Fast Multi-Focus Ultrasound Image Recovery Using Generative Adversarial Networks

Sobhan Goudarzi, Amir Asif *Senior Member, IEEE,* and Hassan Rivaz *Senior Member, IEEE*

***Abstract*—In conventional ultrasound (US) imaging, it is common to transmit several focused beams at multiple locations to generate a multi-focus image with constant lateral resolution throughout the image. However, this method comes at the expense of a loss in temporal resolution, which is important in applications requiring both high-frame rate and constant lateral resolution. Moreover, relative motions of the target with respect to the probe often exist due to hand tremors or biological motions, causing blurring artifacts in the multi-focus image. This paper introduces a novel approach for multi-focus US image recovery based on Generative Adversarial Network (GAN) without a reduction in the frame-rate. Herein, a mapping function between the single-focus US image and multi-focus version for having a constant lateral resolution everywhere is estimated through different GANs. We use adversarial loss functions in addition to Mean Square Error (MSE) to generate more realistic ultrasound images. Moreover, we use the boundary seeking method for improving the stability of training, which is currently the main challenge in using GANs. Experiments on simulated and real phantoms as well as on *ex vivo* data are performed. Results confirm that having both adversarial loss function and boundary seeking training provides better results in terms of the mean opinion score test. Furthermore, the proposed method enhances the resolution and contrast indexes without sacrificing the frame-rate. As for the comparison with other approaches which are not based on NNs, the proposed approach gives similar results while requiring neither channel data nor computationally expensive algorithms.**

***Index Terms*—Ultrasound imaging, focal point, frame-rate, GAN, adversarial loss.**

I. Introduction

THE main Ultrasound (US) imaging techniques are: (1) Classical focused transmission (also known as line-per-line acquisition) which is the transmit configuration used in current study; (2) Element-by-element transmissions synthetic aperture imaging [1], [2], and; (3) Plane-wave transmissions (also known as ultrafast imaging) [3]. Synthetic aperture imaging generally has a limited depth of penetration and also poor signal-to-noise ratio since it uses a single element for emission [4]. In plane-wave transmission, however, high frame-rate as well as optimal multi-focus quality can be accomplished. More specifically, plane-waves transmitted with different angels can be coherently compounded to reconstruct US images which are focused everywhere [3]. Notwithstanding, clinical application of this method is costly because it needs high data transfer bandwidth, powerful data acquisition cards, and powerful parallel processing units.

Focused transmission provides better SNR than synthetic aperture imaging and has a lower computational cost as compared to plane-wave. More specifically, in focused transmission, transmitted beams are focused in order to have higher intensity and better lateral resolution at a specific depth. Indeed, focusing means aligning the pressure fields of all elements of the aperture to simultaneously arrive at a specific field point [5]. Focusing can be done through a physically curved aperture or electronic beamforming. Focused beams have a complex bowtie shape with side lobes and grating lobes [6]. In classical focused transmission, it is assumed that received echoes are brought about by scatterers from within the main transmitted US beam. However, if there is a strong reflector outside of the main beam, it may cause detectable echoes for transducer and will be falsely displayed. This problem is called beam width artifact [7]. Hence, the narrower the transmitted beam, the lower the beam width artifacts.

When the beam is focused, the quality of the image is optimal at the focal point and progressively degrades away from it. Therefore, in order to preserve optimal lateral resolution everywhere along the axial direction, several beams focused at different depths are often transmitted. Consequently, the multi-focus US image can be recovered. However, this approach drastically reduces the frame-rate which is inversely proportional to the number of transmissions. Therefore, there is a trade-off between the lateral resolution and frame-rate in classical focused transmission. It has to be mentioned that when the depth of imaging is limited, image degradation due to beam divergence is limited. Therefore, if there is no clearly discernible target such as a cyst or hyperechoic region, the difference between the quality of single and multi-focus images is difficult to observe. Another issue arising in this method is the assumption of having no relative motion between the tissue and the probe while transmitting several beams. This assumption is not practical in several applications such as in imaging regions close to the heart or a major artery and in obstetric sonography. For example, in cardiac sonography, the motion blur is large even in between different lines, which has led to the advent of multi-line acquisition (MLA) methods [8]. Hand motion and tremor are additional sources of relative motions. Inspired by the success of deep learning algorithms, we propose a data-driven method for multi-focus line-per-line US imaging with only a single focused transmission and without a loss in frame-rate. More specifically, we train a Generative Adversarial Network (GAN) [9] to form a mapping function between non focused and focused US images.

Convolutional Neural Networks (CNNs) are able to effi-

Sobhan Goudarzi, Amir Asif, and Hassan Rivaz are with the Department of Electrical and Computer Engineering, Concordia University, Montreal, QC, H3G 1M8, Canada. Email: s_oudarz@ece.concordia.ca , amir.asif@concordia.ca , and hrivaz@ece.concordia.ca

This paper has supplementary downloadable material available at http://ieeexplore.ieee.org., provided by the author. The material includes some additional results. This material is 4.64 MB in size.

ciently extract necessary features from raw data, and there is no need to engineer hand-crafted features anymore [10]. CNNs have been successfully used in variety of applications such as classification, super resolution, denoising, and etc. Defining a proper objective function to minimize during the training phase is a critical factor that influences the performance of the network, and is currently an active field of research [11], [12], [13].

GANs address this issue by using CNNs to automatically learn an objective function appropriate for satisfying the specific task. More specifically, GANs consist of generator and discriminator networks, which compete with each other. Generally, the generator does a mapping from input space to a real desired space, and discriminator specifies the quality of generated data. Hence, the discriminator is the objective function for generator, and the generator tries to fool the discriminator by generating more realistic data [9]. Both networks are interestingly trained during the training process, which entails solving a minimax game to find the Nash equilibrium of these two competing networks.

Improving the training process and test performance of GANs is an active field of research [14]. Training dynamics of GANs were theoretically investigated in [15], which has led to several contributions in improving the training process [16], [17], or finding a working architecture [18] tailored for specific applications. Subsequently, Arjovsky et al. [19] exploited the concept of integral probability metric [20] and introduced Wasserstein GAN (WGAN). Although it resolved some issues, it has a limited success because of using weight clipping to enforce a Lipschitz constraint on the discriminator. This problem was solved by penalizing the norm of gradient of the discriminator over interpolation between generated and real data [21], [22]. Another notable contribution was proposed by Roth et al. [23] where a gradient norm penalty similar to [21] is introduced, except that there is no interpolation and f-divergences is instead used.

In spite of such important theoretical contributions, there is still no clear understanding on why the discriminator objective function is critical in stable training of GANs. Moreover, it has been shown that most of reviewed models can reach similar scores with non-saturating GAN introduced in [9], and there is no evidence that any of them consistently outperforms the non-saturating GAN. Using a different approach, another method for training GANs was proposed entitled Boundary-Seeking GANs (BSGANs) [24]. BSGAN is based on providing a policy gradient for training the generator that forces the generator to produce samples which are near the decision boundaries (i.e., the discriminator cannot distinguish real or generated data). In addition to better training behavior, BSGAN works for discrete as well as continuous data.

Application of GANs to different tasks such as classification and regression, image synthesis, image to image translation, and super-resolution is also experiencing a rapidly growing interest. Herein, we confine our literature review on most important contributions in the filed of medical imaging. Yang et al. [25] used WGANs for denoising Low-Dose Computed Tomography (LDCT) images. They also took advantage of pretrained VGG-19 network [26] for feature extraction and defining a perceptual loss function instead of MSE loss function. However, VGG-19 was trained on color images, and they duplicated the gray-scale channels to be able to feed CT images to VGG-19 network [25]. Simultaneously, another work on LCDT denoising was published which utilizes a Conveying Pathbased Convolutional Encoder-decoder (CPCE) network as the generator in a WGAN structure [27]. In another application, conditional GANs were used for reconstruction of magnetic resonance imaging (MRI) data recorded for a compressed sensing scenario [28]. The main idea in conditional GANs is conditioning both the generator and discriminator networks on some extra information [29]. In this work, frequency-domain information were used for conditioning the networks in order to have results that are similar in both time and frequency domains [28]. Nie et al. [30] used GANs for medical image synthesis. Their method was validated on reconstruction of MRI images from CT images and also generating 7T MRI from 3T MRI images. Recently, Mardani et al. [31] proposed a compressed sensing framework that uses GAN to remove the aliasing artifacts of undersampled MRI images.

In line-per-line US imaging, multilayer perceptron (MLP) was used for correction of phase aberrations [32] a long time ago. After many years, a deeper version of MLP was used for US beamforming [33], which trained several networks in frequency subbands to suppress off-axis scattering and remove clutter from channel data. This work used fully connected networks, which are prone to overfitting compared to CNNs. The reconstruction of B-Mode images from sub-sampled Radio-Frequency (RF) data using CNNs was investigated in [34]. Recently, CNNs were used for speckle reduction [35]. In ultrafast imaging, Gasse et al. [36] recovered high-quality plane-wave images from a limited number of transmitted angles using CNNs in a pilot study. Zhou et al. [37] improved the same idea and used multi-scale structure CNNs on different channels for recovery. To preserve the speckle information, wavelet postprocessing was added to the output of the network.As for the application of GANs in US imaging, in [38], a context-conditional GAN was used to acquire the quality of 128-channel B-Mode images from 32 channels. Speckle reduction was done using GAN in [39]. Recovery of high quality plane-wave images from a limited number of transmitted angles using GANs was performed in [40].

