An image-to-image translation approach for speckle noise reduction in US

António Rafael Ramiro Azeitona
antoniorrazeitona@tecnico.ulisboa.pt

Instituto Superior Técnico, Lisboa, Portugal

November 2021

Abstract

Ultrasound images are plagued by a characteristic type of noise named speckle noise. Although filtering techniques that reduce this type of noise exist, they result in low quality images. A method that has garnered attention recently is the use of generative adversarial networks (GAN) to create images from other images (image-to-image translation). The most popular of these networks is the pix2pix network, which can be trained with image pairs of the same ground truth image and learn to transform one image into the other. This process requires a large number of images matched pixel to pixel. Two pix2pix networks were trained. The first network was trained to transform phantoms into ultrasound images so that a paired dataset of MRI-Ultrasound images could be created. The second network was trained to transform ultrasound images into MRI-like versions. Several real and simulated ultrasound images were processed by this denoising network and quality assessment metrics were extracted from the results, as well as compared to those obtained from classical filtering methods. All of the tests point to this work as an absolute success, with possibilities of being applied to other imaging modalities and further improved.
Keywords: Ultrasound, GAN, pix2pix, Denoising, Speckle, Image-to-Image Translation

1. Introduction

To better understand this work, some concepts should be introduced.

1.1. Ultrasound

The technique that is most relevant to this work is ultrasound (US), although the work can be adapted to other imaging modalities. The use of ultrasound as an imaging technique is widespread and considered a basic necessity for any hospital. Important factors for this are its low cost, unparalleled safety rapport and simplicity of use. One area that benefits greatly from the use of ultrasound is the diagnosis of cardiovascular diseases (CVD), more specifically, the diagnosis of atherosclerotic cardiovascular disease. The two main predictors used for the diagnosis and assessment of atherosclerotic CVD risk are the carotid intima-media thickness and the analysis of the carotid arterial plaque. Both of these metrics rely on the use of US imaging, reinforcing the importance of improving the quality of the resulting images [1].

1.2. Neural Networks and Images

Artificial neural networks (ANN) started as a model for a real neural network, a system of connected neurons that receive stimuli and signal the neurons on the next layer accordingly. Convolutional Neural Networks (CNN) have been shown to be extraordinarily well suited for image processing. The key lies in the convolutional layers of these networks that work by feeding only the convolution of certain windows of information with certain kernels to the next layer, allowing for fewer weights to be calculated without a loss of performance[2]. CNNs are very promising and have shown incredible usefulness in medical image analysis, opening the door to other types of networks that also work magnificently well with image processing. One of these is the Generative Adversarial Network (GAN), which is composed of two different networks, a generator network and a discriminator network. The basic idea is that the generator network will learn to fool the discriminator network, and the discriminator network will learn to distinguish real data from fake data. In this case, creating images and determining their authenticity. These networks are known to incorporate aspects of CNNs into their architecture [3]. In this manner, the generator network learns how to produce realistic images based on the error of the discriminator. In these models, only the discriminator has access to the real images, forcing the generator to learn solely on the interaction between

the two [3]. It can be thought of as the discriminator learning an adequate error function, and, based on this error, the generator will learn to minimize it [4]. In essence, this means that a GAN can be trained by using a set of paired images and learn to generate images of one kind based on images of another kind. As a matter of fact, this has already been done successfully with the pix2pix network [4].

1.3. Objectives

The goal of the work done here is the creation of a pipeline that, given US images, can produce images that are closer to MRI images, which lack speckle noise and possess increased sharpness. This pipeline is built with the use of two GANs. The first GAN is trained to produce US counterparts of simplified phantoms and is referred to as the US simulator GAN. This will allow for an increased sample of US images, the generation of close approximations of image pairs obtained through real medical images, and an additional tool to simulate the noise associated with US images. The second GAN will be trained with medical and real-life objects so that the input will be an US image and the output an MRI counterpart of that same image This network will be referred to as Denoising GAN. Much of the work done was centered around the search for the correct datasets to use, the image processing required and the architecture of the overall pipeline. The value of a GAN such as this comes, not only from the improvement of the image quality, but also from an improvement in future automatic and semi-automatic segmentation of structures. The datasets used for the training and the parameters used both in the training and processing of the images are detailed in this document. The weights of both GANs are available to be used directly in the Google Colaboratory code for pix2pix available online[1] [4, 5].

2. Problem Formulation

With the basic concepts needed for the understanding of this work detailed above, a more technical view of the work is presented.

2.1. Speckle Noise

Speckle noise has the characteristic of being a type of multiplicative noise. This means that it follows the relation expressed in Equation. 1,

$$g_{i,j} = f_{i,j} \cdot u_{i,j}, \tag{1}$$

where $g_{i,j}$ is the observed image, $f_{i,j}$ is the real image and $u_{i,j}$ is the multiplicative component of speckle noise while i and j are the indices of the pixel [6].

[1]Cycle GAN - Colaboratory URL: https://colab.research.google.com/github/junyanz/pytorch-CycleGAN-and-pix2pix/blob/master/CycleGAN.ipynb (Visited 02/09/2021)

The relationship shown in Equation. 1 is only applicable in the radiofrequency (RF) domain, while most images are in brightness mode (B-Mode), which involves a process of log-compression, as well as other unknown processing steps. This process is modelled in Equation. 2 [7].

$$z_{i,j} = \alpha \cdot log(y_{i,j} + 1) + \beta, \tag{2}$$

where i and j are the indices of the pixel, $z_{i,j}$ is the pixel in the B-Mode image, $y_{i,j}$ is the pixel in the RF image while α and β are variable parameters dependent on the contrast and brightness of the image during acquisition, in this order.

