RESEARCH ARTICLE

Ultrasound Despeckling With GANs and Cross Modality Transfer Learning

DIOGO FRÓIS VIEIRA [ID] **¹, AFONSO RAPOSO** [ID] **¹, ANTÓNIO AZEITONA¹, MANYA V. AFONSO** [ID] **²,**
LUÍS MENDES PEDRO [ID] **³, AND J. SANCHES** [ID] **¹**

¹Institute for Systems and Robotics, LARSyS, Department of Bioengineering, Instituto Superior Técnico, 1049-001 Lisbon, Portugal
²Wageningen University and Research, 6708 PB Wageningen, The Netherlands
³Faculdade de Medicina, Universidade de Lisboa, 1649-028 Lisbon, Portugal

Corresponding author: Diogo Fróis Vieira (diogovieira16@tecnico.ulisboa.pt)

ABSTRACT Ultrasound images are corrupted by a type of signal-dependent noise, called speckle, difficult to remove or attenuate with the classical denoising methods. On the contrary, structural Magnetic Resonance Imaging (MRI) is usually a high resolution low noise image modality that involves complex and expensive equipment and long acquisition times. Herein, a deep learning-based pipeline for speckle removal in B-mode ultrasound medical images, based on cross modality transfer learning, is proposed. The architecture of the system is based on a pix2pix Generative Adversarial Network (GAN), D, able to denoise real B-mode ultrasound images by generating synthetic MRI-like versions by an image-to-image translation manner. The GAN D was trained using two classes of image pairs: i) a set consisting of authentic MRI images paired with synthetic ultrasound images generated through a dedicated ultrasound simulator based on another GAN, S, designed specifically for this purpose, and ii) a set comprising natural images paired with their corresponding noisy counterparts corrupted by Rayleigh noise. The denoising GAN proposed in this study demonstrates effective removal of speckle noise from B-mode ultrasound images. It successfully preserves the integrity of anatomical structures and avoids reconstruction artifacts, producing outputs that closely resemble typical MRI images. Comparative tests against other state-of-the-art methods reveal superior performance of the proposed denoising strategy across various reconstruction quality metrics.

**INDEX TERMS** Ultrasound, denoising, deep learning, GANs, modality translation.

I. INTRODUCTION

Imaging modalities such as ultrasound, synthetic aperture radar, LASER, and optical coherence tomography are affected by a type of (pseudo) multiplicative noise, called speckle, that results from constructive and destructive interference of the back-scattered radiation [1], [2]. A relevant step in image processing is noise removal, to improve overall image quality and better visualization of morphological features [3]. In the case of ultrasound, the speckle noise contaminating the Radio Frequency (RF) image is modeled

as multiplicative and follows the Rayleigh statistical distribution [2], [4].

Many existing approaches rely on ultrasound images in the RF domain rather than the displayed B-mode images [5], [6]. However, RF data is not usually available in commercial equipments, requiring its estimation, which introduces an additional source of error. In this paper, we present a deep learning-based denoising pipeline [7], [8] that eliminates the intermediate step of estimating RF, streamlining the process and mitigating potential sources of error. The proposed method inputs the available B-mode data to generate its corresponding MRI-like denoised image, as showcased in Fig. 1. Deep learning autoencoders have been successfully used to learn structures in clean natural images and remove

---

The associate editor coordinating the review of this manuscript and approving it for publication was Ramakrishnan Srinivasan [ID].
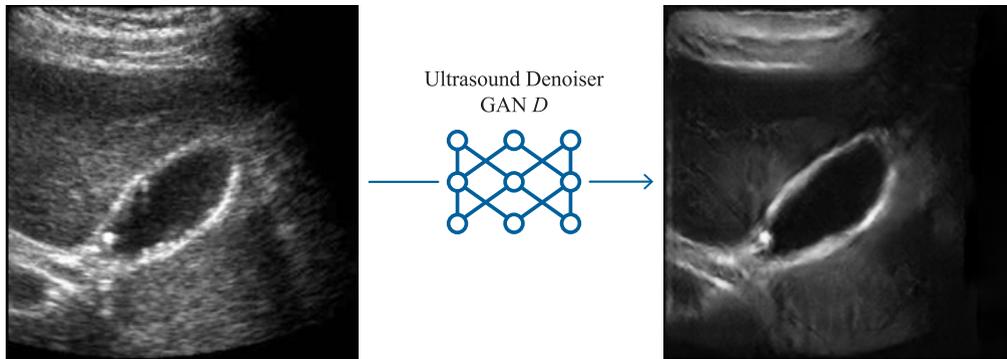
FIGURE 1. Ultrasound image of a gallbladder and its denoised version obtained with the trained denoising GAN *D*, showcasing its ability to remove speckle noise from ultrasound images, transforming them into an image with characteristics similar to MRI.

noise [9], but require a training phase with ground truth data consisting of pixel-wise pairs of noisy and clean images. Our proposed pipeline addressed this requirement by generating a synthetic dataset from a baseline dataset of a different modality, which was used to train the denoising deep learning model.

Generative Adversarial Networks (GANs) are a type of neural network whose architecture comprises two networks: the generator and the discriminator, that compete with each other, in the sense that the generator tries to generate fake data that the discriminator would find indistinguishable from real data. Such network architecture can be applied for image-to-image translation. Image-to-image translation is when the task of the generator is to create an image from an input image. In other words, translating the input image to the output as a new different modality image.

Pix2pix [10] is a well-known GAN that has produced impressive results in transformations between pixel-wise paired images such as satellite images, maps or natural images and their edges. This network is a conditional adversarial network and learns its own loss function instead of using only ℓ_1 or ℓ_2, which would lead to blurry results. Thus, training this model with a set of noisy ultrasound images and their corresponding noise-free images would provide a tool that could be applied to a new noisy image and generate a clean version as the (synthetic) output.

A. RELATED WORK

Existing methods for ultrasound despeckling include techniques of different natures. Filter-based methods, such as the Speckle Reducing Anisotropic Diffusion (SRAD) filter [11], the non-local means filter [12], and the Block Matching 3D filter (BM3D) [13], are among the used classical ultrasound denoising methods in the field. In particular, Coupé et al. [14] proposed an Optimized Blockwise Non-Local Means (OBNLM) denoising filter, which was used for noise removal in MRI images. In [13], cardiac and fetal ultrasound images were denoised using the BM3D method [6] to group

similar image patches into a 3D grid, which was then denoised using a thresholding operation on the brushlet coefficients. A Non-Local Low-Rank (NLLR) filter-based despeckling method, proposed in [15], selects significantly speckled patches which are then refined using a low-rank minimization using the truncated weighted nuclear norm and structured sparsity. In more recent works, Wang et al. [16] proposed a denoising formulation comprising a Kullback-Leibler difference term and a sum of total variation terms of different orders (denoted as TVHTVM). In [17], a filter-based framework using anisotropic diffusion and total variation was proposed. Yet another despeckling method Despeckling by Quantum Interactive Patches (denoted as De-QuIP) from Duta et al. [18], inspired by the quantum many-body theory.

More recent approaches focus on deep learning [7], [8], which is rapidly emerging as the state-of-the-art technique for several computer vision problems, especially segmentation. Stacked denoising autoencoders, trained on simulated ultrasound signals, were previously used for the compression and reconstruction of ultrasound signals [19]. In [20], a trained Convolutional Neural Network (CNN) with pairs of Inphase and Quadrature images and conventionally despeckled images was reported to be able to despeckle ultrasound images and also approximate the corresponding CT image, by exploiting the fact that both X-ray CT and ultrasound imaging are affected by physical properties of the tissue [20].

A 3D U-Net architecture-based method was used in [21] to filter clutter artifacts in echocardiographic images. A deep learning-based despeckling method for liver ultrasound volumes [22] used a 3D dense U-Net model to process 3D ultrasound B-mode data, for 3D registration. A CNN was proposed for Plane Wave Imaging (PWI) in [23], which combines the image formation (beamforming) and speckle reduction stages, with the parameters learned from training on simulated data.