As for the purpose of multifocal imaging, Bottenus [41] proposed a method based on formulating a new frequency domain transmit encoding matrix that incorporates both delay and apodization to recover synthetic transmit aperture dataset. This method allows for synthetic transmit focusing at all points in the field of view. However, it is originally designed for phased array sequences in which the radial scan lines increase in separation in the axial direction. Consequently, this method was demonstrated on a walking aperture curvilinear sequence [42]. Using the regularized inverse of encoding matrix, the possibility of recovering synthetic transmit aperture dataset at each frequency for walking sequences was demonstrated in [43]. Recently, Ilovitsh et al. [44] proposed an approach which relies on superposition of axial mutifoci waveforms in a single transmission. Despite substantially

advancing the state-of-art, this method has two limitations. First, superposition can only be completed on a subset of probe crystals because of the piezoelectric maximal element response producing nonuniform quality in the axial direction. Second, it leads to an increase in thermal index due to transmissions of longer durations.

Herein, the central idea is generating several focal points by sending only one focused transmit beam. The nonlinear propagation pattern of the US beam is not stationary along the axial direction. Accordingly, in order to achieve a narrow beam everywhere, a mapping function between the single-focus US image and multi-focus version is estimated through different GANs. More specifically, the optimal focus depth of the transmit beam is found to be in the middle of imaging depth. The number of networks depends on the depth of imaging. In current study, we consider two networks to recover shallow and deep regions.

A preliminary version of this work was presented in ISBI 2019 [45]. The comparative analysis between the current work and previous work presented in [45] can be summarized as follows:

1) The generator network is significantly improved by adding residual connections, which enabled us to reduce the number of generator parameters by a factor of 7. This leads to better training and testing performance.
2) The training is now performed using BSGAN. The changes in the architecture and training technique substantially improved the results. To keep the manuscript concise, we did not include a comparison with the results of the ISBI version.
3) The training and test data are substantially extended. Four different shapes of cysts and highly scattering regions are simulated. For each shape, five different sizes are considered. And, for each size, 40 independent realizations of scatterer are simulated. In comparison, the ISBI version only contains one shape and size. Moreover, we wrote a script in Python to collect three consecutive images at different focal points without altering other imaging settings. This allowed us to collect consecutive images at a very high frame rate (i.e., more than 50 frames per second) to minimize the probability of misalignment between images.
4) The results of the Mean Opinion Score (MOS) test are added to strengthen validation and allow comparison on the perception of images.
5) The validation step is extended and now includes ex vivo experiments with Monte Carlo simulations.
6) As standard image processing metrics alone are not sufficient to assess US image quality, the results are extended to assess the performance of the proposed method in terms of the contrast to noise ratio (CNR) and full width at half maximum (FWHM) parameter, which are specialized ultrasound assessment indexes.
7) The results of the proposed approach are now compared to other multifocal methods which are not based on NNs.

Comprehensive experiments show that high quality multi-focus US images can be generated without sacrificing the frame-rate. As part of this manuscript, we will also make our code as well as all data available online. A copy of it is now available online for the reviewers in this link: Data-Codes.

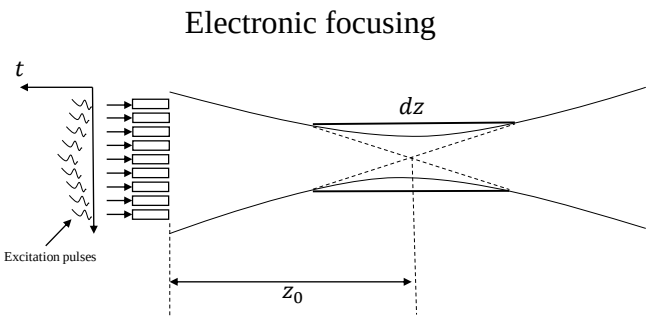


Fig. 1: Electronic focusing of the transmit beam by applying the time delays shown in left.

II. METHODOLOGY

A. Focusing

In electronic beamforming, in order to focus at a specific axial depth (z_0), a set of excitation pulses with proper time delays are applied to the crystals. This method, as shown in Fig. 1, is always used in classical line-per-line imaging. The highest amplitude of acoustic potentials is achieved at focus point. Therefore, the distance between two points where the field on axis is 3dB less than at the focal point is defined as depth of focus (dz) [6]. The lateral resolution is optimum in this region. In order to preserve the lateral resolution (having optimal multi-focus image), the maximum distance between transmitted focal points has to be equal to the depth of focus. We formulate our problem as finding a nonlinear mapping function which transforms the bowtie-shaped focused beam (with one focal point) to a thin cylindrical beam. However, this nonlinear function is nonstationary along the axial direction. In other words, this function varies with depth and cannot be estimated through only one network. Therefore, different networks should be trained that correspond to different depths. Consequently, the proposed method is based on partial estimation of nonlinear function for multiple depth intervals. This is a common solution for addressing nonstationary problems such as spectrum estimation. Therefore, we break the image into limited number of intervals along the axial direction such that we get closer to the stationary assumption in training convolutional neural networks and have a lower amount of variation, and subsequently train a BSGAN for each interval.

B. Proposed recovery method

Let x be a sample of input space, $\{x^{(i)} \in \mathbb{R}^{r\times c}\}_{i=1}^{m}$, which is an US image with single focus point (m denotes the number of samples. symbols r and c, respectively, denote the number of rows and columns), and y be the corresponding sample of output space, $\{y^{(i)} \in \mathbb{R}^{r\times c}\}_{i=1}^{m}$, which is a multi-focus US image. We formulate the problem as:

$$y = \mathcal{F}(x) \tag{1}$$

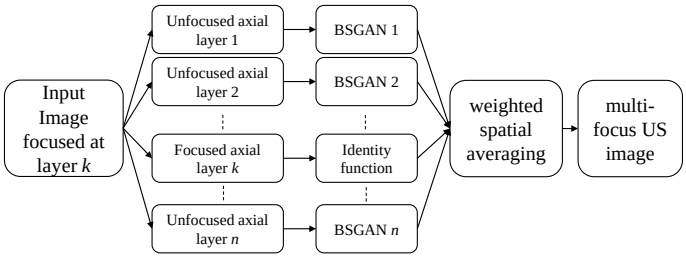


Fig. 2: The proposed recovery scheme. The input image is focused at layer k on which no transformation is applied. All other unfocused axial layers are transformed through distinct BSGANs - one for each layer.

where $\mathcal{F} : \mathbb{R}^{r\times c} \rightarrow \mathbb{R}^{r\times c}$ denotes the recovery function. Herein, a few main points have to be considered regarding the proposed recovery scheme. First, we assume that the recovery function $\mathcal{F}$ does really exist which means that it is possible to recover the multi-focus US image from a single-focus observation. Second, we assume that $\mathcal{F}$ can be estimated, with proper upper bound error, through GANs. In fact, the manifolds of input and output are in an unknown high dimensional space ($r \times c$), and the problem is ill-posed. However, it has been shown that CNNs are able to efficiently represent the input data in middle layers and estimate any nonlinear function with a desired upper bound error [46], [47]. These are reasonable assumptions that need to be commonly made for deep learning, and their mathematical proof is beyond the scope of this paper.

Our proposed recovery scheme is summarized in Fig. 2. First, the single-focus input B-Mode image is broken into a few axial layers. Then, for all layers where the transmitted beam is not focused, the mapping to the corresponding focused layer is achieved through a distinct BSGAN (i.e. a different network is trained for each axial layer). As the input image is focused at layer k, the output of this layer is the same as the input (i.e. an identity function is applied to this layer). Finally, all of the axial layers are merged together by minimal blending in small overlapping regions between layers in order to remove border effects as is the common practice, and multi-focus B-mode image is recovered.

C. Generative adversarial networks

Our aim is to estimate a nonlinear function that maps the input space to the target space. This aim can be fulfilled through CNNs. However, CNNs need an explicit differentiable objective function which scores the quality of results. Therefore, we need a distance measure $Dist$ between estimated output $\widehat{y}$ and desired output y. The problem can then be formulated as:

$$\widehat{\theta} = \arg\min_{\theta} \; Dist(\widehat{y}, y) \tag{2}$$

where θ is the parameters of the CNN. A long-running problem with CNNs is defining an appropriate distance measure. In other words, we still need to specify what we wish to minimize. As we will show in the Results Section, the commonly used MSE produces blurry results [48] because it averages across pixels. In the context of US imaging, this leads to incoherent averaging of the data which destroys the speckle pattern [3]. Fortunately, GANs give us the chance of reaching the desirable results only by specifying a high-level goal. What GANs learn is a loss function which classifies whether output is real or fake (the discriminator network) and a mapping function to minimize this loss (the generator network). Therefore, GANs consist of generator and discriminator networks, which compete with each other.