This same model can be inverted, as shown in Equation. 3.

$$y_{i,j} = e^{(\frac{z_{i,j}-\beta}{\alpha})} - 1 \tag{3}$$

The statistical analysis of speckle noise reveals that it can follow either a Rayleigh or Nakagami probability distribution function (PDF)[8, 9].

The Rayleigh PDF is mathematically written as presented in Equation. 4.

$$\rho(y_{i,j}) = \frac{y_{i,j}}{\sigma^2} e^{-\frac{y_{i,j}^2}{2\sigma^2}}, \tag{4}$$

where ρ is the probability density function, $y_{i,j}$ is the intensity value of the pixel at i and j indices in the grayscale RF ultrasound image and σ is a scale factor dependent on the scattering amplitude of the medium.

The equation that describes the Nakagami PDF is shown below as Equation. 5

$$\rho_{\eta_{i,j}}(\eta_{i,j}) = \frac{2L^L}{\Gamma(L)(2\sigma^2)^L} \eta_{i,j}^{2L-1} e^{\left(\frac{-L\eta_{i,j}^2}{2\sigma^2}\right)}, \eta_{i,j} > 0, \tag{5}$$

where $\eta_{i,j}$ is the pixel intensity at position i and j, $\Gamma()$ is the gamma function, L is the Nakagami shape parameter and $2\sigma^2$ is a scaling parameter [10].

One other interesting tool used was the Field II program, that simulates all the elements of the probe and allows for the calculation of the impulse response, culminating in a powerful tool for the simulation of realistic ultrasound images [11, 12].

2.2. Denoising State of the Art

The most common denoising methods for US images available at the present are spatial domain filters.

Anisotropic Diffusion Filter

Some of these filters are based on anisotropic diffusion, and are known as anisotropic diffusion filters.

They can be modelled by the differential equation:

$$\begin{cases} \frac{dI}{dt} = div(D \cdot \nabla I) = D \cdot \nabla^2 I, \\ I(i,j,0) = g(i,j), \end{cases} \tag{6}$$

where $g(i,j)$ is the pixel in the original noisy image at indices i and j, the ∇ is the Laplacian operator and t is the artificial time for this "diffusion" of the image, analogous to the number of iterations. Solving the equation produces $I(i,j,t)$, which is the pixel at indices i and j in the filtered image at a specific moment of the artificial time. The functions for D are as such [13]:

$$d(|\nabla I|) = e^{-\left(\frac{|\nabla I|}{k}\right)^2} \tag{7}$$

This first equation (7) results in a D with a very low value in areas with high contrast differences.

$$d(|\nabla I|) = \frac{1}{1+\left(\frac{|\nabla I|}{k}\right)^2} \tag{8}$$

In both equations, $D = (|\nabla I|)$ and k is a factor that adjusts sensitivity to edges. The use of an anisotropic diffusion filter for speckle noise removal is commonplace, earning it the name of Speckle Reducing Anisotropic Diffusion Filter (SRAD).

Adaptive Median Filter

Another type of filter that has proven useful at removing speckle noise is the adaptive median filter, proposed by *Loupas et al,* [14]. A median filter works by analysing a patch of the image through a window, taking the intensity values of that window into consideration and substituting the center pixel with the median. This filter is non-linear and has the advantage of preserving edges. A modified version of this filter is the weighted median filter, where taking the window into account, each position in said window has a weight associated with it. This filter can further be modified to become an adaptive weighted system that makes use of the parameters of variance (σ) and mean (m) obtained locally from the window being analysed. The slope is then adjusted for the window based on the ratio of these two parameters in conjunction with a scaling constant (c) and the distance from the point i,j to the center of the window $K+1, K+1$ (d), as shown in Equation. 9.

$$w_{i,j} = \left[w_{K+1,K+1} - \frac{cd\sigma^2}{m} \right] \tag{9}$$

2.3. Generative Adversarial Networks

The basis for the denoising method presented in this work are Generative Adversarial Networks [15]. The core concept is based on the simultaneous development of 2 neural networks [15].This approach is powerful but it can be directed more precisely by using conditional GANs (cGAN) [16], which have a slightly different assumption than GANs: let x be the data of the sample, $D(x)$ the probability that x belongs to the sample as given by the D network, z is a noise vector and $G(z)$ is the data generated by the G network, as would be the case with regular GANs. But also, give some information about the sample desired, a label for example, together with the input of the generator network G and also provide that label (y) to the discriminator network D. This is modelled by the following equation [16]:

$$\begin{aligned}\min_G \max_D V(D,G) = \mathbb{E}_{x\sim p_{data}(x)}[\log D(x|y)] + \\ \mathbb{E}_{z\sim p_z(z)}[\log(1 - D(G(z|y)|y))]\end{aligned} \tag{10}$$

This leads us to pix2pix [4], the GAN architecture used in this work. Pix2pix is a cGAN which accepts as input two images, a sketch and the object from which the sketch was made. Following the convention above, the sketch would be y and the real object image would be x. As a result the discriminator network has access to the sketch and the image and attempts to discern if the image was made from the sketch by the generator network or if it is part of the sample, while the generator attempts to create images of the real objects from the sketch. This GAN harmonizes methods proven to work in image processing and the first way in which it does this is by adding the $L1$ norm to the error to be minimized. More adaptations based on image processing techniques learned from CNNs are the use of U-net architecture for the generator network, which ferries information across layers of the network, and the use of PatchGAN in the discriminator network, which analyses the image patch by patch, taking into account all the contributions into the final result. Both the generator and discriminator network are made up of modules of the form convolution-BatchNorm-ReLu [4].