Among GAN-based works, MimickNet [24] is a cycle GAN-based deep learning framework that approximates post-processed images with reduced speckle, getting around
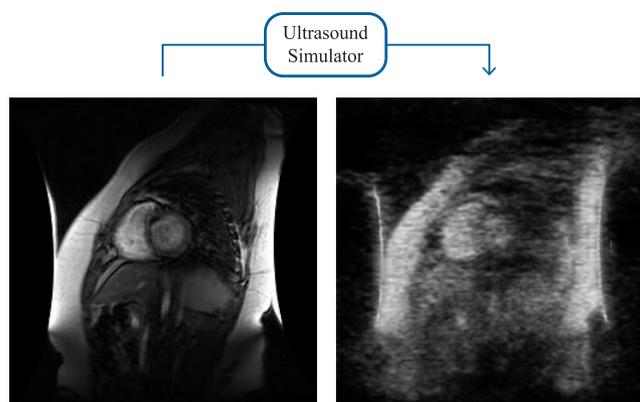
FIGURE 2. Functioning of the ultrasound simulator, composed of a preprocessing block and another GAN (*S*), able to generate synthetic ultrasound images from MRI images, needed for the training of *D*.

the proprietary closed-source B-mode mappings used by ultrasound equipment manufacturers [25]. In [26], GANs trained on a dataset of cardiac images and their despeckled versions obtained using the NLLR method [15], along with a U-Net type of generator and a patch-wise discriminator, were found to be able to learn to denoise speckle patterns even for non-cardiac ultrasound images. GANs were also used in [27] to improve lateral resolution in RF data in order to improve strain image quality in quasi-static ultrasound strain elastography. Finally, a GAN-based method was proposed in [28] to recover multi-focus ultrasound images from blur, speckle, and other artifacts during acquisition. This work improved the realism of the ultrasound images by using adversarial loss functions in addition to Mean Square Error (MSE) as well as the boundary-seeking method for improving the stability of training. Another recent work [29] used pix2pix trained with ultrasound images and their denoised versions obtained using existing denoising algorithms. An important work by Kong et al. [30] is the RegGAN, which is composed of pix2pix and cycle GANs designed for medical image-to-image translation and speckle removal in ultrasound images.

B. CONTRIBUTIONS

In this work, a method for denoising real B-mode ultrasound images is proposed. It is based on a trained pix2pix GAN model, denoted as *D*, which translates noisy (speckled) input ultrasound images into clean images resembling MRI, as shown in Fig. 1. To train *D*, another GAN, *S*, was designed to produce realistic speckled synthetic ultrasound data from iconic, simplified, and abstracted representations of anatomical structures, obtained by a preprocessing step of real MRI images. This pipeline, termed as ultrasound simulator, is illustrated in Fig. 2. The quality of the results, assessed through visual inspection and various quality metric functions, was outstanding and outperformed the other methods used for comparison purposes.

II. PROPOSED METHOD

The proposed method is based on two pipelines: i) a B-mode ultrasound simulator, composed of a preprocessing block and a GAN, called *S*; and ii) a B-mode ultrasound denoiser, composed of another GAN, *D*. Each of these GANs consists of a stand-alone pix2pix network trained appropriately. The following sections will describe the datasets, strategies and experimental setup used to train both networks. The procedure to quantitatively and qualitatively evaluate the denoising performance will also be discussed.

A. DATASETS

This work utilized three sets of data. The first set consisted of clinical ultrasound images of the carotid, aiming to instruct the system about the specific features characteristic of carotid ultrasound images. The data used in this work (images and frames of videos) comprised 1024 ultrasound images of carotid arteries with different morphologies, in different probe positions, along different anatomical segments and some even displaying pathologies, more specifically arterial plaques or stenosis. A subset of 670 images was selected after a selection process that excluded images according to a list of criteria: images with low quality, with several artifacts, with too many regions with very little ultrasound information, images that included Doppler information and images taken from a transverse viewpoint. This exclusion was based on the lack of structural information, either caused by the presence of the Doppler, or by the negligible amount of it in the transverse images.

The second set comprised a collection of public natural images that were corrupted with Rayleigh noise and log-compressed according the model described in [4]. This set was employed to familiarize the system with the statistical distribution associated with speckle noise. To do that, a subset of the Columbia Image Library (COIL-100) dataset [31] was gathered as a source of images of natural objects, used to train the ultrasound denoiser network *D*. COIL-100 consists of 72000 color images of 100 different objects such as mugs and other retail items, taken at different angles between 0 and 360 degrees, in steps of 5 degrees. These images have a resolution of 128×128 pixels. From this set, 375 images were randomly selected, which were then converted to grayscale and resized to 256×256 pixels using linear interpolation of neighboring points.

To train the GAN *S* that composes the ultrasound simulator, as well as to train the denoising network (*D*), a dataset of 375 MRI images of the human heart [32] was used, each image being of size 256×256 pixels. Data augmentation was applied by rotating the images by random values between -90 and 90 degrees, followed by random translations of between -50 and 50 pixels in the *x* direction, and between -25 to 25 pixels in the *y* direction. Finally, the images were flipped randomly on the *x* and *y* axes. This is necessary to avoid artifacts caused by overfitting, since higher pixel values are concentrated near the center of the original MRI images and the borders have lower values.
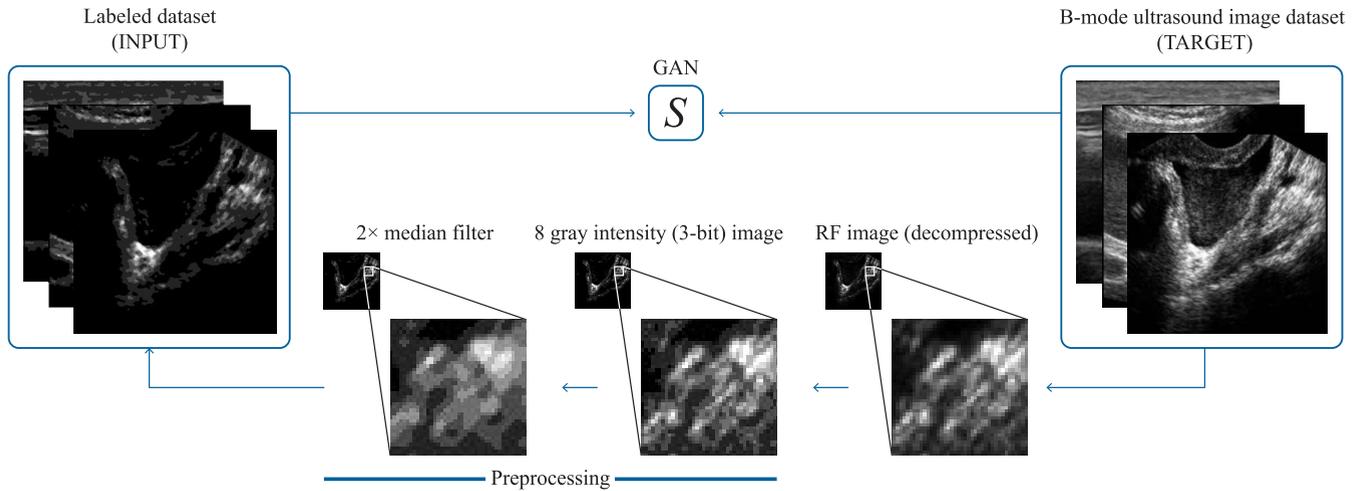
FIGURE 3. Pipeline for the training of *S*, with an example image of a uterus and a zoomed-in detail to show the preprocessing effect. Each ultrasound image from the dataset (used as target) was preprocessed (decompression, quantization to 3-bit values and two iterations of median filtering). After preprocessing, a phantom-like image was obtained and used as input. *S* is trained using the original ultrasound image paired with its iconic/labeled image counterpart.

To assess the performance of *D*, a test dataset of 148 MRI images was collected from the aforementioned MRI image dataset with images of the human heart not used during the training phase [32].

B. GENERATING SYNTHETIC ULTRASOUND IMAGES FROM MRI

The goal of the GAN *S* was to produce simulated B-mode ultrasound images given labeled images representing low complexity "phantoms" as input. In other words, iconic/labeled images can be obtained from MRI images (after preprocessing) and used as input, producing a synthetic version with characteristics typical of ultrasound images. The images produced by *S* will be crucial to create training pairs to train the denoising GAN, *D*, as will be covered in the next section.

Fig. 3 illustrates how *S* was trained. For the network to learn characteristic patterns of ultrasound images, a dataset of images of this type of modality must be used as target. First, each original B-mode ultrasound image, denoted as *y*, was converted to the RF domain, y_{RF}, using the decompression function for each pixel (i, j),

$$y_{RF}(i,j) = e^{\frac{y(i,j)-\beta}{\alpha}} - 1, \quad (1)$$

following the method described in [4]. The mapping parameters α and β were calculated according to [33].

Images now in the RF domain were preprocessed in two steps. First, the intensity depth (the number of bits to represent each pixel value) was reduced by multiplying and dividing each integer grayscale pixel by 32. This process discards information from 5 of the bits, representing the intensities in the remaining 3 bits, allowing for only 8 levels. By discarding the 5 least significant bits, a coarser/iconic representation of the original image was obtained. Finally, the resulting image was smoothed by 2 iterations of median

filtering with a squared kernel of size 3×3 pixels. At the end of this discretization process, each pixel gets assigned an integer value, or label, between 0 and 7 proportional to its original pixel intensity. Overall, the resulting labeled image represents simplified anatomical structures with a phantom-like appearance, as shown in Fig. 3.