In classical form, GANs training is a min-max game between the generator and the discriminator [9]:

$$\begin{aligned}\min_{G}\max_{D} V(D,G) = \mathbb{E}_{y\sim p_{data}(y)}\left[log\ D(y)\right] \\ +\mathbb{E}_{x\sim p_X(x)}\left[log(1-D(G(x))\right]\end{aligned} \tag{3}$$

where y and x are the desired and input respectively with $\widehat{y} = G(x)$ the estimated/generated output. E denotes the expected value, and D and G are the discriminator and generator, respectively, and $V(D,G)$ denotes the objective function for GAN training. $y \sim p_{data}(y)$ means y is a sample of data generating distribution while $x \sim p_X(x)$ means x is a sample of input distribution.

D. Boundary seeking generative adversarial networks

It can be shown from Eq. 3 that the optimal discriminator $D^*(y)$ is given by [9]:

$$D^*(y) = \frac{p_{data}(y)}{p_{data}(y) + p_g(y)} \tag{4}$$

Hence, if the optimal discriminator with respect to generator is known, the global minimum of generator training will be $p_g = p_{data}$, wherein the desired distribution of output data is perfectly estimated by the generator, and the generator produces samples that are indistinguishable for discriminator. In practice, however, we are far from optimal case and the true data distribution, $p_{data}(y)$, could be achieved by weighting with the ratio of optimal discriminator as follows [24]:

$$p_{data}(y) = p_g(y)\frac{D^*(y)}{1-D^*(y)} \tag{5}$$

As the optimal discriminator is also unknown and hard to estimate, we always work with an approximation of $D^*(y)$. The intuition in training of GANs is that as we train the discriminator, it gets closer to $D^*(y)$, and consequently, the results improve. Eq. 5 means that the optimal generator is what makes the discriminator 0.5 everywhere, or a coin toss. In fact, $D(y) = 0.5$ is the decision boundary for a discriminator. So, BSGANs are a specific form of GANs in which generated data are close to the decision boundary of the discriminator [24].

The discriminator parameters ω are trained through the following optimization problem:

$$\begin{aligned}\widehat{\omega} &= \arg\min_{\omega} \; L_D(\widehat{y}, y) = \\ &= \arg\min_{\omega} \; L_{BCE}(D(y),1) + L_{BCE}(D(\widehat{y}),0)\end{aligned} \tag{6}$$

where $L_D(\widehat{y}, y)$ is the loss function for the discriminator. Herein, we used binary cross entropy (BCE) which is defined as follows:

$$L_{BCE}(D(y),l) = -\sum_i [l_i log(D(y_i)) + (1-l_i)log(1-D(y_i))] \tag{7}$$

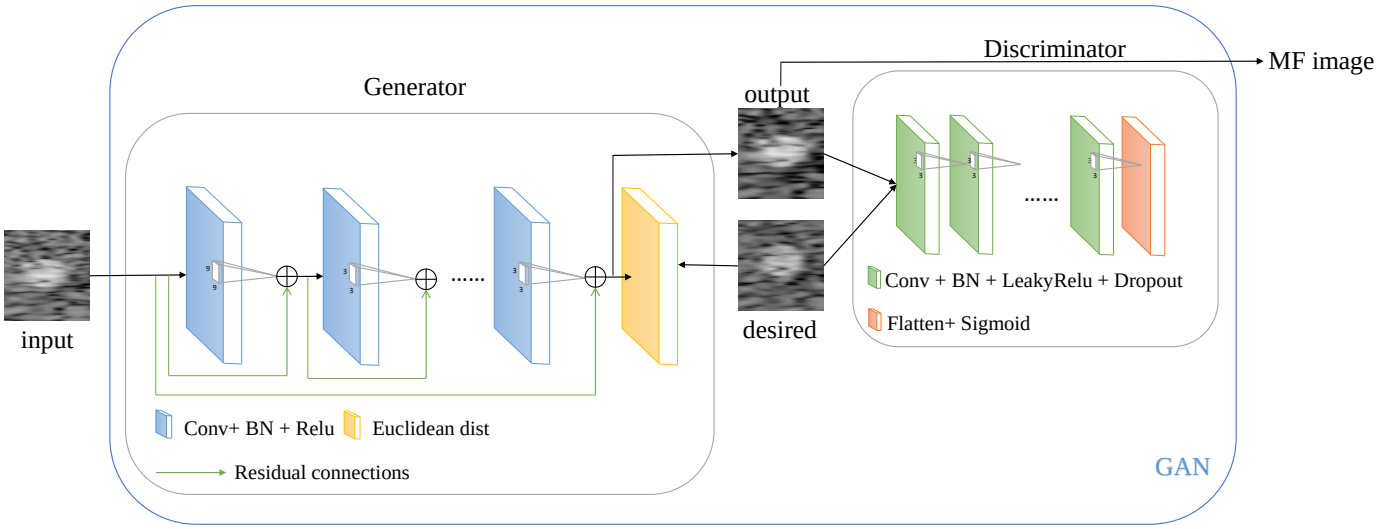


Fig. 3: The structure of the proposed BSGANs.

where l is the output label with values of $\{0, 1\}$. As can be seen from Eq. 6, the discriminator is trained to complete a two-class classification problem wherein the generated (i.e., not real) data is assigned to 0, and real data is assigned to 1.

The generator parameters θ are trained through the following optimization problem:

$$\begin{aligned}\widehat{\theta} &= \arg\min_{\theta} L_G(\widehat{y}, y) = \\ &= \arg\min_{\theta} \lambda_1 \parallel \widehat{y} - y \parallel^2 + \lambda_2 L_{BS}(\widehat{y})\end{aligned} \tag{8}$$

where $\parallel . \parallel$ is the second order norm, and λ_1 and λ_2 are regularization coefficients. The first term is the classical MSE loss function, and the second one $L_{BS}(\widehat{y})$ is the boundary seeking loss function which is defined as following:

$$L_{BS}(\widehat{y}) = \frac{1}{2m}\sum_{i=1}^{m}[log(D(\widehat{y}_i)) - log(1 - D(\widehat{y}_i))]^2 \tag{9}$$

In other words, we take advantage of both MSE and adversarial objective function to reach desirable results.

E. Proposed network

Our proposed network is shown in Fig. 3. The generator in Fig. 3 is a fully convolutional network with residual connections [49] consisting of 9 layers, where the first 8 layers contain 32 filters, and the last layer contains 1 filter. The first layer contains kernels of size 9×9, and other layers contain kernels of size 3×3. Each layer also contains batchnorm layer and a ReLU (Rectified Linear Unit) activation function except for the last layer which uses tanh activation function in order to map the output values between $[-1, 1]$. As shown in Fig. 3, we used both overall and local residual connections. The discriminator in Fig. 3 consists of 4 layers containing 32, 64, 128, and 256 convolution filters with the same kernel size of 3. The first 3 layers have stride of 2, and the fourth layer has a stride of one. Each layer also contains LeakyReLU, batchnorm, and dropout (rate = 0.25) layers. The last layer is flattening with sigmoid activation for getting the output label. The number of filters and layers was chosen to maintain a minimum number of parameters for preserving the generalization performance and a more stable training. Kernel sizes were chosen empirically. We did not encounter checkerboard artifacts because the input and output patches have the same size. In summary, total number of trainable parameters for generator and discriminator networks are 68,000 and 400,000, respectively.

III. EXPERIMENTS

A. Datasets

1) Simulated phantom: This dataset contains phantom simulations using the Field II program [50], [51]. The transducer configuration is described in Table I. The transducer configuration is described in Table I. The sampling frequency is reduced to 10 MHz after envelope extraction to reduce the size. The phantoms typically consist of 100,000 scatterers (more than 30 scatterers per wavelength to ensure fully developed speckles) and a collection of three point targets, three cyst regions, and three highly scattering regions in three different axial depths. Four different shapes of cysts and highly scattering regions are simulated. For each shape, five different sizes are considered. Finally, for each size 40 independent realizations of scatterers are simulated. For each realization (i.e., each phantom), three different images were simulated by changing the location of

TABLE I: Field II simulation setting

Parameter	Value	Unit
Array geometry	Linear	-
Number of elements	192	elements
Center frequency	3.5	MHz
Element width	0.44	mm
Element height	5	mm
Kerf	0.05	mm
Sampling frequency	100	MHz
Number of scan lines	50	lines
Speed of sound	1540	m/s

Fig. 4: Real phantom experiment setup.

the focal point. Therefore, we have 4*5*40*3=2,400 different simulated images in total. The size of images is fixed as 40 mm lateral * 60 mm axial. We use line-per-line imaging with delay and sum beamforming.

2) Real phantom: Multi-Purpose Multi-Tissue Ultrasound Phantom (CIRS model 040GSE, Norfolk, VA) was used as real phantom. US images were collected using an E-CUBE 12 Alpinion machine with L3-12H high density linear array and a centre frequency of 8.5 MHz. The sampling frequency of the RF data was 40MHz, and 384 RF lines were collected for each image. 20 independent images were collected at different locations of the phantom. At each location, three images with different focal points were collected, while the probe was held with a mechanical arm to prevent any probe movement during changing the transmit focus point. This ensured that images with different focal depths were collected at the same location. Our setup is shown in Fig. 4. Although more images can be collected from a phantom, only independent images are of significance in training process, and repeated similar images from the same location do not help the generalization ability of the network.

