Fan, Fan Zhang, Hui Fan, and Caiming Zhang. Brief review of image denoising techniques. *Visual Computing for Industry, Biomedicine, and Art*, 2 (1), July 2019. doi: 10.1186/s42492-019-0016-7.

[3] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros. Image-to-image translation with conditional adversarial networks, 2018.

[4] Amer M. Johri, Vijay Nambi, Tasneem Z. Naqvi, Steven B. Feinstein, Esther S.H. Kim, Margaret M. Park, Harald Becher, and Henrik Sillesen. Recommendations for the assessment of carotid arterial plaque by ultrasound for the characterization of atherosclerosis and evaluation of cardiovascular risk: From the american society of echocardiography. *Journal of the American Society of Echocardiography*, 33(8):917–933, 2020. doi: https://doi.org/10.1016/j.echo.2020.04.021.

[5] P Krishna Kumar, Tadashi Araki, Jeny Rajan, John R Laird, Andrew Nicolaides, and Jasjit S. Suri. State-of-the-art review on automated lumen and adventitial border delineation and its measurements in carotid ultrasound. *Computer Methods and Programs in Biomedicine*, 163:155–168, September 2018. doi: 10.1016/j.cmpb.2018.05.015.

[6] Joseph F. Polak and Daniel H. O'Leary. Carotid intima-media thickness as surrogate for and predictor of CVD. *Global Heart*, 11(3):295, September 2016. doi: 10.1016/j.gheart.2016.08.006.

[7] Jose Seabra and Joao Sanches. Modeling log-compressed ultrasound images for radio frequency signal recovery. In *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, August 2008. doi: 10.1109/iembs.2008.4649181.

[8] José Carlos Rosa Seabra. *Medical Ultrasound B-Mode Modeling, Despeckling and Tissue Characterization Assessing the Atherosclerotic Disease*. PhD thesis, Instituto Superior Técnico, 2011.

[9] R.F. Wagner, S.W. Smith, J.M. Sandrik, and H. Lopez. Statistics of speckle in ultrasound b-scans. *IEEE Transactions on Sonics and Ultrasonics*, 30(3): 156–163, May 1983. doi: 10.1109/t-su.1983.31404.