The B-mode image and the corresponding 3 bit (or 8 level) image pairs so obtained were used to train the GAN *S*. 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 pix2pix model trained this way can now be applied on the 8-level version of a MRI image (prepared after the decompression and preprocessing steps), producing an output image that resembles a B-mode ultrasound image. This process is shown on the top row of Fig. 4, as it was an indispensable framework to train *D*.

C. DENOISING OF ULTRASOUND IMAGES

The GAN *D* aimed to despeckle input B-mode ultrasound images by converting them to clean MRI-like versions. For this, the GAN must learn i) the main characteristics of MRI images, and ii) the statistical properties of speckle noise. Therefore, two distinct datasets had to be designed for training, to account for these two aspects. Fig. 4 illustrates how *D* was trained.

First, the GAN must learn how to reproduce the main characteristics and patterns of MRI images, including their high resolution details and low noise. For this, a set of 375 synthetic B-mode ultrasound images were generated from original MRI images using the trained *S*, as shown on the process of the top half of Fig. 4. The labeled/iconic input image generation for *S* was done using the preprocessing steps described before with a single difference: MRI images do not require decompression and are directly intensity depth
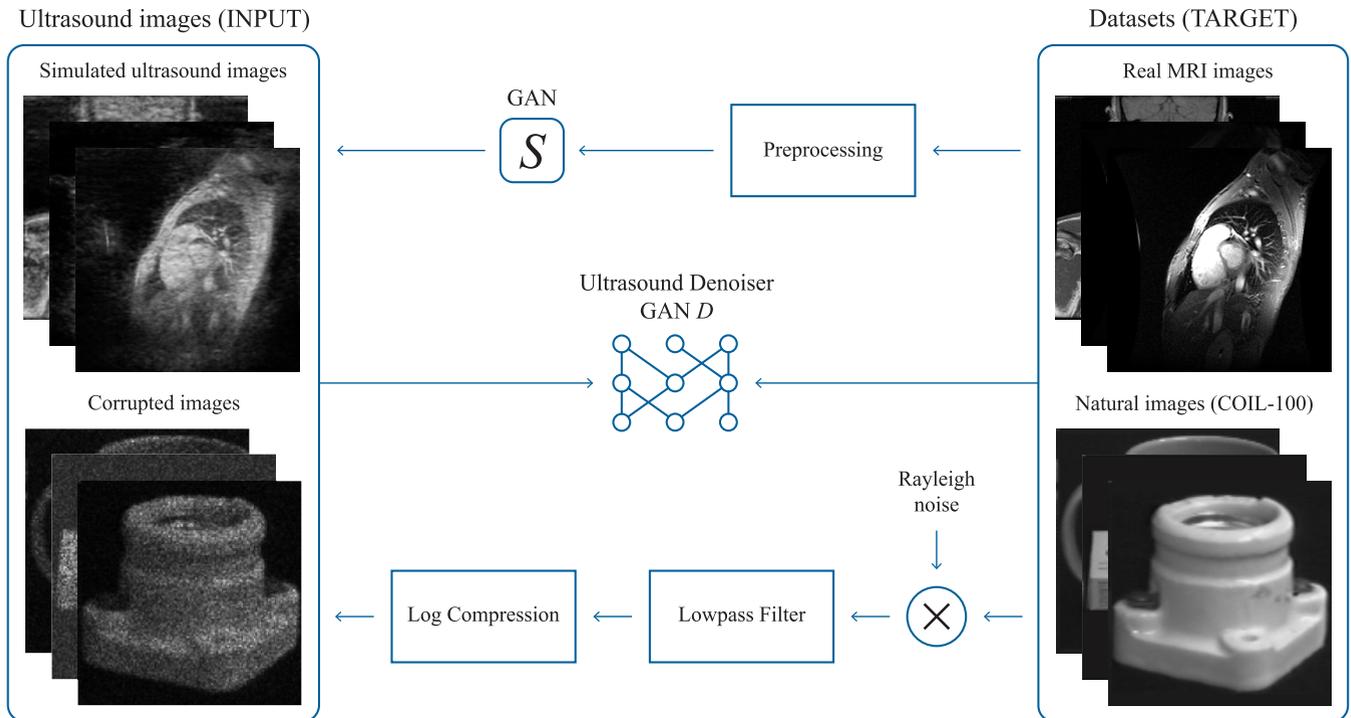
FIGURE 4. Training of *D*. Each MRI image from the dataset was preprocessed (quantization to 3-bit and median filtering (x2)) to obtain its phantom representation, used as input for the trained *S*, producing its synthetic ultrasound version. Additionally, each natural image from the COIL-100 dataset was corrupted with speckle noise and log-compressed, obtaining a noisy result.

reduced and filtered as described in Section II-B. The labeled image can now be inputted into *S*, producing the desired set of simulated ultrasound images when applied to all 375 original MRI images.

Second, the GAN must learn the appropriate statistical models for speckle noise. For this, a set of 375 natural images from the COIL-100 dataset [31] were contaminated with noise, following the sequence of operations illustrated on the bottom half of Fig. 4. First, grayscale images of natural objects were selected from the COIL-100 dataset [31]. The pixel values were normalized to one (by dividing by the maximum pixel value), and then multiplied pixel-wise by random values drawn from a Rayleigh distribution with a scale factor of 1. Then, to simulate the interpolation process performed by commercial ultrasound equipment the image obtained in the previous step was decimated by a factor of 2 along both axes, and then interpolated back to the original size using bilinear interpolation. This operation works as a lowpass filter. Finally, logarithmic compression was applied to each pixel (i, j) to compress the image obtained in the previous step, denoted as k, using

$$y(i, j) = \alpha \log(k(i, j) + 1) + \beta \,, \quad (2)$$

according to the B-mode image generation model in [4]. The parameters α and β in this step were randomly chosen between the values of 10 and 50 for α and 0 and 50 for β. Thus, several different equipment settings were taken into account and presented to the network. This sequence of

TABLE 1. Hyperparameters and information related to training, for the two GANs.

GAN	Train / Val. split	Epochs	λ	Initial gain	Learning rate	Training time
Simulator, S	500 / 170	700	100	0.02	10^{-4}	17 hours
Denoiser, D	600 / 280	550	100	0.02	10^{-4}	15 hours

operations produced a set of paired images of real-life objects and their statistically simulated B-mode counterparts, with the parameter variation mimicking the real-life scenario that these parameters are not typically known.

Hence, training the pix2pix model with the described combined set as per the scheme shown in Fig. 4 allowed it to learn typical features of MRI images (using the paired MRI - simulated ultrasound subset) and the common statistical models for speckle noise (using the natural - corrupted images subset). After training *D*, it can be applied to a noisy B-mode ultrasound image to obtain a denoised (despeckled) image as its output, resembling a MRI image, as illustrated in Fig. 1.

D. EXPERIMENTAL SETUP
The setup used to train the pix2pix networks was Google Colab environment with GPU and PyTorch. The default U-Net architecture and patch GAN were used for the generative and discriminator networks, respectively, for both the simulator (*S*) and denoising (*D*) networks. The set of hyperparameter values, train/validation image split, number of epochs, and total training times for each of the networks

are presented in Table 1. Colab was also used on the same system to perform statistical evaluations and apply histogram matching to assess the results of ultrasound simulation and denoising.

The image preprocessing operations, namely the compression of the ultrasound and MRI images to 3 bits and corruption of the natural images with Rayleigh noise was done in MATLAB version R2020a on a Windows Intel i7 system with 8 GB of RAM.

E. EVALUATION CRITERIA

Speckle noise in the RF domain is a multiplicative type of noise, following the Rayleigh distribution [4]. According to this image formation model, the ultrasound RF image y_{RF} ($N \times M$ pixels) which is related to the respective log-compressed B-mode image, y through (1), is corrupted by a truly multiplicative model of noise. According to this model, the RF speckled image is obtained by multiplying the noise free image, x, by a statistically independent field of Rayleigh distributed random variables, η. Thus, having the clean x, and the noisy RF image, y_{RF}, the Rayleigh noise field can be obtained by pixel-wise division, $\eta(i,j) = y_{RF}(i,j)/x(i,j)$ [34]. The normalized histogram of this noise field, h, was compared against the analytical Rayleigh distribution fitted to the data, with the following Probability Density Function (PDF)

$$p(\eta) = \frac{\eta}{\sigma^2} e^{\frac{-\eta^2}{2\sigma^2}},$$

where σ is the distribution parameter [2]. This parameter defines the position of the maximum of this function, which should coincide or at least lay on a small neighborhood around the (experimental) maximum of the noise field histogram to guarantee the best fit and lower $KL(p||h)$. The Kullback-Leibler divergence [35] between the analytical and experimental density functions, $KL(p||h)$, was computed (in dB) to assess the similarity of both distributions, using

$$KL(p\|h) = \sum_i p(b_i) \log \left(\frac{p(b_i)}{h_i} \right), \qquad (3)$$

where h_i and b_i are the bin counts and centers, respectively, and i represents the i'th intensity value bin.