3) Ex vivo data: These images were collected from a fresh lamb liver. Imaging parameters are the same as phantom experiments. Instead of placing the liver in a gel phantom to minimize its motion during data collection, which may lead to some loss of blood and other tissue changes, we placed the liver on a plate and wrote a script in Python (which is the Alpinion interface for using the machine in research mode) to collect three consecutive images at different focal points without altering other imaging settings. This allowed us to collect consecutive images at a very high frame rate (i.e., more than 50 frame per second) to minimize the chance of misalignment between images. In addition, we attemped to hold the probe steady during data collection. These steps lead to a collection of images at different focusing depths with minimal relative motion between the probe and tissue. To have independent data points, we repeated the experiment five times by collecting images from different locations of the lamb liver.

B. Evaluation setting

For evaluation, we placed three real equispaced focal points in the axial direction of the US images, and blended the resulting three images by weighted spatial linear averaging as in commercial US machines. As such, the multi-focus image (desired) has 3 layers with 2 blended regions (Fig. 5 (b)). One of the images (Fig. 5 (a)) with the middle focal point is the input of our model. Therefore, the middle layer of the output (Fig. 5 (c-f)) is equal to the input, and two other layers are estimated from related layers of input through two BSGANs. Each layer was broken into 52×52 patches and fed to the network. During the test phase, we did not break the image, and each layer was fed to the generator to prevent the blocking artifact. For quantitative analysis, we tried to compare the results of the proposed method in terms of all image quality metrics. General metrics including Peak Signal to Noise Ratio (PSNR), Normalized Root Mean Square Error (NRMSE), and Structural Similarity (SSIM) index were calculated between ground truth and both of the output of proposed network and input. Additionally, MOS test was performed to show which form of network and which type of training is more successful in recovery of perceptually convincing images. Monte Carlo simulation was performed on *ex vivo* data to investigate the ability of the proposed method on recovering the sharpness of images in terms of Mean Gradient (MG) index. Afterward, the proposed method, using specialized ultrasound assessment indexes including Contrast to Noise Ratio (CNR) and Full Width at Half Maximum (FWHM), is compared with other approaches which are not based on NNs. The next subsection describes details of the MOS comparison.

C. Mean Opinion Score (MOS) testing

As common indices for image quality assessment have a limited potential to indicate how much an image is perceptually convincing, we performed an MOS test to improve the validation step. More specifically, 20 graduate students who work in the field of US imaging, as raters, were asked to assign a score from 1 (bad quality) to 5 (excellent quality) to images. 6 versions of simulated phantom image (Fig. 5 (a-f)) were rated. Images were presented in a randomized fashion to raters. Raters very consistently rated ground truth image as excellent quality and the original input image (with only single focal point) as bad quality. Moreover, we put two identical images in questionnaire to make sure that answers are reliable. The summary of all results is reported in Table II.

D. Network training

The entire database was broken into three sets of training, validation, and test groups with sizes of 70, 15, and 15 percent of the total size of images, respectively. We first normalized the intensity input US images to [-1,1]. As it is common in training GANS [9], in each iteration, the discriminator is trained 3 times (N_D), and the generator one time. In all experiments, the Adam algorithm with learning rate ($\alpha = 10^{-4}$) was used for optimization [52]. The training procedure of the proposed BSGAN is shown in Algorithm 1. The code is implemented

Algorithm 1 Minibatch stochastic gradient descent training of BSGANs. The number of steps to apply to the discriminator $N_D = 3$. All experiments in the paper used Adam parameters, $\alpha = 10^{-4},\ \beta_1 = 0.9,\ \beta_2 = 0.99$.

Require: set $\lambda_1 = 0.4$, $\lambda_2 = 90$.
Require: set the number of total epochs, $N_{epoch} = 100$, the batch size $m = 64$.
calculate the number of iteration in each epoch
$N_{iter} \leftarrow total\ number\ of\ training\ samples/m$
Require: ω_0, initial discriminator parameters. θ_0, initial generator parameters.
for N_{epoch} **do**
  for N_{iter} **do**
    for N_D **do**
      sample a batch of input patches $\{x^i\}_{i=1}^m$
      sample a batch of ground truth patches $\{y^i\}_{i=1}^m$
      update the discriminator by descending its stochastic gradient:
      $\nabla_\omega[-\frac{1}{m}\sum_{i=1}^m log(D(y_i)) + log(1 - D(G(x_i)))]$
    end for
    sample a batch of input patches $\{x^i\}_{i=1}^m$
    sample a batch of ground truth patches $\{y^i\}_{i=1}^m$
    update the generator by descending its stochastic gradient:
    $\nabla_\theta[\frac{\lambda_1}{m}\sum_{i=1}^m (y_i - G(x_i))^2 + \frac{\lambda_2}{2m}\sum_{i=1}^m [log(D(G(x_i))) - log(1 - D(G(x_i)))]^2]$
  end for
  calculate average SSIM index over the validation set.
end for
select the model with highest SSIM index for test.

using Keras library using TensorFlow back-end, and training was done with an Nvidia Titan Xp GPU.

The solution to training a BSGAN network (which is a game between two players) is a Nash equilibrium. In fact, by having the optimal discriminator, the global minimum of generator's loss function is achieved if and only if $p_g = p_{data}$, which means that the discriminator gives the same probability of 0.5 to both generated and real data. Although the two players may suddenly reach an equilibrium, the training process oscillates between two modes and players repeatedly undo each other. In fact, as we never reach the perfect case (in which $p_g = p_{data}$), after finishing training process for a specific number of epochs, the model which has the best structural similarity to desired on validation dataset is chosen as final model. The final model of training is saved and applied to the test set.

For real data (i.e., real phantom, *ex vivo* experiments), we used transfer learning to fine tune the networks trained on simulated data. Transfer learning was done in the same adversarial manner as before and used for fine tuning the weights of whole of the layers in generator and discriminator. More specifically, weights of the best network on simulated phantom data was used as initial point of training on new data. As before, model selection was done based on SSIM evaluation on validation data. Finally, selected generator was used for test part.

IV. RESULTS

A. Experimental methods

The first network used in comparison is a well-known structure named Super-Resolution CNN (SRCNN), a relatively shallow network with 3 layers without any residual connections, about which details can be found in [53]. The second, entitled Residually connected Fully CNN (RFCNN), is the generator in our proposed structure shown in Fig. 3, which is deeper and also has residual connections compared to SRCNN. Consequently, proposed RFCNN is used in a basic non-saturating GAN structure [9]. Finally, the basic GAN is extended to boundary seeking version. It has to be mentioned that non-GAN networks are only trained with MSE loss function. It is worth mentioning that in order to better illustrate the results, the difference maps of Fig. 6, 7, and 9 are provided in the supplementary materials.

B. Comparison on simulated phantom

In the first experiment, the performance of different networks is evaluated on the simulated phantom data. As can be seen in Fig. 5, both SRCNN and RFCNN do not perform very well and lead to over smooth images. The main reason for the loss of fine details is that the network is trained with only MSE as the loss function. In the GAN structure, however, the image quality is enforced indirectly by the discriminator in adversarial training, as the generator network tries to make images that look like real images. Between the basic GAN structure and the boundary seeking version, the latter works better because the training process of BSGAN is more stable and the discriminator is nearer to the optimal case. As can be seen in Fig. 5, the GAN result has some artifacts in the middle part of the cyst region. Furthermore, the GAN result has more contrast, but even more than the ground truth (b). So, as it has been shown in [24] for general images, GAN results are more artificial while results of BSGAN are more natural.

In order to provide better comparison among different methods, quantitative results are presented in Table II in which the input is a single focus image, and the desired output is a multi-focus image. As mentioned, common indices do not illustrate how much an image is perceptually convincing as these metrics are not developed for US images. Therefore, SRCNN and RFCNN have better values on some of those general indices because their results are very smooth. However, their poor quality is uncovered by the expert raters, and GAN-based networks get much better scores. Comparing basic GAN and BSGAN, the second one has better results with lower standard deviation.