2.4. Assessment Methods

The full reference methods used in this work were signal-to-noise ratio (SNR), peak signal-to-noise ratio (PSNR), Structural Similarity Index (SSIM) and the Multiscale Structural Similarity Index (MS-SSIM).

The SNR is given by :

$$SNR = 10\log_{10}\left(\frac{\sigma^2}{\sigma^2_{g-f}}\right), \tag{11}$$

where σ^2 is the variance of the ground truth image and σ^2_{g-f} is the variance of the difference between the ground truth image, g, and the noisy image, f. [17].

The PSNR is given by:

$$PSNR = 10\log_{10}\left(\frac{peakvalue^2}{MSE}\right), \tag{12}$$

where *peakvalue* is the highest value possible for the intensity (taking the value of 255 in this case)

and MSE is the mean square error between the images shown in Equation. 13 [17].

$$MSE = \frac{1}{M}\sum_{j=1}^{M}(g_{i,j} - f_{i,j})^2, \quad (13)$$

where M is the total number of pixels in the image, $g_{i,j}$ is the pixel at positions i and j on the noisy image and $f_{i,j}$ is the pixel in the ground truth image [17].

The SSIM, is calculated as:

$$SSIM(g, f) = \frac{(2\mu_g\mu_f + C_1)(2\sigma_{gf} + C_2)}{(\mu_g^2 + \mu_f^2 + C_1)(\sigma_g^2 + \sigma_f^2 + C_2)}, \quad (14)$$

where σ_g and σ_f are the standard deviation for the image g and f, σ_{gf} is the covariance between the images, $C_1 = (K_1L)^2$ and $C_2 = (K_2L)^2$, where $K_1 = 0.01$, $K_2 = 0.03$ and L is the dynamic range of the image (in this case $L = 255$) [18].

The MS-SSIM score is calculated by Equation. 15:

$$MS\text{-}SSIM(g, f) = \left[\frac{(2\mu_g\mu_f + C_1)}{(\mu_g^2 + \mu_f^2 + C_1)}\right]_M \prod_{j=1}^{M}\left[\frac{(2\sigma_{gf} + C_2)}{(\sigma_g^2 + \sigma_f^2 + C_2)}\right]_j, \quad (15)$$

where g, f, σ_g, σ_f, σ_{gf}, C_1 and C_2 have the same meaning as in Equation. 14, and j represents the number of filter and downsampling procedures done to obtain that image, where $j = 1$ is the original image and $j = M$ is the image after M iterations of low-pass filtering and downsampling [19].

Another way to assess the results from the denoising GAN is based on a more statistical view. With a small manipulation, Equation. 4 can be used to, given the ground truth image ($g_{i,j}$) and the US image (in the RF domain, $f_{i,j}$), extract the speckle noise ($u_{i,j}$). Afterwards, the distribution of the histogram of the speckle noise can be studied.

These methods, however, depend on the existence of a ground truth image, which is an impossibility in the field of medical imaging. One of the most used methods that do not require ground truth images and also boasts a good performance is the Blind/Referenceless Image Spatial QUality Evaluator, also known as BRISQUE. BRISQUE is based on concepts regarding natural images. There are some factors that strain the credibility of BRISQUE when using it on medical images. As such, a control test is made using US and MRI images, where the notion that the BRISQUE score supports the convention that MRI images have better quality than US images is tested. Furthermore, it has also been shown that BRISQUE presents good dataset independence [20].

3. Methods

The high dependency of the ANN on the dataset used for its training is an issue present in all machine learning models and the pix2pix GAN is not an exception. Therefore, the careful choice of the training images is important for the success of the network. In the case of medical images, there is a scarcity of data unlike the other fields.

3.1. US Simulator GAN

This hurdle in machine learning in the medical image processing field highlights the value of the US Simulator GAN. The existence of a method that creates realistic US images using simplistic phantoms as a basis and can also allow for the development of large annotated datasets is very valuable.

Dataset

All of the data used in the training of this US Simulator GAN was from cardiovascular diagnostic images and videos. The data was provided by physicians of the Santa Maria Hospital's department of Cardiology, and was comprised of 1024 US images of carotid arteries with different morphologies, in different probe positions, along different anatomical segments and some even displaying arterial plaques or stenosis. Of these images, 670 were chosen for the dataset after a selection process that excluded images with low quality, several artifacts, and abundance of zones with little information. The images used were grayscale, with intensity normalized to values between 0 and 255 and the resolution is 256x256 pixels.

Preprocessing

The preprocessing itself follows a certain sequence of smaller and simpler steps. As a reminder, the objective is to create a dataset composed of simplistic phantoms of the dataset with corresponding US images. This correspondence must be in the order of pixel-to-pixel. The first step of the preprocess deals with the fact that US images in B-Mode do not represent the full range of information obtained from a US scan. As shown in Equation. 3, the RF image can be obtained from the B-Mode image with relative simplicity. There are, however, two parameters that must be estimated: α and β. According to the work done in this article by *J. Sanches et al.*[21], these parameters can be approximated by the following equations:

$$\hat{\alpha} = \sqrt{\frac{24}{\pi^2}\sigma_z^2} \quad (16)$$

$$\hat{\beta} = \min(z), \quad (17)$$

where σ_z^2 is the variance of the image and $\min(z)$ is the minimum value of the image in B-Mode.