The performance of the proposed denoising method was tested by comparing its results with those obtained with other state-of-the-art algorithms, such as RegGAN [30], De-QuIP [18], OBNLM [14], and TVHTVM [16]. For this, each method was applied on the test dataset of 148 MRI images of longitudinal and transversal slices of the human heart dataset [32]. Each test MRI image (ground truth) was used to generate a speckled ultrasound version with the ultrasound simulator proposed herein. This synthetic ultrasound image is the input of the proposed denoiser and the result is compared with the original noise free MRI image. In order to compare images with different dynamic ranges and intensity distributions, the Nyúl's histogram matching normalization technique was applied to the set of MRI,

TABLE 2. Average values obtained for the reconstruction quality evaluation metrics applied to the test dataset. MAE = Mean Average Error; RMSE = Root Mean Squared Error; PSNR = Peak Signal-to-Noise Ratio; BRISQUE = Blind/Referenceless Image Spatial Quality Evaluator. Values in bold correspond to the best results for each metric.

Method \ Metric	MAE	RMSE	PSNR (dB)	BRISQUE
RegGAN	15.4	25.2	20.1	43.3
De-QuIP	8.14	11.0	27.4	36.2
OBNLM	8.26	11.1	27.3	37.3
TVHTVM	8.21	11.1	27.3	50.7
Proposed	**5.83**	**8.90**	**29.2**	**18.2**

noisy and denoised images [36] using the Python package `intensity-normalization` (version 2.2.4) [37]. This harmonization technique aims to align MRI images acquired from different scanners and imaging protocols, ensuring that variations in image acquisition do not confound the results and allowing more accurate and reliable comparisons. Normalized images could now be visually compared and using the four objective metrics adopted in this work to assess the results. Denoting x as the original (MRI) image and $\hat{x}$ as the denoised image, the metrics extracted were the Mean Average Error (MAE), Root Mean Squared Error (RMSE) and Peak Signal to Noise Ratio (PSNR) [38], according to the following equations, respectively:

$$MAE = \frac{1}{MN} \sum_{i=1}^{M} \sum_{j=1}^{N} |x(i,j) - \hat{x}(i,j)|, \qquad (4)$$

$$RMSE = \sqrt{MSE}, \qquad (5)$$

$$PSNR = 10 \log_{10} \left(\frac{\max \hat{x}}{MSE} \right), \qquad (6)$$

where MSE is the mean squared error, defined as

$$MSE = \frac{1}{MN} \sum_{i=1}^{M} \sum_{j=1}^{N} (x(i,j) - \hat{x}(i,j))^2. \qquad (7)$$

The last metric considered was the Blind/Referenceless Image Spatial Quality Evaluator (BRISQUE) [39], computed with the Python package `brisque` (version 0.0.15). The average values of these metrics were used to quantitatively compare the performance of each denoising method on this test dataset.

III. RESULTS AND DISCUSSION

In this section, experimental results regarding the generation of synthetic ultrasound images (using S) and the denoising of B-mode ultrasound images (using D and other state-of-the-art methods) are presented and discussed.

A. GENERATION OF SYNTHETIC B-MODE ULTRASOUND DATA

The process of obtaining a simulated B-mode ultrasound image from an input MRI image applied to our ultrasound simulator is shown in Fig. 2. This example shows that S is capable of generating characteristics typical of real ultrasound images, particularly the speckled noise profile and

(a)

(b)
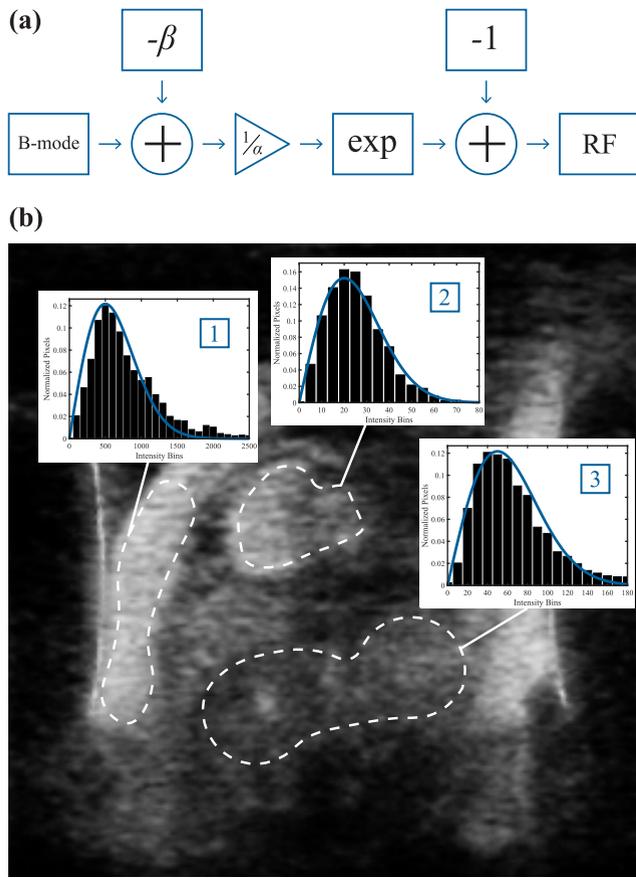
FIGURE 5. (a) Block diagram illustrating the decompression process to convert a B-mode image to the RF domain [4] (see (1)). With the image now in the RF domain, the noise field can be estimated by element-wise division of both noisy (simulated ultrasound in RF domain) and original (MRI) images. (b) Simulated B-mode obtained with the ultrasound simulator from the source MRI image in Fig. 2. The histograms refer to the intensity distribution of the noise field estimated within three regions of interest (1, 2 and 3). The fitted analytical Rayleigh distributions are also plotted in blue. Regions of interest contain distinct organs: (1) ribs, with $KL_1(p||h) = 1.24 \cdot 10^{-1}$ dB; (2) heart, with $KL_2(p||h) = 7.95 \cdot 10^{-2}$ dB; (3) other organs, with $KL_3(p||h) = 8.86 \cdot 10^{-2}$ dB. h refers to the histogram of the estimated noise field, while p refers to the fitted analytical Rayleigh distribution.

low overall resolution, since the contours lose definition and anatomical structures are roughly preserved.

Our ultrasound simulator has an advantage over other ultrasound image conversion strategies such as Field II [40] in terms of computational speed. To produce one ultrasound image, Field II took 12 hours on a MATLAB cluster of ten 3 GHz CPUs, whereas our simulator took on average 0.25 seconds per image.