The second question that should be answered is about the selection of the input. In fact, we want to know whether the proposed method depends on the place of focus point in the input or not. To this end, we ran the algorithm for different scenarios. Results showed that the best selection is when the input image is focused at the middle position of axial direction, as expected. More specifically, we found that when the single focus point is in the middle part of the image, the similarity with multi focus image is the highest value. So, this input is

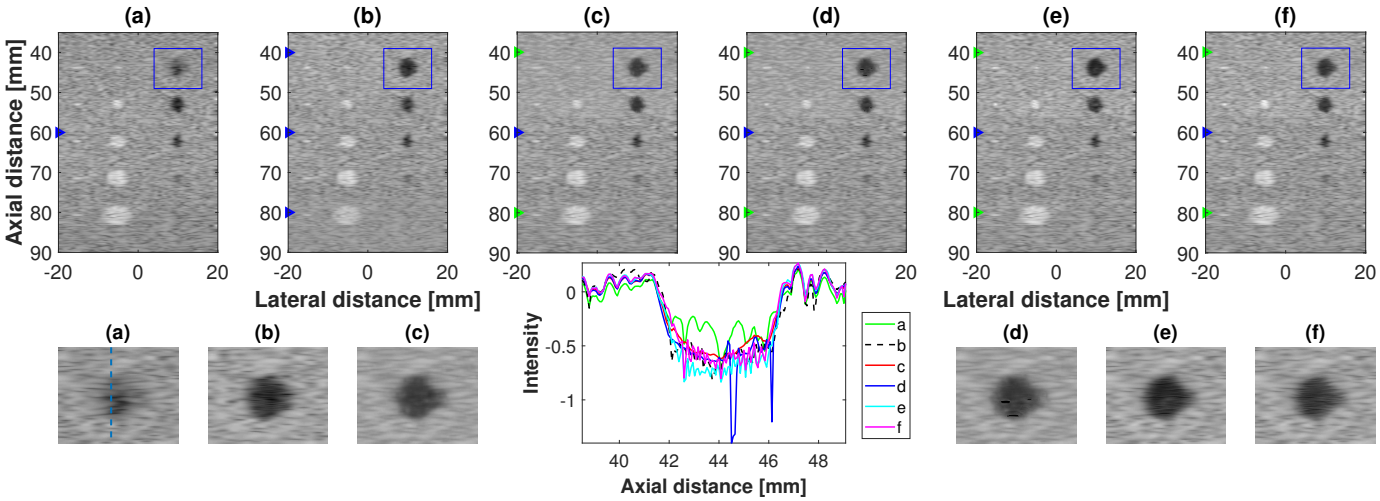


Fig. 5: Results of the different methods on the simulated phantom data. Blue triangles indicate real transmit focal points, and green triangles indicate focal points added by the network. (a) Input image with a single focal point. (b) Desired image with 3 focal points. (c) Output of the SRCNN (d) RFCNN (e) GAN (f) the proposed BSGAN. The second row shows a zoomed in view of the blue rectangle in the first row, and edge spread function of different methods across the vertical line shown in zoomed view of (a) is in the middle.

the most informative one and the mapping function from the input space to the output space is more straightforward.

The other important point regarding training the network on simulation data is training history. As for the MSE loss, the training and validation curves are provided in supplementary materials. Please note that the adversarial loss function does not reveal useful information in training GANs, and, as such, is not presented in this paper. To check whether the training has converged or not, generating a few samples and looking at them during the training phase is instead commonly performed [16], [54]. In this way, the evolution of what network learns, as a mapping function between input and output domains, is illustrated as the output of the network for a validation sample after different epochs in supplementary materials. Finally, in addition to the results listed in Table II, the training history of the non-saturating GAN is also provided in the supplementary materials to better illustrate the superiority of using BSGAN.

C. Real phantom results

The proposed method was also validated on real phantom data. As for real phantom data, whenever the focus point was set on first or second axial layer, image of the last layer had a very low quality. Consequently, it is understood as noise by the network, and discriminator gives the probability of 0.5 to it which means that the discriminator is uncertain whether it is real or generated data. For real phantom data, therefore, the image focused on third axial layer was used as input and two other layers were estimated using BSGANs although this was not the best scenario as discussed in last subsection. Fig. 6 shows the result of different methods on test data, which depicts the sharp borders of cysts as well as the hyperechoic regions are preserved only in the output of the proposed method as the desired image. It can be easily understood that the proposed method outperforms other approaches noticeably.

As for fine tuning using the real phantom experiment, the number of images is limited compared to simulation data. To reduce the risk of overfitting, two common approaches of training the weights of a specific layer or training for few epochs are commonly used [55], [56]. We chose the latter. In this way, we multiplied the learning rate with 0.1 and limited the number of epochs to 10. This ensures that weights only change slightly. Finally, in order to understand how much the fine-tuning impacts the final result, the output of the trained network on simulation data without fine tuning is presented in supplementary materials.

D. Ex vivo results

In real tissues, there are two main limitations preventing the multi-focus desired image to have a noticeable difference

TABLE II: The results of PSNR, NRMSE, SSIM, and MOS between input-desired and output-desired pairs. The best values (highest mean and lowest std) are in bold font.

data	input				SRCNN				RFCNN				GAN				BSGAN			
index	PSNR	NRMSE	SSIM	MOS	PSNR	NRMSE	SSIM	MOS	PSNR	NRMSE	SSIM	MOS	PSNR	NRMSE	SSIM	MOS	PSNR	NRMSE	SSIM	MOS
mean	23.27	0.034	0.622	1	26.46	0.023	0.782	3.15	26.78	0.023	0.794	3.15	24.69	0.029	0.773	3.92	25.32	0.027	0.769	**4.07**
std	1	0.004	0.02	0	0.95	0.002	0.018	0.688	0.932	0.002	0.016	0.89	0.795	0.002	0.01	0.64	0.919	0.003	0.017	**0.49**
min	20.77	0.026	0.574	1	22.98	0.019	0.729	2	23.44	0.018	0.74	2	22.71	0.022	0.725	3	22.9	0.021	0.723	**3**
max	25.62	0.045	0.684	1	28.29	0.035	0.824	4	28.56	0.033	0.826	5	26.9	0.036	0.803	5	27.38	0.035	0.797	**5**
median	23.16	0.034	0.621	1	26.57	0.023	0.784	3	26.92	0.022	0.796	3	24.77	0.028	0.778	4	25.46	0.026	0.775	**4**

compared to single-focus input. First, there is no specific cyst or hyperechoic region in the tissue which makes the comparison more difficult to clearly visualize the improvement in the image quality. Second, the depth of imaging is limited which means the amount of degradation in image quality, because of beam divergence, may be difficult to notice.

Based on aforementioned reasons, Monte Carlo simulation is used to better investigate the performance of the proposed method on *ex vivo* data. More specifically, a PSF is convolved with the image to simulate large imaging PSF away from the focal point. As mentioned in Section II-A, we assume that changes within each axial layer is negligible and for each axial layer one GAN is trained. The standard deviation (STD) of the Gaussian PSF is the parameter which specifies the level of blurriness and is composed of two deterministic and random parts as follows:

$$c = c_{det} + c_{rand} \tag{10}$$

where c indicates STD of the Gaussian PSF. The deterministic part of STD (c_{det}) specifies a minimum level of blurriness which is set to 1. A positive random number taken from $\mathcal{N}(0,\, \sigma^2)$ is used as the random component of STD (c_{rand}). The random part is added to the deterministic component to specify the level of blurriness in each run. Consequently, Monte Carlo simulation is done for 10 different values of σ. For each value of σ, 100 runs are performed. Fig. 7 illustrates the results for *ex vivo* data. Fig. 7 is shown after convolving with a Gaussian PSF having a STD of 8. The blurring is not applied on the correct focused layer because there is no modification on that. As can be seen in Fig. 7, the proposed method successfully recovered the multi-focus image, very similar to the ground truth, while other methods failed to recover fine details from the blurry input. Fig. 8 summarizes the observed changes in image quality as the STD of the simulated Gaussian PSF is increased. More specifically, Fig. 8 illustrates the box plot of image quality indices obtained from a Monte Carlo simulation comprising of 100 runs for each value of σ. We want to make sure that the proposed method preserves its performance over a wide range of simulated blurriness. As shown in the first row of Fig. 8, the SSIM index between the blurred input image and desired multi-focus image rises as the amount of blurriness (c) increases. Therefore, other indices, such as the Mean Gradient (MG) index, which reflects the sharpness and texture changes of the image should be used. As observed in the second row of Fig. 8, the output of the proposed method is substantially sharper than the input and much closer to the desired image for all levels of blurriness that we tested.

E. Comparison with other methods

In this subsection, the result of the proposed method is compared with other multifocal methods which are not based on NNs. As reviewed in section I, two multi-focal methods were proposed before us. Comparison with the method proposed by Ilovitsh et al. [44] was not possible for us because of two main reasons. More specifically, their method is based on the summation of electrical stimulation corresponding to different focused transmissions. So, one multi-focal beam which has a longer duration of time is transmitted instead of transmitting several single focus beams. However, the summation causes not only cross-talk, but also it can only be completed in a subset of probe crystals because of the piezoelectric maximal element response which causes nonuniform quality in the axial direction. This problem makes the comparison impossible. Moreover, we cannot implement the method on our research machine. However, the method proposed by Bottenus et al. named Retrospective Encoding For Conventional Ultrasound Sequences (REFoCUS) could be applied using a conjugate transpose (REFoCUS adjoint) [41], or a regularized inverse (REFoCUS inverse) [43]], of the transmit encoding matrix at each frequency. Fig. 9 shows the results of our comparison based on a simulated phantom data with imaging details