With a close representation of the ground truth image, an attempt to simplify the information in the image is made by narrowing the intervals of intensity represented. This consists of dividing the image and then multiplying it by the same number, being that number a power of 2, based on the number of intensity intervals desired. In this case, the image is divided and then multiplied by 32, as is the case, the information on 5 of the bits is discarded and the colors are represented in only 8 intensity ranges ($256/32 = 8$). The information on these 5 bits is of the finer details, and as such, when disregarded, the result is a coarser representation of the original image. To further simplify the image, without causing significant damage to the contours, a median filter was applied 2 times. The entirety of the preprocessing pipeline can be seen in Figure. 1.

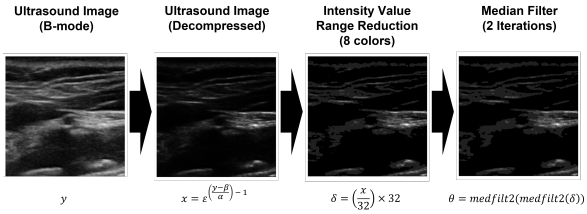

Figure 1: Diagram of the preprocessing of the dataset.

Training

The preprocessing was done to the full chosen dataset of 670 ultrasound images. The created pairs were divided into 500 training pairs and 170 validation pairs. The network was then allowed to train for 700 epochs using the default U-net architecture as the generative network and Patch GAN as the discriminator network. Furthermore, the λ that controls the L_1 part of the loss was set to 100, the initial gain was set to 0.02, the learning rate was set to 0.0001. The training was done using the pix2pix PyTorch code and Google CoLab, making use of one of their GPUs of the Intel(R) Xeon(R) CPU @ 2.30GHz model. The training time is about 17 hours using the hardware with these specifications. Several epoch values were tested by taking four images from the validation set and analysing the evolution of the results through the plotting of the MSE between the results from the epoch in question and the results from 50 more epochs of training. It is not wise to base the epoch choice only on image metrics when it comes to GANs, and as such, direct observation of the images was also a key component in this choice of the epoch number.

3.2. US Denoising GAN

As mentioned in the analogous section of the previous chapter, medical images are few and far between. Unlike that section, however, the images needed for the training of this network are MRI images.

Dataset

Fortunately, a series of 1024 MRI slices of the heart was made available from *A. Andreopoulos et al.* [22]. Nonetheless, many of these images contained undesirable features, such as large dark areas or spots of extreme intensity. A manual selection of these images was made to ensure that the data fed to the network was of the best quality within the available dataset. The final number of images chosen was 440 structural MRI images of the heart. To increase the versatility of the network and ensure the filtering aspect in regards to speckle noise removal, the dataset was balanced with grayscale images of real-world objects corrupted with the Rayleigh noise model. The idea behind the incorporation of these images into the training of the GAN was to allow for the learning of an array of somewhat different contours that are not very easily found in medical images or, when they are indeed found, are somewhat unclear. The everyday object image dataset used was the Columbia Object Image Library (COIL-100) Dataset [23], which is composed by color images of 100 different mundane objects, taken at 5° intervals until a full rotation around the object is accomplished. This dataset contains 7200 images, and from these, 440 images where sampled randomly to maintain an equilibrium between both approaches.

Preprocessing

Starting with the preprocessing of the medical images, the first step was to create the simplistic phantoms that the first network accepts as input. This was done by recycling components of the method detailed in the earlier section (section. 3.1). More specifically, the intensity range reduction into values stored in 3 bits, followed by two iterations of the adaptive median filter for smoother and more rounded contours. With the phantoms of MRI images in hand, it is simply a matter of running these images thru the network. It is noteworthy to mention that the time necessary for the network to process these images rounded the 30 seconds mark. Beforehand, however, data augmentation was performed on the raw MRI images. This data augmentation entailed applying rotations, translations or reflections to the images.

The preprocessing of the COIL-100 dataset begins with a random sampling of 440 images. These

images are converted from RGB to grayscale and, as their resolution is different from the standard resolution used in this work of 256x256 pixels, a linear interpolation was used. The COIL-100 dataset, besides being vast, contains images from different angles of the objects, and as such, data augmentation was not used. These images were then considered the ground-truth images. Firstly the Rayleigh noise model in Equation. 4 was applied. This was followed by a decimation and immediate upsample of the images using a factor of 2. What this means is that the image was reduced to half the size, discarding the intensities in the even entries of the image. Afterwards, the image is returned to the original size, with the help of a linear interpolation. The last step of the preprocessing is the logarithmic compression model. This model is reliant on two unknown parameters, that vary with the acquisition options of the device. To increase the robustness of the network, these parameters were chosen randomly for each image from the intervals $[10, 50]$ for α and $[0, 40]$ for β. The images were then normalized and stored.