Furthermore, the synthetic RF-ultrasound image confirms to the statistical characteristics of the modality. Speckle noise in ultrasound images is multiplicative and follows a Rayleigh distribution when in the RF domain [4]. Since S outputs images in B-mode, an RF conversion was performed as described in Section II-B and illustrated in Fig. 5 (a). The mapping parameters α and β were estimated as a function of the simulated B-mode image according to [33]. Since
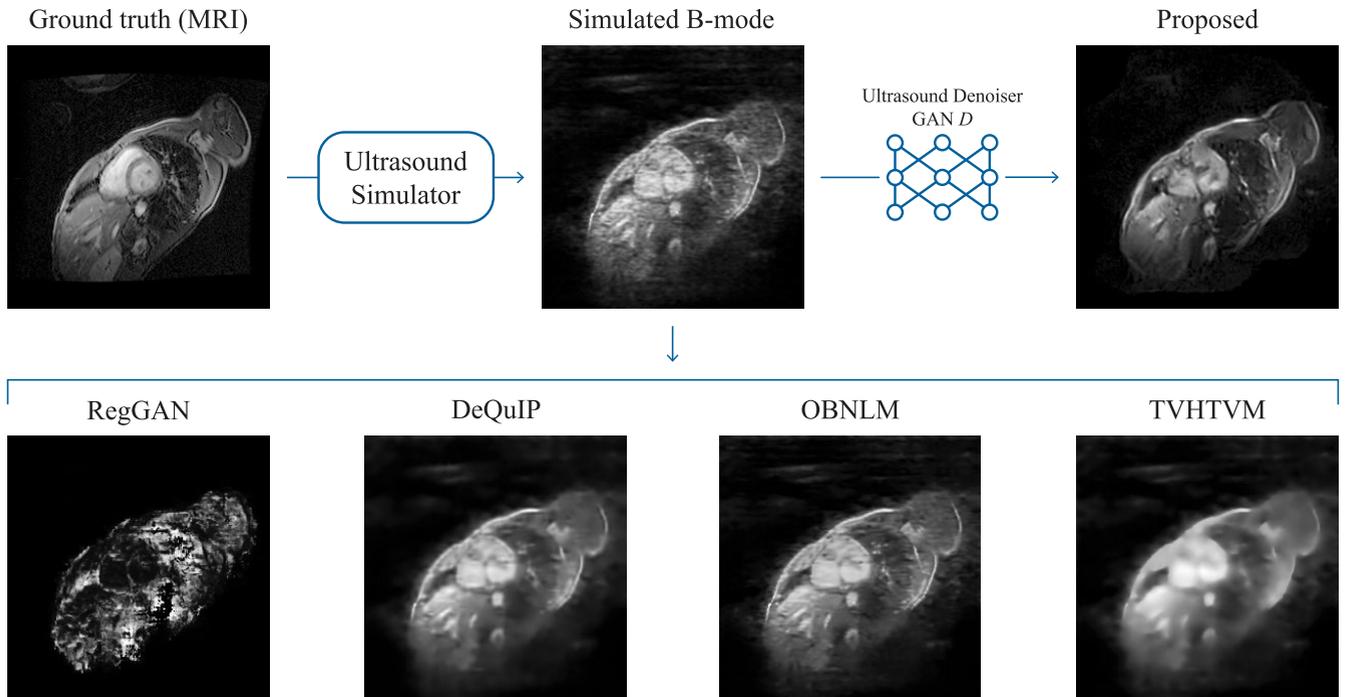
in the RF image the noise is pixel-wise multiplicative, the noise field can be estimated by element-wise division of the RF image and the original (MRI) image. To assess how close the intensity distribution of the noise field is from the analytical Rayleigh distribution, the PDF of the former was roughly estimated using a normalized histogram and compared to the latter by fitting the histogram data. To fit the analytical distribution, the intensity bin associated with the histogram's maximum was identified. Then, three distribution parameters were tested: the center of the intensity bin corresponding to the maximum value, as well as the centers of the intensity bins to the left and right of this particular bin. From this set of three distribution parameters, the one of the fitted analytical distribution was chosen to minimize the KL divergence. Fig. 5 (b) shows the distribution of the estimated noise field inside 3 regions of interest and their respective Rayleigh fits and KL divergences (computed using (3)). The estimated PDFs and the fitted Rayleigh distributions were similar. To confirm this finding, an uniform image with the same size as Figure 5 was created, where each pixel intensity was generated from a Rayleigh distribution. When analyzed within similar regions of interest, the KL divergence between this noise field's histogram and the fitted Rayleigh distribution was around 10^{-2} (-40 dB). From this, we confirmed quantitatively that the values obtained in Figure 5 ($KL_1(p||h) = 1.24 \cdot 10^{-1}$ (-18 dB), $KL_2(p||h) = 7.95 \cdot 10^{-2}$ (-22 dB) and $KL_3(p||h) = 8.86 \cdot 10^{-2}$ (-21 dB) are considered low.

### B. COMPARISON WITH OTHER DENOISING METHODS

To be able to compare the performances of the different methods, Nyúl's histogram matching method [36] was applied to the set of ground truth, noisy and the five denoised versions (one from the proposed and four from the other methods described in the literature). This was a crucial step to guarantee that different sources were appropriately aligned in terms of intensity values. These different sources refer to the proposed and other methods used for comparison, which interfered with the dynamic range of the input image due to different image processing strategies, causing different histogram shifts, posteriorly aligned using Nyúl's normalization [36].

In this section, we show the comparison results with the following state-of-the-art methods described in the literature: RegGAN [30], De-QuIP [18], OBLNM [14] and TVHTVM [16].

The average values for the MAE, RMSE, PSNR (computed with (4), (5) and (6), respectively) and BRISQUE metrics were evaluated using the whole test dataset of 148 images, and the results are shown in Table 2. MAE and RMSE measure pixel-to-pixel intensity similarities. Low values point to similar intensity distributions of the original and denoised images. PSNR, in (6), measures in dB the ratio between the maximum value of the denoised image and the MSE, comparing the denoised image with the original one. High PSNR values indicate successful denoising processes [38].

Ground truth (MRI)   Simulated B-mode   Proposed

Ultrasound Simulator

Ultrasound Denoiser GAN *D*

RegGAN   DeQuIP   OBNLM   TVHTVM

FIGURE 6. Pipeline to compare the performance of the various denoisers applied to an example image of the test dataset. With the ultrasound simulator, the ground truth MRI image was preprocessed and converted to its synthetic B-mode ultrasound representation using the GAN *S*. Then, using this image as input, the proposed denoising method was tested (using the *D* network) and compared with the clean images obtained using the other 4 methods. The set of images shown was subject to Nyúl histogram matching to allow their comparison.

(a)   (b)   (c)

(d)   (e)   (f)

(g)   (h)
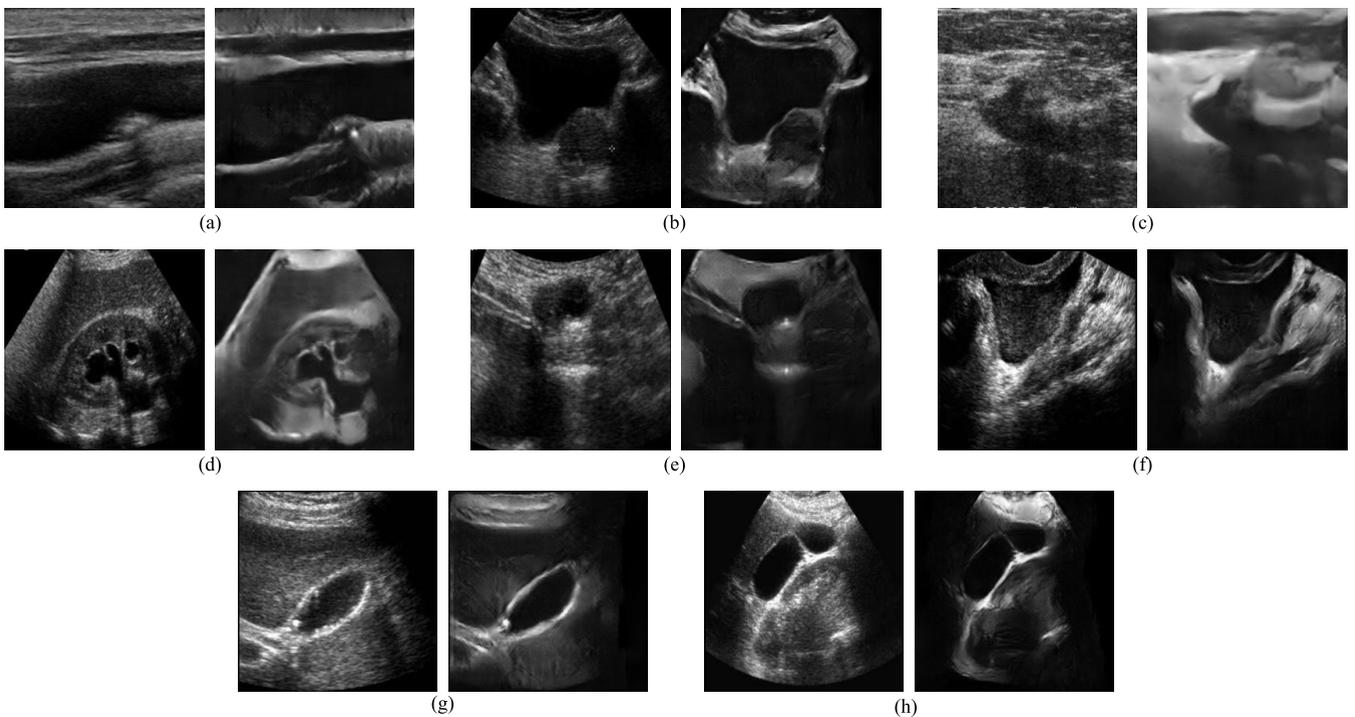
FIGURE 7. Eight examples of real B-mode ultrasound images and their denoised versions obtained using the proposed method: (a) carotid artery, (b) bladder, (c) breast, (d) pancreas, (e,f) uterus, (g,h) gallbladder. Denoised images resulting from the other 4 state-of-the-art methods can be found in Appendix A.