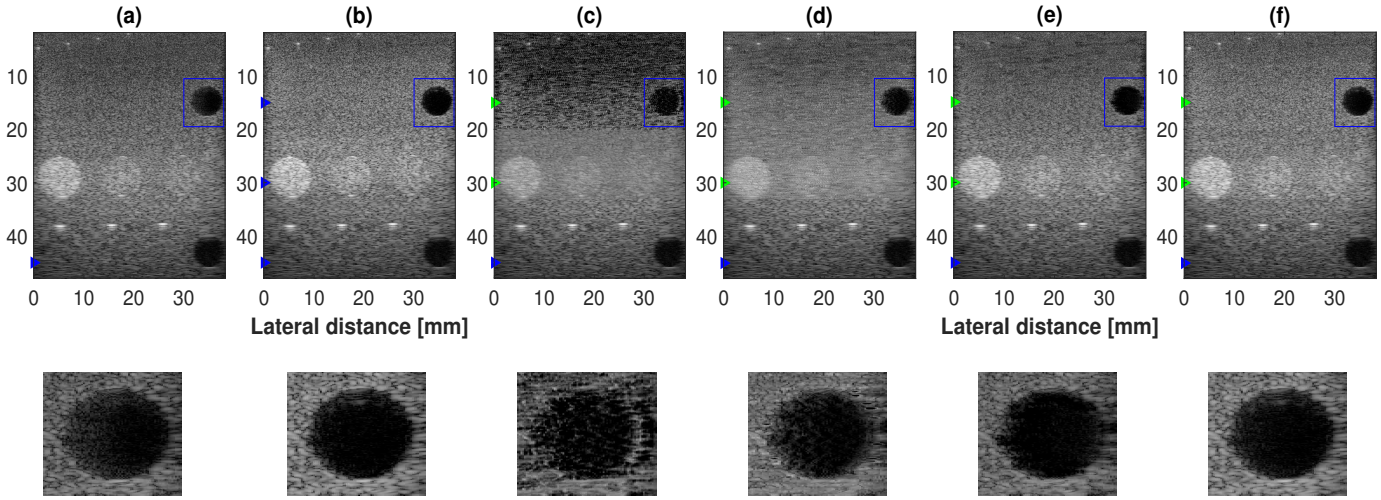


Fig. 6: Results of the different methods on real phantom data. Blue triangles indicate real transmit focal points, and green triangles indicate focal points added by the network. (a) Input image with a single focal point. (b) Desired image with 3 focal points. (c) Output of the SRCNN (d) RFCNN (e) GAN (f) the proposed BSGAN. The second row shows a zoomed in view of the blue rectangle in the first row.

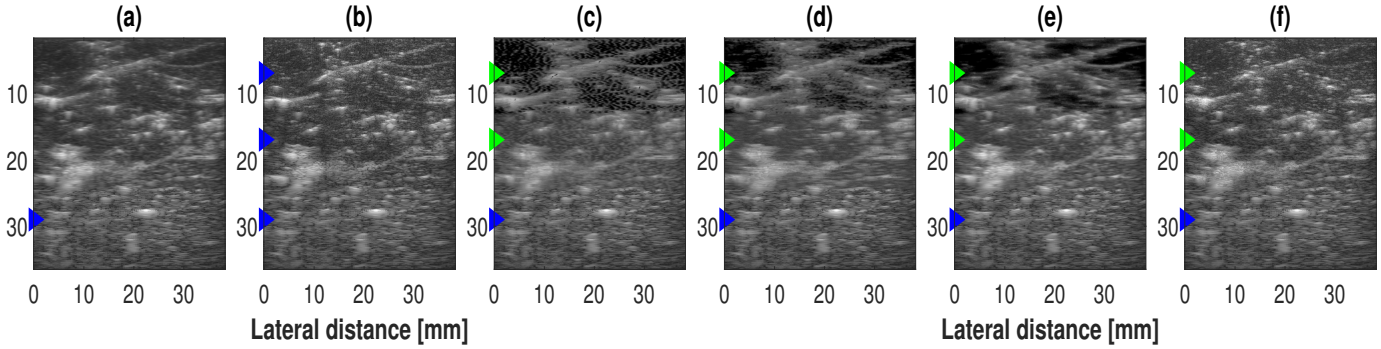


Fig. 7: Results of the different methods on *ex vivo* data. Blue triangles indicate real transmit focal points, and green triangles indicate focal points added by the network.(a) Input image with a single focal point. (b) Desired image with 3 focal points. (c) Output of the SRCNN (d) RFCNN (e) GAN (f) the proposed BSGAN.

exactly the same as for the real experiments. Herein, we used fine-tuned networks on real phantom data. As observed in Fig. 9, the result of proposed method is much more similar to the ground truth while other methods exhibit poor performance at the first row of cysts. In deeper regions, our method gives result similar to the ground truth while other methods are even better than the ground truth. In order to better illustrate the superiority of proposed method, the difference map of Fig. 9 is included in the supplementary material. Our method which is based on NNs does not require either channel data or any sort of matrix inversion, which is worth noting for practical applications since improved inversion matrix techniques are computationally expensive and time consuming while NNs in test case are on-line.

F. Ultrasound image quality metrics

The importance of using the adversarial loss function (GAN structure) as well as the boundary seeking method of training compared to other cases is demonstrated in subsection IV-A. The last subsection of the results is dedicated to assess proposed methods in terms of specialized ultrasound assessment indexes [57]. To this goal, the CNR [57] and the FWHM indices are calculated. As our method is proposed to preserve the lateral resolution, we only calculate the FWHM in the lateral direction. More specifically, for the simulation experiment, the point spread function of the imaging system is simulated by placing point targets on different focus points along the axial direction. Consequently, the FWHM is calculated for the input single-focus image, ground truth, and the results of the REFoCUS method. For real experiments, the calculation is performed using the point targets in the real phantom as shown in Fig. 4. It has to be mentioned that the comparison with the REFoCUS method only was possible for the simulation data. Moreover, we did not have the cyst region on the second axial layer of the real phantom, so, the CNR is only reported for the first layer. As it can be seen from Table III, the REFoCUS method provides better resolution in terms of FWHM while the proposed method has a better performance in terms of contrast. However, as it is shown in supplementary materials, the lower FWHM (narrower main lobe) value for the REFoCUS method comes at the expense of worse side lobes. Table III also confirms the improvement of indexes for the real phantom experiment.

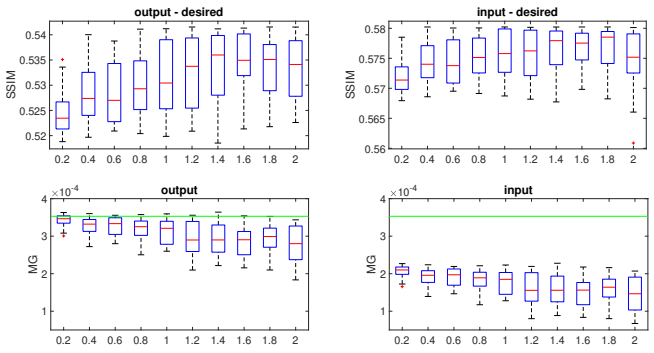


Fig. 8: Results of Monte Carlo simulation. First row contains the box plot of SSIM versus σ, and second row illustrates the box plot of MG versus σ. Green line shows the MG of desired image.

V. DISCUSSION

The results have illustrated that the proposed method based on BSGAN noticeably outperforms our implementations of SRCNN, RFCNN and GAN learning structures. Having residual connections in the fully convolutional generator network provides better performance because it is easier to learn the difference between the input and output [49]. The necessity of using adversarial objective function in training besides basic MSE loss function for having a sharper image, which is more perceptually convincing, is rather significant. Moreover, using the boundary seeking method for training provides a policy gradient for training the generator, and generates samples near the decision boundary. This ultimately leads to improved stability in training.

The proposed method was also tested on real applications. Transfer learning was successfully performed from the simulated space to real space. In order to provide desired training data in real experiments, two different settings were used. First, we used a mechanical arm to prevent any probe movement during changing the transmit focus point. Second, we wrote a data collection script in Python to alleviate the problem

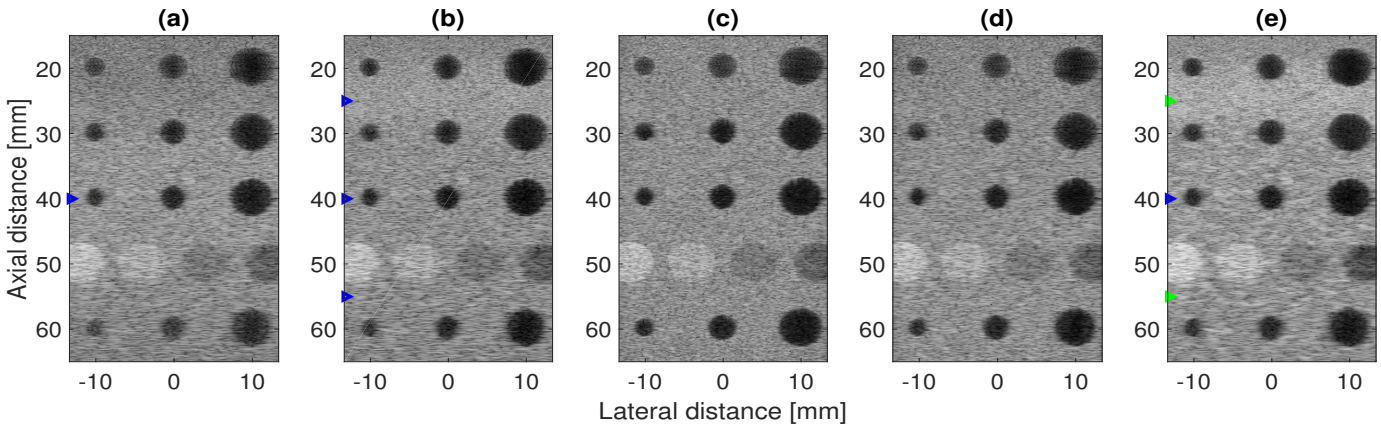


Fig. 9: Comparison of results with other methods. Blue triangles indicate real transmit focal points, and green triangles indicate focal points added by the network.(a) Input image with a single focal point. (b) Desired image with 3 focal points. (c) Output of REFoCUS inverse (d) REFoCUS adjoint (e) the proposed BSGAN.

of unavoidable movements during data collection. The network also works well in real experiments. Comprehensive qualitative and quantitative results with Monte Carlo simulations also verified the higher quality of recovered multi-focus images as compared to the single focus inputs.