Training

For the training of the network, the dataset was once again divided into training images (600 images) and validation images (280 images). The ratio was 1:1 for the distribution of medical images to COIL-100 images in each subset. As was the case with the first network, the default U-net architecture for the generative network and Patch GAN for the discriminator network were used. The number of epochs chosen was 550 epochs. This value was also chosen from calculating the MSE of 4 images resulting from one epoch to a previous epoch (in intervals of 50 epochs) and plotting the values. When the MSE remained somewhat constant across the plots of the 4 chosen images after a given epoch, that epoch was selected. In this case, the evaluation encompassed 1000 epochs. The training was done using the pix2pix PyTorch code and Google CoLab, making use of one of their GPUs of the Intel(R) Xeon(R) CPU @ 2.30GHz model. Once again, the λ that controls the L_1 part of the loss was set to 100, the initial gain used was 0.02, the learning rate chosen was 0.0001. The time that it took for the training of this Denoising GAN was also similar to the US Simulator GAN, approximately 15 hours. A simplification of the training pipeline can be seen in Figure. 2.

4. Results and Assessment

One of the goals of this work was the removal of speckle noise from US images. This type of noise can be characterized by the Rayleigh and the Nakagami distributions. The analysis of the histograms

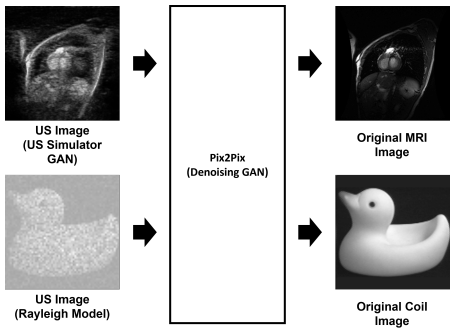

Figure 2: Diagram of the training of the Denoising GAN.

of the multiplicative noise can be used to determine if the networks are fulfilling their respective purposes.

4.1. Statistical Analysis of the Noise of the US Simulator GAN

Starting with the US Simulator GAN, whose purpose is to transform simplistic phantoms of MRI images into corresponding US images, the methodology described above is applied. In more practical terms, this method consists in using the output of the network (simulated US image), transforming it into the RF domain, and divide it, pixel by pixel, by the original MRI image. The images used for this test were not part of the training process and are nothing like carotid artery images. To ease the interpretation and showcase the similarity of the distributions with the Rayleigh PDF, this distribution was superimposed onto the histograms in Figure. 3.

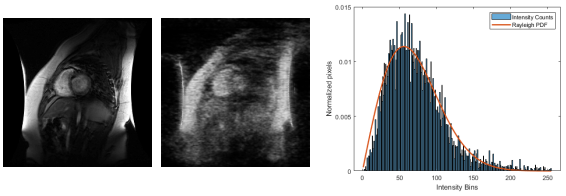

Figure 3: The original MRI image (left), the image resulting from the US Simulator GAN (right) and the histogram extracted from the multiplicative noise

As it can be seen, the distribution of the histogram shows an uncanny likeness to the Rayleigh distribution, but does not perfectly align with it. This can be due to the fact that the RF domain image is obtained with α and β estimated from the B-Mode image, and not the actual parameters from the acquisition. Another factor is that the network was trained to replicate every kind of noise and transformation inherent to the US acquisition process, not only speckle. Nonetheless, the histogram displays strong components alike to the Rayleigh distribution, indicating a strong presence of speckle noise.

4.2. Statistical Analysis of the Noise of the Denoising GAN

The properties of the Denoising GAN will be tested in this section.

US Simulator GAN

The first tests involve the outputs of the US Simulator GAN based on real MRI images.

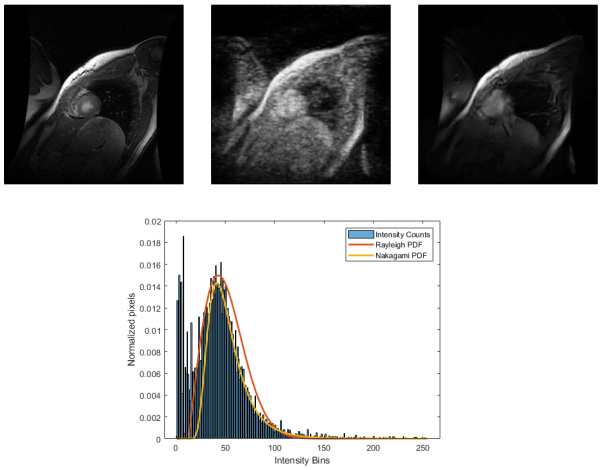

Figure 4: The top row shows the original (left), the US Simulator GAN (middle) and the Denoising GAN (right) images. The bottom image shows the histogram extracted by the division of the Rayleigh model and Denoising GAN outputs.

Analysing Figure. 4, it is interesting to note that, although the smaller details were not recovered with the Denoising GAN, the larger-scale structures seem continuous and sharp with little blurring. The shape of the histogram is not a perfect fit of the distributions, but instead forms an intermediate shape between them. There is also an increased number of counts in the lower intensity ranges. A reasonable explanation would be that the US Simulator GAN is simulating noise that does not perfectly follow statistical distributions. The quality of the denoised image was also be assessed by more conventional means such as SNR, PSNR, SSIM, and MS-SSIM.

Table 1: Values of SNR, PSNR, SSIM and MS-SSIM for the noisy (n) and denoised (d) images of the heart. The noisy images were result of the US Simulator GAN.

	Heart (d)	Heart (n)
SNR	12.3958	0.2359
PSNR	28.001	15.8411
SSIM	0.725	0.3752
MS-SSIM	0.9098	0.7038

It is obvious from the analysis of Table. 1, that the Denoising GAN improves the quality of the image considerably, the higher the values of these metrics, the better the quality of the image.