Finally, BRISQUE is a no-reference image quality assessment algorithm designed to evaluate the perceptual quality of images without relying on a reference (original) image: low values indicate better image quality [39]. Table 2
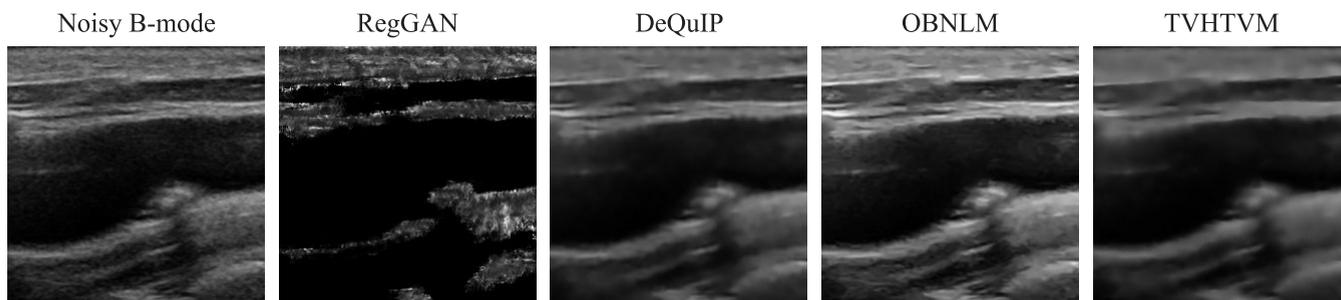
FIGURE 8. Real ultrasound image of a carotid artery, along with the denoised versions obtained from each studied state-of-the-art method. Images were normalized with Nyúl histogram matching.
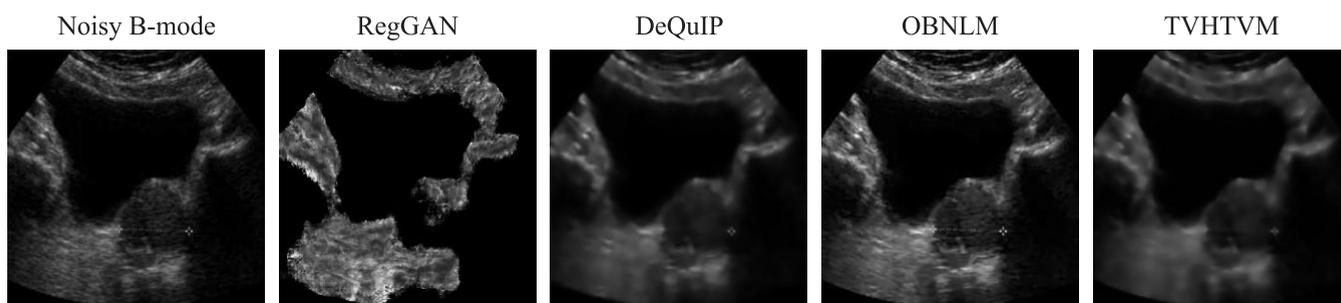
FIGURE 9. Real ultrasound image of a bladder, along with the denoised versions obtained from each studied state-of-the-art method. Images were normalized with Nyúl histogram matching.

suggests that the proposed method outperformed all the others in the four evaluations metrics adopted for comparison purposes.

Fig. 6 depicts an example MRI image (top left) from the test dataset, then preprocessed and converted to its synthetic B-mode representation (with the ultrasound simulator) in order to test the proposed and the other 4 state- of-the-art denoising methods. Notably, this example showed, once again, the ability of the ultrasound simulator to generate realistic synthetic ultrasound images. These synthetic ultrasound images were then used as input to test the denoising performance of the proposed and other methods. Indeed, the other methods did not achieve the same despeckle degree and contour maintenance capacity as compared to the proposed method. RegGAN, another GAN designed for medical image-to-image translation and noise removal in medical images, was the method that surprisingly produced the worst results. With this method, the distribution of intensities was completely altered compared to the original image, as reinforced by the high MSE and RMSE. The anatomical contours lost coherence and the image lacked overall quality, as indicated by the high BRISQUE score and low PSNR. It must be noted that the trained model provided by the authors was used, without training on our dataset. The results of the remaining three methods were very similar to each other quantitatively, and none outperformed the proposed one. In these examples, De-QuIP, a despeckling algorithm, seemed to spread the noise by smearing the image, rather than removing it. OBNLM did not remove speckle

noise well enough, on visual inspection. Finally, TVHTVM was also not capable of removing speckle noise and keep smooth organ contours.

The proposed method stands out from the others due to its ability to recover and highlight the integrity of some organs and other anatomical structures, while at the same time drastically reduced speckle noise. Although it is very difficult to recover a resolution at the level typical of MRI images, the improvement in quality of denoised images compared to synthetic ultrasound images is evident, as proven by quantitative evaluation criteria and by visual inspection.

C. DENOISING RESULTS
The proposed denoising method aimed to translate noisy (speckled) input ultrasound images into clean images resembling MRI. Fig. 7 depicts examples of original ultrasound images and their denoised versions, obtained with the proposed method. Notably, the coherent morphological structures appeared clear whereas the speckle noise pattern was greatly reduced across all examples. Ultrasound images included several organs of the human body, including the carotid artery, bladder, breast, pancreas, uterus, and gallbladder, thus highlighting the robustness of the denoiser GAN.

The proposed pipeline was capable of producing denoised images of good visual quality, in the absence of training data for the modality without knowing parameters associated with the ultrasound sensor.
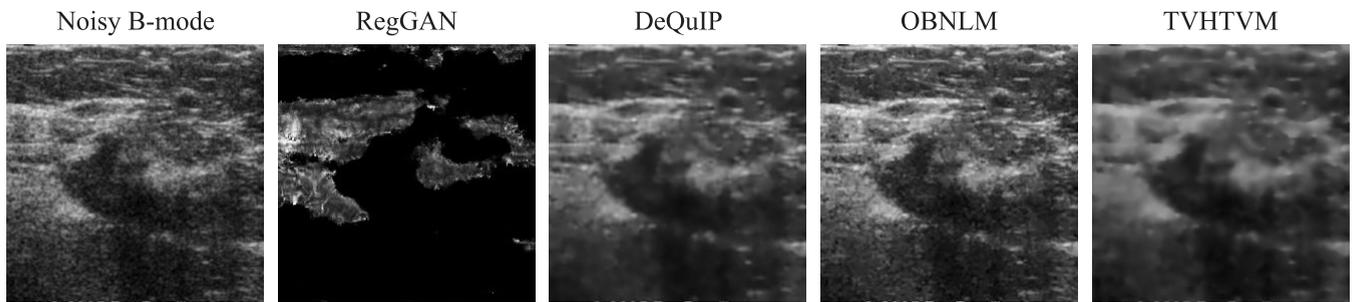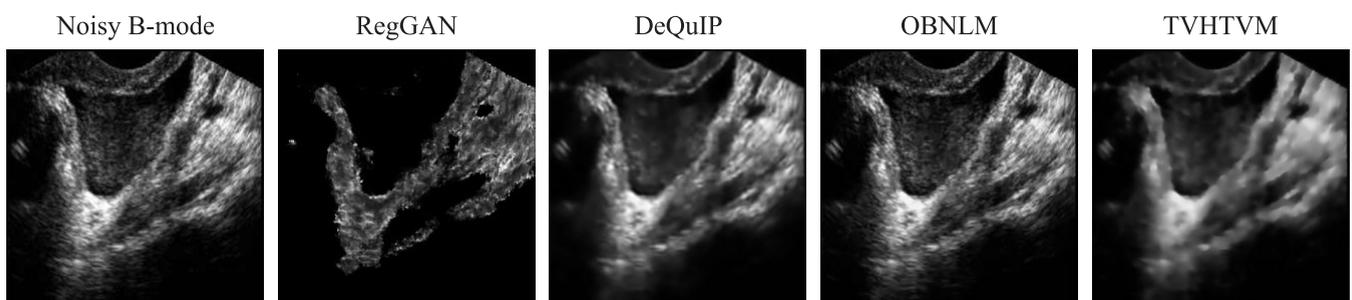
FIGURE 10. Real ultrasound image of a breast, along with the denoised versions obtained from each studied state-of-the-art method. Images were normalized with Nyúl histogram matching.