The results of comparison with other non-NN based method (REFoCUS) show that the proposed method achieves similar results while it is faster and does not require matrix inversions. Moreover, a comparison on specialized ultrasound assessment indexes shows that the proposed method is able to simultaneously improve both resolution and contrast. Moreover, it is possible to combine the ideas in the proposed method and REFoCUS to further improve image quality, which is an interesting avenue for future work. For example, the output of REFoCUS method can be considered as the ground truth in the training step.

Currently, most of the commercial scanners are still using line-per-line acquisition, and plane-wave imaging is prohibitively expensive for affordable point-of-care ultrasound scanners. Therefore, most future scanner designs are likely to rely on line-per-line acquisition technique. For example, several next generation pocket-size ultrasound machines such as Extend R2 (GE Vscan), Sonon (Healcerion, USA) and Clarius all cost less than $5K. In comparison, only the data acquisition board for plane-wave imaging usually costs approximately $10K. In addition, plane-wave imaging also requires expensive computations using high-end GPUs. The proposed method in this manuscript requires a GPU for training and can be easily implemented on a CPU in the test phase making it a cost-effective choice for the next generation pocket-size ultrasound machines.

As the proposed method works on B-Mode images, its application to Doppler imaging and motion estimation algorithms, which are based on RF data, is limited. Moreover, an important issue in using machine learning methods for different medical image processing tasks, such as image synthesis, denoising and image reconstruction, is the reliability of generated results for the sake of diagnosis and surgical planning and guidance. In other words, are these results misleading or helpful for clinicians? In future, we plan to extend the proposed method to work on pre-beamformed RF data and test the performance of the proposed method in in-vivo applications and perform MOS tests with radiologists. In addition, we will explore training conditional GAN structures to be able to recover US images with a specific amount of reliability.

VI. CONCLUSIONS

A reduction in the frame-rate and motion blurs are the main challenges associated with multi-focus line-per-line imaging technique. Herein, the proposed alternative works as a non-linear mapping function from the input space (US image with one transmitted focused beam) to the optimum multi-focus output space. As shown above, GANs with boundary seeking method of training have been adapted to achieve the quality of multifocus US images without any loss in frame-rate or appreciable drop in quality due to probe movement. The experiments confirm that the proposed approach provides perceptually convincing images with a higher resolution and contrast, while it is computationally efficient and does not require channel data. The proposed approach can potentially be used in applications that require both high frame rate and lateral resolution.

TABLE III: The results of CNR and FWHM indexes for simulation and real phantom experiments.

method		input		desired		BSGAN		ReFOCUS (adjoint)		ReFOCUS (inverse)	
index		FWHM	CNR	FWHM	CNR	FWHM	CNR	FWHM	CNR	FWHM	CNR
simulation	layer 1	1.3	7.2	1.01	8.32	1.09	8.02	1.15	7.2	1.04	7.56
	layer 3	2.13	6.29	1.88	7.3	1.95	6.95	1.37	6.9	.94	7.7
real phantom	layer 1	1.52	9.6	1.37	11.7	1.44	11.1	-	-	-	-
	layer 2	.88	-	.74	-	.83	-	-	-	-	-

ACKNOWLEDGEMENT

The authors would like to thank Dr. N. Bottenus for providing us with the results of the REFoCUS method and the associated Field II code and data. We acknowledge the support of the Natural Sciences and Engineering Research Council of Canada (NSERC) RGPIN-2020-04612 and RGPIN-2017-06629. We would like to thank current and former graduate students of IMPACT lab for their help in the MOS test. Finally, we would like to thank NVIDIA for donation of the Titan Xp GPU.

REFERENCES

[1] I. Trots, A. Nowicki, and M. Lewandowski, "Synthetic transmit aperture in ultrasound imaging," *Archives of Acoustics*, vol. 34, no. 4, pp. 685–695, 2009.

[2] J. Liu, Q. He, and J. Luo, "Compressed sensing based synthetic transmit aperture imaging: Validation in a convex array configuration," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 65, no. 3, pp. 300–315, March 2018.

[3] G. Montaldo, M. Tanter, J. Bercoff, N. Benech, and M. Fink, "Coherent plane-wave compounding for very high frame rate ultrasonography and transient elastography," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 56, no. 3, pp. 489–506, March 2009.

[4] B. Lokesh and Arun K. Thittai, "Diverging beam transmit through limited aperture: A method to reduce ultrasound system complexity and yet obtain better image quality at higher frame rates," *Ultrasonics*, vol. 91, pp. 150 – 160, 2019.

[5] J. A. Jensen, "Linear description of ultrasound imaging systems," *Notes for the International Summer School on Advanced Ultrasound Imaging, Technical University of Denmark July*, vol. 5, pp. 54, 1999.

[6] G. S. Kino, *Acoustic waves: devices, imaging, and analog signal processingPrentice-Hall Signal Processing Series*, Englewood Cliffs, Prentice-Hall, 1987.

[7] M. K. Feldman, S. Katyal, and M. S. Blackwood, "Us artifacts," *RadioGraphics*, vol. 29, no. 4, pp. 1179–1189, 2009, PMID: 19605664.

[8] O. Senouf, S. Vedula, G. Zurakhov, A. Bronstein, M. Zibulevsky, O. Michailovich, D. Adam, and D. Blondheim, "High frame-rate cardiac ultrasound imaging with deep learning," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2018, pp. 126–134.

[9] I. Goodfellow, J. Pouget-Abadie, M. Mirzai, B. Xu, D. Warde-Farley, S. Ozairl, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Advances in neural information processing systems*, pp. 2672–2680. 2014.

[10] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, pp. 436, 2015.

[11] Y. You, C. Lu, W. Wang, and C. Tang, "Relative cnn-rnn: Learning relative atmospheric visibility from images," *IEEE Transactions on Image Processing*, vol. 28, no. 1, pp. 45–55, Jan 2019.

[12] G. Cheng, J. Han, P. Zhou, and D. Xu, "Learning rotation-invariant and fisher discriminative convolutional neural networks for object detection," *IEEE Transactions on Image Processing*, vol. 28, no. 1, pp. 265–278, Jan 2019.

[13] J. Liang, Q. Hu, C. Dang, and W. Zuo, "Weighted graph embedding-based metric learning for kinship verification," *IEEE Transactions on Image Processing*, vol. 28, no. 3, pp. 1149–1162, March 2019.

[14] A. Creswell, T. White, V. Dumoulin, K. Arulkumaran, B. Sengupta, and A. A. Bharath, "Generative adversarial networks: An overview," *IEEE Signal Processing Magazine*, vol. 35, no. 1, pp. 53–65, Jan 2018.

[15] M. Arjovsky and L. Bottou, "Towards principled methods for training generative adversarial networks," *arXiv preprint arXiv:1701.04862*, 2017.

[16] D. Berthelot, T. Schumm, and L. Metz, "Began: boundary equilibrium generative adversarial networks," *arXiv preprint arXiv:1703.10717*, 2017.

[17] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen, "Improved techniques for training gans," in *Advances in Neural Information Processing Systems*, 2016, pp. 2234–2242.

[18] A. Radford, L.Metz, and S.Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," *arXiv preprint arXiv:1511.06434*, 2015.

[19] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein gan," *arXiv preprint arXiv:1701.07875*, 2017.

[20] B. K. Sriperumbudur, K. Fukumizu, A. Gretton, B. Schölkopf, and G. RG Lanckriet, "On integral probability metrics,\phi-divergences and binary classification," *arXiv preprint arXiv:0901.2698*, 2009.

[21] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. C. Courville, "Improved training of wasserstein gans," in *Advances in Neural Information Processing Systems 30*, pp. 5767–5777. Curran Associates, Inc., 2017.

[22] Jiqing Wu, Zhiwu Huang, Janine Thoma, Dinesh Acharya, and Luc Van Gool, "Wasserstein divergence for gans," in *The European Conference on Computer Vision (ECCV)*, September 2018.

[23] K. Roth, A. Lucchi, S. Nowozin, and T. Hofmann, "Stabilizing training of generative adversarial networks through regularization," in *Advances in Neural Information Processing Systems 30*, pp. 2018–2028. Curran Associates, Inc., 2017.

[24] R. D. Hjelm, A. P. Jacob, T. Che, A. Trischler, K. Cho, and Y. Bengio, "Boundary-seeking generative adversarial networks," *arXiv preprint arXiv:1702.08431*, 2017.

[25] Q. Yang, P. Yan, Y. Zhang, H. Yu, Y. Shi, X. Mou, M. K. Kalra, Y. Zhang, L. Sun, and G. Wang, "Low-dose ct image denoising using a generative adversarial network with wasserstein distance and perceptual loss," *IEEE Transactions on Medical Imaging*, vol. 37, no. 6, pp. 1348–1357, June 2018.