Rayleigh Speckle noise and Logarithmic Compression model

The subsequent step in the assessment of the Denoising GAN was to use it to process MRI images that were subjected to the Rayleigh noise model, shown in Figure. 5.

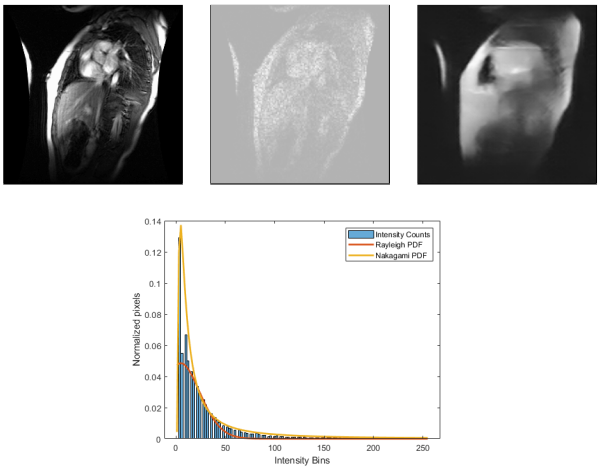

Figure 5: The top row shows the original (left), the Rayleigh model (middle) and the Denoising GAN (right) images. The bottom image shows the histogram extracted by the division of the Rayleigh model and Denoising GAN outputs.

In these simulated US images the destruction of the finer details by the noise is evident. As a consequence, the denoising network was only able to recover the larger structural details. The analysis of the histogram and comparison with the curves of both Rayleigh and Nakagami PDFs indicates that the multiplicative noise contains similarities to both distributions.

Table 2: Values of SNR, PSNR, SSIM and MS-SSIM for the noisy (n) and denoised (d) images of the heart. The noisy images were result of the Rayleigh and compression models.

	Heart (d)	Heart (n)
SNR	6.07	-6.03
PSNR	16.57	4.47
SSIM	0.30	0.14
MS-SSIM	0.77	0.55

The full reference image quality metrics were extracted and are displayed in Table. 2. As was the previous case, when compared to the noisy image, the denoised image improved across the board in terms of these metrics.

Field II Program

The next step is the analysis of images simulated with the Field II program. The image used was the example of a kidney phantom given by the authors of the program.

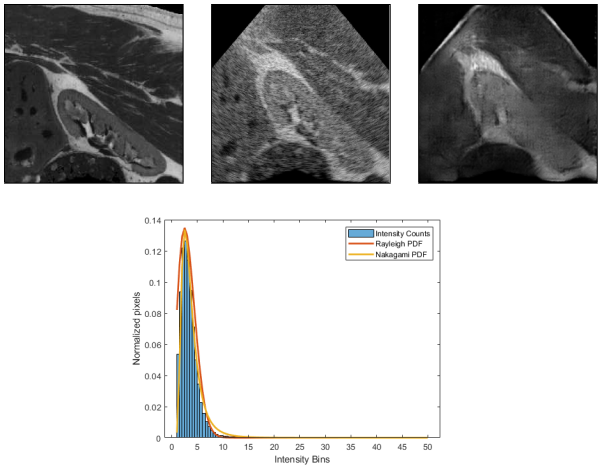

Figure 6: The top row shows the original (left), the Field II (middle) and the Denoising GAN (right) images. The bottom image shows the histogram extracted by the division of the Field II and Denoising GAN outputs.

The resulting image loses smaller structural information, as seen in Figure. 6, which could be explained by the differences in scale between the network training images and the high-resolution original image. The larger structures, however, are recovered with a smoothness and sharpness of contours very agreeable to the goals of this work. From the histogram, it can be seen that the Nakagami distribution fits almost perfectly. This confirmed likeness is very encouraging in terms of expected results.

The distortion of the US images caused by the Field II program means that the comparison metrics used in the previous sections are not useful. As such, the use of a no-reference image quality metric is introduced, the BRISQUE score, shown in Table. 3.

Table 3: BRISQUE score values for the versions of the kidney image: original, output of the Field II program and output of the Denoising GAN.

	Original	Field II	Denoised
Kidney	30.71	36.83	33.37

From these values, a significant drop in the BRISQUE score is noticeable, near to half the value of the difference between worst (Field II) and best (Original) case scenario.

Real US images

The final tests focus on real US images, the most important of the results, seeing as the denoising of this type of image is the purpose of the network. Due to the lack of a ground truth image, the full-reference metrics are not used as evaluation metrics. As such the BRISQUE scores are calculated for the original US images, denoised counterparts and the MRI dataset. The scores of the MRI will serve as reference and certification that a lower BRISQUE is related to better quality.

Table 4: Metric extracted from the BRISQUE scores of the 3 datasets, MRI, Denoised and US, namely the mean and median.

	Mean	Median
US	37.47	38,61
Denoised	34.20	33.92
MRI	**30.13**	**31.57**

From the values in Table. 4, it can be seen that, in most cases, the MRI dataset exceeds the US and Denoised dataset, in terms of lowest mean and median, as was intended to prove. Moving on, some examples of the Denoising GAN outputs are shown in Figure. 7.