FIGURE 11. Real ultrasound image of a pancreas, along with the denoised versions obtained from each studied state-of-the-art method. Images were normalized with Nyúl histogram matching.
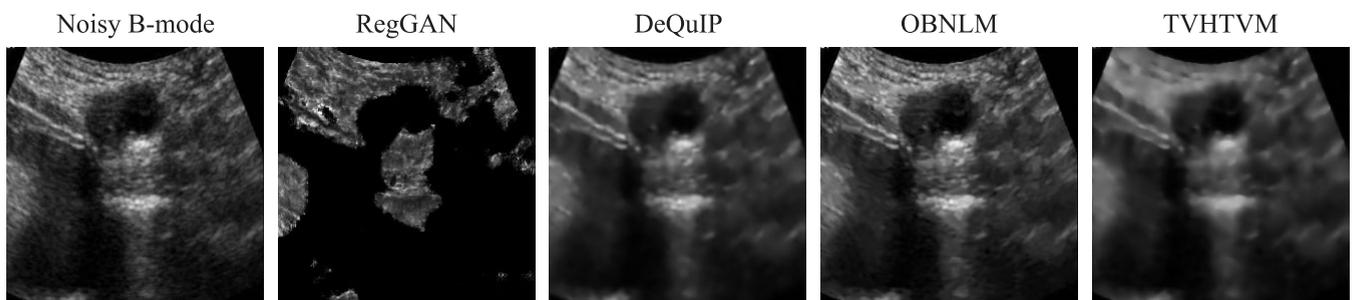
FIGURE 12. Real ultrasound image of a uterus, along with the denoised versions obtained from each studied state-of-the-art method. Images were normalized with Nyúl histogram matching.
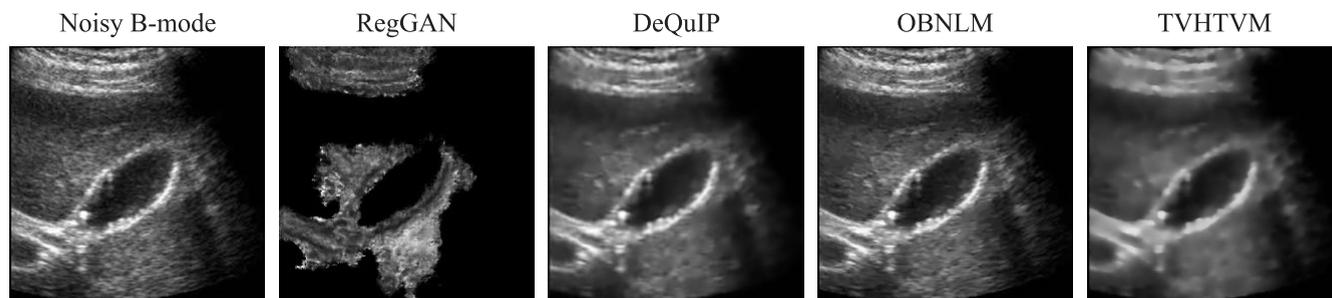

FIGURE 13. Real ultrasound image of a uterus, along with the denoised versions obtained from each studied state-of-the-art method. Images were normalized with Nyúl histogram matching.

IV. CONCLUSION

This paper proposed a trained pix2pix GAN model for B-mode ultrasound image despeckling. The GAN was trained with datasets of paired natural - speckled images and MRI - simulated B-mode ultrasound images to learn the statistical model of speckle noise and MRI features. Thus,

FIGURE 14. Real ultrasound image of a gallbladder, along with the denoised versions obtained from each studied state-of-the-art method. Images were normalized with Nyúl histogram matching.
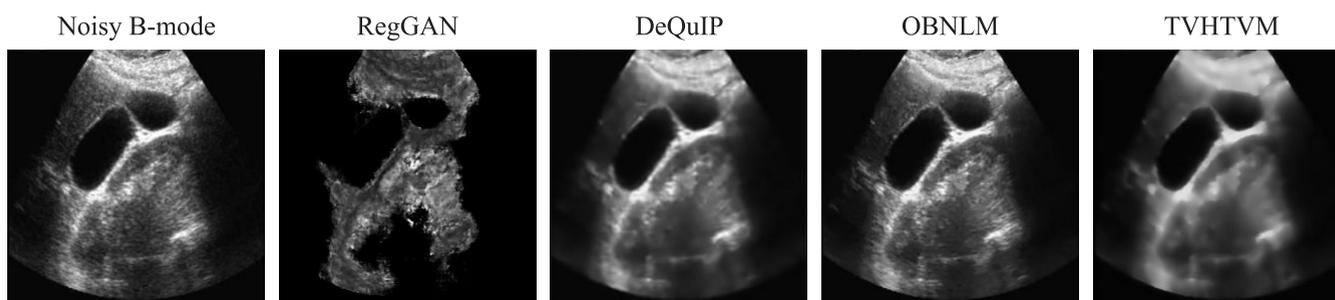
FIGURE 15. Real ultrasound image of a gallbladder, along with the denoised versions obtained from each studied state-of-the-art method. Images were normalized with Nyúl histogram matching.

this methodology comprised a GAN-based denoising method for speckled images in a cross modality manner between ultrasound and MRI. Denoising results of several real and simulated ultrasound images show that the GAN is capable of efficiently remove speckle noise, while maintaining the integrity of anatomical structures and avoiding reconstruction artifacts. It generates outputs that resemble typical MRI images due to the cross modality transfer property. The proposed method also outperforms other 4 state-of-the-art methods under several reconstruction quality metrics. Concluding, this denoising method may ultimately aid diagnosis, clinical decision-making or further image processing tasks. Generalizing the proposed pipeline to use other methods for generating synthetic images such as stable diffusion, transformers, and unpaired approaches will be addressed in future work.

APPENDIX A
DENOISING OF REAL ULTRASOUND IMAGES WITH STATE-OF-THE-ART METHODS
See Figures 8–15.

REFERENCES
[1] T. Eltoft, "Modeling the amplitude statistics of ultrasonic images," *IEEE Trans. Med. Imag.*, vol. 25, no. 2, pp. 229–240, Feb. 2006.

[2] O. Michailovich and A. Tannenbaum, "Despeckling of medical ultrasound images," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 53, no. 1, pp. 64–78, Jan. 2013.

[3] J. A. Noble and D. Boukerroui, "Ultrasound image segmentation: A survey," *IEEE Trans. Med. Imag.*, vol. 25, no. 8, pp. 987–1010, Aug. 2006.

[4] J. Seabra and J. Sanches, "Modeling log-compressed ultrasound images for radio frequency signal recovery," in *Proc. 30th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Aug. 2008, pp. 426–429.

[5] J. C. R. Seabra and J. M. R. Sanches, "A 3D graph-cut based algorithm for evaluating carotid plaque echogenicity and texture," in *Recent Advances in Biomedical Engineering*. London, U.K.: IntechOpen, 2009.

[6] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-D transform-domain collaborative filtering," *IEEE Trans. Image Process.*, vol. 16, no. 8, pp. 2080–2095, Aug. 2007.

[7] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.

[8] E. Akleman, "Deep learning," *Computer*, vol. 53, no. 9, p. 17, Sep. 2020.

[9] J. Xie, L. Xu, and E. Chen, "Image denoising and inpainting with deep neural networks," in *Proc. 25th Int. Conf. Neural Inf. Process. Syst.*, vol. 1. Lake Tahoe, NV, USA: Curran Associates, 2012, pp. 341–349.

[10] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proc. CVPR*, 2017, pp. 1125–1134.

[11] Y. Yu and S. T. Acton, "Speckle reducing anisotropic diffusion," *IEEE Trans. Image Process.*, vol. 11, no. 11, pp. 1260–1270, Nov. 2002.

[12] P. Coupe, P. Hellier, C. Kervrann, and C. Barillot, "Nonlocal means-based speckle filtering for ultrasound images," *IEEE Trans. Image Process.*, vol. 18, no. 10, pp. 2221–2229, Oct. 2009.

[13] Y. Gan, E. Angelini, A. Laine, and C. Hendon, "BM3D-based ultrasound image denoising via brushlet thresholding," in *Proc. IEEE 12th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2015, pp. 667–670.

[14] P. Coupe, P. Yger, S. Prima, P. Hellier, C. Kervrann, and C. Barillot, "An optimized blockwise nonlocal means denoising filter for 3-D magnetic resonance images," *IEEE Trans. Med. Imag.*, vol. 27, no. 4, pp. 425–441, Apr. 2008.

[15] L. Zhu, C.-W. Fu, M. S. Brown, and P.-A. Heng, "A non-local low-rank framework for ultrasound speckle reduction," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 20170, pp. 493–501.

[16] S. Wang, T.-Z. Huang, X.-L. Zhao, J.-J. Mei, and J. Huang, "Speckle noise removal in ultrasound images by first- and second-order total variation," *Numer. Algorithms*, vol. 78, no. 2, pp. 513–533, Jun. 2018.

[17] K. Mei, B. Hu, B. Fei, and B. Qin, "Phase asymmetry ultrasound despeckling with fractional anisotropic diffusion and total variation," *IEEE Trans. Image Process.*, vol. 29, pp. 2845–2859, 2020.

[18] S. Dutta, A. Basarab, B. Georgeot, and D. Kouamé, "A novel image denoising algorithm using concepts of quantum many-body theory," *Signal Process.*, vol. 201, Dec. 2022, Art. no. 108690.

[19] D. Perdios, A. Besson, M. Arditi, and J.-P. Thiran, "A deep learning approach to ultrasound image recovery," in *Proc. IEEE Int. Ultrason. Symp. (IUS)*, Sep. 2017, pp. 1–4.

[20] S. Vedula, O. Senouf, A. M. Bronstein, O. V. Michailovich, and M. Zibulevsky, "Towards CT-quality ultrasound imaging using deep learning," 2017, *arXiv:1710.06304*.

[21] M. Tabassian, X. Hu, B. Chakraborty, and J. D'hooge, "Clutter filtering using a 3D deep convolutional neural network," in *Proc. IEEE Int. Ultrason. Symp. (IUS)*, Oct. 2019, pp. 2114–2117.

[22] H. Li, T. Mezheritsky, L. Vazquez Romaguera, and S. Kadoury, "3D B-mode ultrasound speckle reduction using deep learning for 3D registration applications," 2020, *arXiv:2008.01147*.

[23] E. Mor and A. Bar-Hillel, "A unified deep network for beamforming and speckle reduction in plane wave imaging: A simulation study," *Ultrasonics*, vol. 103, Apr. 2020, Art. no. 106069.

[24] O. Huang, W. Long, N. Bottenus, M. Lerendegui, G. E. Trahey, S. Farsiu, and M. L. Palmeri, "MimickNet, mimicking clinical image Post-processing under black-box constraints," *IEEE Trans. Med. Imag.*, vol. 39, no. 6, pp. 2277–2286, Jun. 2020.

[25] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2242–2251.

[26] F. Dietrichson, E. Smistad, A. Ostvik, and L. Lovstakken, "Ultrasound speckle reduction using generative adversial networks," in *Proc. IEEE Int. Ultrason. Symp. (IUS)*, Oct. 2018, pp. 1–4.

[27] L. He, B. Peng, T. Yang, and J. Jiang, "An application of super-resolution generative adversary networks for quasi-static ultrasound strain elastography: A feasibility study," *IEEE Access*, vol. 8, pp. 65769–65779, 2020.

[28] S. Goudarzi, A. Asif, and H. Rivaz, "Fast multi-focus ultrasound image recovery using generative adversarial networks," *IEEE Trans. Comput. Imag.*, vol. 6, pp. 1272–1284, 2020. [Online]. Available: https://ieeexplore.ieee.org/document/9178454

[29] S. Cammarasana, P. Nicolardi, and G. Patanè, "A universal deep learning framework for real-time denoising of ultrasound images," 2021, *arXiv:2101.09122*.

[30] L. Kong, C. Lian, D. Huang, Z. Li, Y. Hu, and Q. Zhou, "Breaking the dilemma of medical image-to-image translation," in *Proc. 35th Conf. Neural Inf. Process. Syst.*, 2021, pp. 1–15.

[31] S. A. Nene, S. K. Nayar, and H. Murase, *Columbia Object Image Library*, document CUCS-006-96, 1996.

[32] A. Andreopoulos and J. K. Tsotsos, "Efficient and generalizable statistical models of shape and appearance for analysis of cardiac MRI," *Med. Image Anal.*, vol. 12, no. 3, pp. 335–357, Jun. 2008.

[33] J. M. Sanches and J. S. Marques, "Compensation of log-compressed images for 3-D ultrasound," *Ultrasound Med. Biol.*, vol. 29, no. 2, pp. 239–253, Feb. 2003.

[34] J. Seabra and J. Sanches, "Ultrasound speckle/despeckle image decomposition for tissue analysis," in *Ultrasound Imaging: Advances and Applications*. Cham, Switzerland: Springer, 2012.

[35] R. Viñals and J.-P. Thiran, "A KL divergence-based loss for in vivo ultrafast ultrasound image enhancement with deep learning," *J. Imag.*, vol. 9, no. 12, p. 256, Nov. 2023.

[36] L. G. Nyul, J. K. Udupa, and X. Zhang, "New variants of a method of MRI scale standardization," *IEEE Trans. Med. Imag.*, vol. 19, no. 2, pp. 143–150, Feb. 2000.

[37] J. C. Reinhold, B. E. Dewey, A. Carass, and J. L. Prince, "Evaluating the impact of intensity normalization on MR image synthesis," *Med. Imag. 2019: Image Process.*, vol. 10949, Mar. 2019, Art. no. 109493H.

[38] P. S. Hiremath, P. T. Akkasaligar, and S. Badiger, "Speckle noise reduction in medical ultrasound images," in *Advancements and Breakthroughs in Ultrasound Imaging*. Chennai, India: InTech, 2013.

[39] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Trans. Image Process.*, vol. 21, no. 12, pp. 4695–4708, Dec. 2012.

[40] J. A. Jensen, "Simulation of advanced ultrasound systems using field II," in *Proc. 2nd IEEE Int. Symp. Biomed. Imaging: Macro. Nano*, Apr. 2004, pp. 636–639.

DIOGO FRÓIS VIEIRA received the B.Sc. and M.Sc. degrees (first semester from the Technical University of Denmark) in biomedical engineering from Instituto Superior Técnico, University of Lisbon, Portugal (IST), in 2020 and 2022, respectively. He is currently pursuing the Ph.D. degree with the Institute for Systems and Robotics, IST, with a focus on bioimaging processing of cytoskeletal structures in cancer and living cells.

During his education, he collected several academic merit and excellence diplomas and was a Teaching Assistant with the Bioimaging and Biosignals Curricular Units, IST. His current research interests include image processing of biomedical images, biomedical signal processing and instrumentation, and machine learning and medicine.

AFONSO RAPOSO received the master's degree in biomedical engineering from Instituto Superior Tècnico (IST), Universidade de Lisboa. He is a dedicated professional. During his academic journey, he immersed himself in the innovative fields of telemedicine and mobile health, exploring the potential of smartphones, and wearables to transform healthcare delivery. He contributed to this work during his studies at IST and documented its initial stages in some scholarly articles accessible through Google Scholar. Currently, he contributes his diverse skill set as a Full Stack Engineer with Prezi Inc., where he continues to blend technical expertise with a passion for advancing digital solutions.

ANTÓNIO AZEITONA received the master's degree in biomedical engineering from Instituto Superior Técnico, Universidade de Lisboa. The work for the thesis focused on generative adversarial networks (GAN) and their application in reducing noise and increasing datasets. The initial stage of this work was presented in the form of a poster at the 2021 Annual Portuguese Conference on Pattern Recognition (RECPAD21).

MANYA V. AFONSO received the B.E. degree in electronics and telecommunication engineering from Goa University, India, in 2003, the M.Tech. degree in communication engineering from IIT Delhi, New Delhi, India, in 2005, and the Ph.D. degree from Instituto Superior Tecnico (IST), Lisbon, Portugal, in 2011.

He was a Researcher with the Institute of Telecommunications (IT) and then a Postdoctoral Researcher with the Institute for Systems and Robotics (ISR), Lisbon. He is currently a Senior Researcher with the Biometric (Applied Mathematics and Statistics) and the Vision and Robotics Groups, Wageningen University and Research, The Netherlands. His research interests include image processing and analysis, inverse problems, optimization, machine learning, computer vision, and deep learning, with an emphasis on application to biomedical image analysis, plant science, and agriculture.

J. SANCHES is currently a Full Professor with the Department of Bioengineering, IST, Universidade de Lisboa. He is also a Senior Researcher and the Deputy President for the Scientific Affairs of the Institute for Systems and Robotics (ISR). His work has been focused on statistical signal and image processing of biomedical data mainly in cancer research and in the development of methods for diagnosis and follow-up of chronic pathologies, such as mental health and cardiovascular diseases. He has published approximately 200 international publications. He is a Senior Member of the IEEE Engineering in Medicine and Biology Society (EMBS).

● ● ●

LUÍS MENDES PEDRO received the M.D. and Ph.D. degrees from the Medical School, University of Lisbon, in 1987 and 2003, respectively. Since 1997, he has been a Specialist in vascular surgery with active clinical experience. In 2000, he achieved approval in the European Board of Vascular Surgery Examination and became a fellow of the European Board of Vascular Surgery (FEBS). He is currently the Head of the Vascular Surgery Department, ULS Santa Maria, Lisbon (Hospital Santa Maria); a Full Professor of ''Vascular Surgery'' and ''Introduction to Clinical Medicine'' with the Medical School, University of Lisbon; and the Deputy President of the Scientific Council. He is also the President of the Portuguese Society of Phlebology and a member of Portuguese Medical Association, Portuguese Society for Angiology and Vascular Surgery (SPACV) (Past President), European Society for Vascular Surgery (ESVS), the International Society of Endovascular Specialists (ISES), and the International Union of Angiology (IUA). He was an Editorial Board Member of *European Journal of Vascular and Endovascular Surgery*, from 2006 to 2010. He was the Past Editor-in-Chief of the journal *Angiologia e Cirurgia Vascular*, from 2006 to 2017.