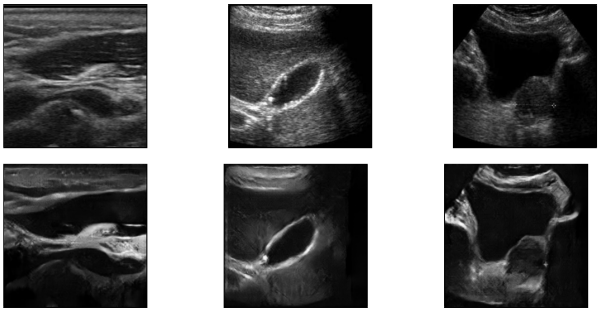

Figure 7: Comparison of the original US image (top row), with the result from the Denoising GAN (bottom row).

From the images, the increase in contrast and the abundance of smooth and well-defined contours when compared to the original US images is clear. Furthermore, analysing the number of times that the denoising process leads to a lower BRISQUE score, it can be said that in 372 images out of the total 530 images, the score lowered after the denoising network was used (in 70% of cases in this dataset).

4.3. Comparison with conventional techniques

The best way to show the relevance and benefits this Denoising GAN brings, is to compare it to methods used presently, namely the adaptive median filter (AMed) and the anisotropic diffusion filter (SRAD).

At first glance, it is immediately noticeable that the Denoising GAN changes the image more drastically. As a consequence of these changes, the images become more visually pleasing. This is noticeable in the smoothness of the contours and structures. The classical filters, on the other hand, tend to preserve the initial visual information of the images while sacrificing the removal of speckle noise. Like in the previous sections, the quality metrics were extracted.

From both these tables, the values that are most indicative of image quality are presented in bold.

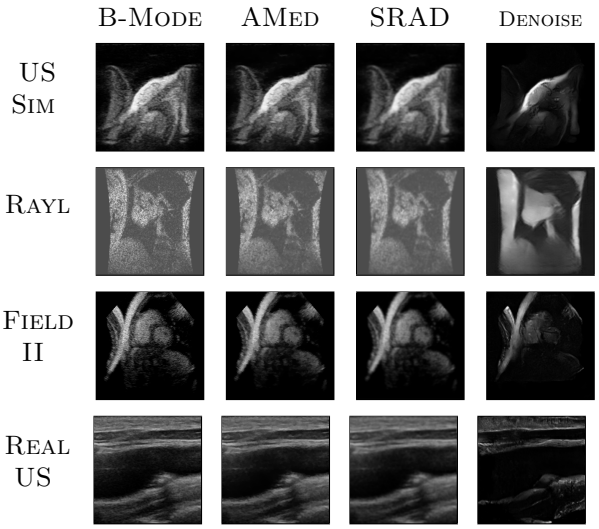

Figure 8: Comparing the B-Mode image (B-Mode), with the Adaptive Median filter (AMed), the Anisotropic Diffusion filter (SRAD) and the Denoising network (Denoise). In each line the origin of the B-Mode images is given, starting with the output of the US simulator network (US-Sim), the Rayleigh Speckle noise model and logarithmic compression model (Rayl), the Field II software (Field II) and real US image acquisition (Real).

Table 5: Values of SNR, PSNR, SSIM and MS-SSIM for the B-Mode, Adaptive median filtered (AMed) and Anisotropic Diffusion Filtered (SRAD) images, as well as the output of the Denoising GAN (Denoise). These images used are shown in the first line of Figure. 8 and the B-Mode image was obtained from the US simulator network.

	B-Mode	AMed	SRAD	Denoise
SNR	0.126	0.189	0.281	**13.037**
PSNR	15.72	15.78	15.87	**28.63**
SSIM	0.381	0.409	0.433	**0.760**
MS-SSIM	0.691	0.711	0.0735	**0.920**

In comparison with each method of denoising, and across the different metrics extracted, the superiority of using the Denoising GAN in regards to these images and metrics is obvious. Next, the BRISQUE score is calculated for both the B-Mode image obtained from the Field II program and a real US image, as well as for their denoised versions.

When speaking of the image simulated by the Field II program, the Adaptive Median filter outperforms the Denoising GAN, while in terms of the real US image, the Denoising GAN produces an image with a much better score.

5. Conclusions

This work was developed as a way to improve speckle noise removal and improve image quality in ultrasound images. The approach chosen to achieve this was based on the use of Generative Adversarial Networks, more specifically the pix2pix network. This network has the purpose of generating images based on other images, a procedure commonly known as image-to-image translation.

Table 6: Values of SNR, PSNR, SSIM and MS-SSIM for the B-Mode, Adaptive median filtered (AMed) and Anisotropic Diffusion Filtered (SRAD) images, as well as the output of the denoising network (Denoise). The images used are shown in the second line of Figure. 8 and the B-Mode image was obtained using the Rayleigh speckle noise and logarithmic compression model.

	B-Mode	AMed	SRAD	Denoise
SNR	3.17	3.22	3.15	**8.50**
PSNR	12.39	12.45	12.38	**17.73**
SSIM	0.278	0.276	0.260	**0.303**
MS-SSIM	0.719	0.719	0.704	**0.746**

Table 7: Values of BRISQUE score for the B-Mode, Adaptive median filtered (AMed) and Anisotropic Diffusion Filtered (SRAD) images, as well as the output of the denoising network (Denoise). The images used are shown in the third and fourth lines of figure.8.

	B-Mode	AMed	SRAD	Denoise
Field II	43.22	**33.18**	46.92	37.37
Real US	40.53	43.36	52.15	**27.89**

Firstly, a network that could create ultrasound images based on simplistic phantoms was developed. This network was trained by using pairs of real ultrasound images and phantoms, and its results are later certified as adequate ways of simulating speckle noise and the ultrasound acquisition process. The certification was based on the noise extracted from these images and through the observation that it followed the distributions that characterize speckle noise. It must be said that the network had a slightly worse performance than the Field II software, at least visually, but the computational and time gains are very much worth it. The Field II software needed 3 days for one image while the network performed the processing of 999 images in 2 to 7 minutes. This network was used to create a dataset of paired US-MRI images of the heart.

The second network, responsible for denoising and image enhancement, was trained with a mixed dataset. Half of the dataset was composed of the pairs of heart images produced by the first network, and the other half was of mundane objects corrupted with Rayleigh speckle noise model and logarithmic compression. This network was proved to reduce speckle noise significantly and improve image quality, a result mirrored by the full reference metrics used to analyse the processed images. In terms of processing real ultrasound images, this network improved the visual quality as observable in the examples and showcased by the BRISQUE metric. There are some tests and examples where the superiority of this network is not as clear, but the overall analysis is positive.

Lastly, the performance of the denoising network was compared to classical filtering algorithms. The

results, both output images, and image quality assessment scores, support the claim that this network is a considerable upgrade in denoising methods. Once again, the results of the BRISQUE score are not as clear when analysing the denoised Field II outputs. This metric, however, is simply a guideline and direct visual analysis can be shown to confirm that the Denoising GAN is a worthwhile improvement. Be it in terms of the simulation or denoising of ultrasound images, the training of the pix2pix network as was described in this work proves to be an effective novel approach that already has merit, but can be further through the use of more specialized training datasets or modifications of the network architecture itself. The weights of the networks used and some more examples of the results are made available at: `https://drive.google.com/drive/folders/16WRXnFvXBVw7DWl1bE12o4j3MPOeDeG9?usp=sharing` and `https://github.com/AntonioAzeitona/DenoisingGANVideos`.

Acknowledgements

This document was written and made publically available as an institutional academic requirement and as a part of the evaluation of the MSc thesis in Biomedical Engineering of the author at Instituto Superior Técnico. The work described herein was performed at the Institute for Systems and Robotics of Instituto Superior Técnico (Lisbon, Portugal), during the period February-November 2021, under the supervision of Prof. João Sanches.

References

[1] 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.

[2] Saad Albawi, Tareq Abed Mohammed, and Saad Al-Zawi. Understanding of a convolutional neural network. In *2017 International Conference on Engineering and Technology (ICET)*. IEEE, August 2017.

[3] Antonia Creswell, Tom White, Vincent Dumoulin, Kai Arulkumaran, Biswa Sengupta, and Anil A. Bharath. Generative adversarial networks: An overview. *IEEE Signal Processing Magazine*, 35(1):53–65, January 2018.

[4] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. In *Computer Vision and Pattern Recognition (CVPR), 2017 IEEE Conference on*, 2017.

[5] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Computer Vision (ICCV), 2017 IEEE International Conference on*, 2017.

[6] O.V. Michailovich and A. Tannenbaum. Despeckling of medical ultrasound images. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 53(1):64–78, 2006.

[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.

[8] 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.

[9] V. Damerjian, O. Tankyevych, N. Souag, and E. Petit. Speckle characterization methods in ultrasound images – a review. *IRBM*, 35(4):202–213, September 2014.

[10] Deepika Koundal, Savita Gupta, and Sukhwinder Singh. Nakagami-based total variation method for speckle reduction in thyroid ultrasound images. *Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine*, 230(2):97–110, 2016. PMID: 26721907.

[11] 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*, 39(2):262–267, 1992.

[12] Jørgen Jensen. Field: A program for simulating ultrasound systems. *Medical and Biological Engineering and Computing*, 34:351–352, 01 1996.

[13] P. Perona and J. Malik. Scale-space and edge detection using anisotropic diffusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(7):629–639, July 1990.

[14] T. Loupas, W.N. McDicken, and P.L. Allan. An adaptive weighted median filter for speckle suppression in medical ultrasonic images. *IEEE Transactions on Circuits and Systems*, 36(1):129–135, January 1989.

[15] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Y. Bengio. Generative adversarial networks. *Advances in Neural Information Processing Systems*, 3, 06 2014.

[16] Mehdi Mirza and Simon Osindero. Conditional generative adversarial nets, 2014.

[17] P.S. Hiremath, Prema T., and Sharan Badiger. Speckle noise reduction in medical ultrasound images. In *Advancements and Breakthroughs in Ultrasound Imaging*. InTech, June 2013.

[18] Li Sze Chow and Raveendran Paramesran. Review of medical image quality assessment. *Biomedical Signal Processing and Control*, 27:145–154, May 2016.

[19] Z. Wang, E.P. Simoncelli, and A.C. Bovik. Multiscale structural similarity for image quality assessment. IEEE, 2003.

[20] Anish Mittal, Anush Krishna Moorthy, and Alan Conrad Bovik. No-reference image quality assessment in the spatial domain. *IEEE Transactions on Image Processing*, 21(12):4695–4708, 2012.

[21] João M. Sanches and Jorge S. Marques. Compensation of log-compressed images for 3-d ultrasound. *Ultrasound in Medicine & Biology*, 29(2):239–253, February 2003.

[22] Alexander Andreopoulos and John K. Tsotsos. Efficient and generalizable statistical models of shape and appearance for analysis of cardiac mri. *Medical Image Analysis*, 12(3):335–357, 2008.

[23] S. K. Nayar S. A. Nene and H. Murase. Columbia object image library (coil-100), February 1996.