[26] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *CoRR*, vol. abs/1409.1556, 2014.

[27] H. Shan, Y. Zhang, Q. Yang, U. Kruger, M. K. Kalra, L. Sun, W. Cong, and G. Wang, "3-d convolutional encoder-decoder network for low-dose ct via transfer learning from a 2-d trained network," *IEEE Transactions on Medical Imaging*, vol. 37, no. 6, pp. 1522–1534, June 2018.

[28] G. Yang, S. Yu, H. Dong, G. Slabaugh, P. L. Dragotti, X. Ye, F. Liu, S. Arridge, J. Keegan, Y. Guo, and D. Firmin, "Dagan: Deep de-aliasing generative adversarial networks for fast compressed sensing mri reconstruction," *IEEE Transactions on Medical Imaging*, vol. 37, no. 6, pp. 1310–1321, June 2018.

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

[30] D. Nie, R. Trullo, J. Lian, L. Wang, C. Petitjean, S. Ruan, Q. Wang, and D. Shen, "Medical image synthesis with deep convolutional adversarial networks," *IEEE Transactions on Biomedical Engineering*, vol. 65, no. 12, pp. 2720–2730, Dec 2018.

[31] M. Mardani, E. Gong, J. Y. Cheng, S. S. Vasanawala, G. Zaharchuk, L. Xing, and J. M. Pauly, "Deep generative adversarial neural networks for compressive sensing mri," *IEEE Transactions on Medical Imaging*, vol. 38, no. 1, pp. 167–179, Jan 2019.

[32] M. Nikoonahad and D. C. Liv, "Medical ultrasound imaging using neural networks," *Electronics Letters*, vol. 26, no. 8, pp. 545–546, April 1990.

[33] A. C. Luchies and B. C. Byram, "Deep neural networks for ultrasound beamforming," *IEEE Transactions on Medical Imaging*, vol. 37, no. 9, pp. 2010–2021, Sep. 2018.

[34] Y. H. Yoon, S. Khan, J. Huh, and J. C. Ye, "Efficient b-mode ultrasound image reconstruction from sub-sampled rf data using deep learning," *IEEE Transactions on Medical Imaging*, vol. 38, no. 2, pp. 325–336, Feb 2019.

[35] D. Hyun, L. L. Brickson, K. T. Looby, and J. J. Dahl, "Beamforming and speckle reduction using neural networks," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 66, no. 5, pp. 898–910, May 2019.

[36] M. Gasse, F. Millioz, E. Roux, D. Garcia, H. Liebgott, and D. Friboulet, "High-quality plane wave compounding using convolutional neural networks," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 64, no. 10, pp. 1637–1639, Oct 2017.

[37] Z. Zhou, Y. Wang, J. Yu, Y. Guo, W. Guo, and Y. Qi, "High spatial–temporal resolution reconstruction of plane-wave ultrasound images with a multichannel multiscale convolutional neural network," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 65, no. 11, pp. 1983–1996, Nov 2018.

[38] C. Huang, O. T. Chen, G. Wu, C. Chang, and C. Hu, "Ultrasound imaging improved by the context encoder reconstruction generative adversarial network," in *2018 IEEE International Ultrasonics Symposium (IUS)*, Oct 2018, pp. 1–4.

[39] F. Dietrichson, E. Smistad, A. Ostvik, and L. Lovstakken, "Ultrasound speckle reduction using generative adversial networks," in *2018 IEEE International Ultrasonics Symposium (IUS)*, Oct 2018, pp. 1–4.

[40] X. Zhang, J. Li, Q. He, H. Zhang, and J. Luo, "High-quality reconstruction of plane-wave imaging using generative adversarial network," in

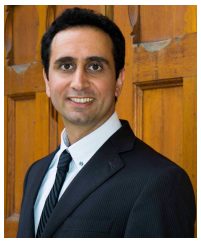
2018 IEEE International Ultrasonics Symposium (IUS), Oct 2018, pp. 1–4.
[41] N. Bottenus, "Recovery of the complete data set from focused transmit beams," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 65, no. 1, pp. 30–38, Jan 2018.
[42] N. Bottenus, "Comparison of virtual source synthetic aperture beamforming with an element-based model," *The Journal of the Acoustical Society of America*, vol. 143, no. 5, pp. 2801–2812, 2018.
[43] R. Ali, J. J. Dahl, and N. Bottenus, "Regularized inversion method for frequency-domain recovery of the full synthetic aperture dataset from focused transmissions," in *2018 IEEE International Ultrasonics Symposium (IUS)*, Oct 2018, pp. 1–9.
[44] A. Ilovitsh, T. Ilovitsh, J. Foiret, D. N. Stephens, and K. W. Ferrara, "Simultaneous axial multifocal imaging using a single acoustical transmission: A practical implementation," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 66, no. 2, pp. 273–284, Feb 2019.
[45] S. Goudarzi, A. Asif, and H. Rivaz, "Multi-focus ultrasound imaging using generative adversarial networks," in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, April 2019, pp. 1118–1121.
[46] Balázs Csanád Csáji, "Approximation with artificial neural networks," *Faculty of Sciences, Etvs Lornd University, Hungary*, vol. 24, pp. 48, 2001.
[47] L. Hongzhou and J. Stefanie, "Resnet with one-neuron hidden layers is a universal approximator," in *Advances in Neural Information Processing Systems 31*, pp. 6172–6181. Curran Associates, Inc., 2018.
[48] D. Pathak, P. Krahenbuhl, J. Donahue, T. Darrell, and A. A. Efros, "Context encoders: Feature learning by inpainting," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 2536–2544.
[49] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016, pp. 770–778.
[50] J. A. Jensen, "Field: A program for simulating ultrasound systems," in *10th 10th Nordic-Baltic Conference on Biomedical Engineering*, 1996, vol. 4, pp. 351–353.
[51] J. A. Jensen and N. B. Svendsen, "Calculation of pressure fields from arbitrarily shaped, apodized, and excited ultrasound transducers," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 39, no. 2, pp. 262–267, March 1992.
[52] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *CoRR*, vol. abs/1412.6980, 2014.
[53] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 2, pp. 295–307, Feb 2016.
[54] I. Goodfellow, "Nips 2016 tutorial: Generative adversarial networks," *arXiv preprint arXiv:1701.00160*, 2016.
[55] Geert Litjens, Thijs Kooi, Babak Ehteshami Bejnordi, Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Mohsen Ghafoorian, Jeroen A.W.M. van der Laak, Bram van Ginneken, and Clara I. Sánchez, "A survey on deep learning in medical image analysis," *Medical Image Analysis*, vol. 42, pp. 60 – 88, 2017.
[56] N. Tajbakhsh, J. Y. Shin, S. R. Gurudu, R. T. Hurst, C. B. Kendall, M. B. Gotway, and J. Liang, "Convolutional neural networks for medical image analysis: Full training or fine tuning?," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1299–1312, May 2016.
[57] H. Liebgott, A. Rodriguez-Molares, F. Cervenansky, J. A. Jensen, and O. Bernard, "Plane-wave imaging challenge in medical ultrasound," in *2016 IEEE International Ultrasonics Symposium (IUS)*, Sep. 2016, pp. 1–4.

Sobhan Goudarzi received his BSc degree from University of Isfahan and MASc from Amirkabir University of Technology, Tehran, Iran. He is currently pursuing his PhD in electrical and computer engineering at IMPACT lab at Concordia University, Montreal, Canada. His research interest are biomedical signal and image processing, medical ultrasound imaging, and machine learning.

Amir Asif (M'97–SM'02) received the M.Sc. and Ph.D. degrees in electrical and computer engineering from Carnegie Mellon University (CMU), Pittsburgh, PA in 1993 and 1996, respectively. He is currently serving as Vice President, Research and Innovation and professor of electrical engineering and computer science at York University. Previously, he has served as the Dean of Engineering and Computer Science at Concordia University (2014 – 2020), professor of electrical engineering and computer science at York University, Toronto, ON, Canada, from (2002 – 2014) and research faculty at CMU (1997 – 1999).

Asif works in the area of statistical signal processing and communications. His current projects include distributed agent networks (autonomy and consensus in complex and contested environments); medical imaging (ultrasound elastography, brain computer interfaces), data science (graph signal processing in social networks); and health management of mission critical systems. He has authored over 175 technical contributions, including invited ones, published in international journals and conference proceedings, and the textbook "Continuous and Discrete Time Signals and Systems" published by the Cambridge University Press.

Asif has served on the editorial boards of numerous journals and international conferences, including Associate Editor for IEEE Transactions of Signal Processing (2014-18), IEEE Signal Processing letters (2002-2006, 2009-2013). He has organized a number of IEEE conferences on signal processing theory and applications. He has received several distinguishing awards including the York University Faculty of Graduate Studies Teaching Award in 2008; York President's University Wide Teaching Award (Full-Time Senior Faculty Category) in 2008 and Science and Engineering Teaching Excellence Award (Senior Faculty Category) from York's Faculty of Science and Engineering in 2006 and in 2004. He sits on several boards of directors, including Research Canada and Making the Shift Inc., and has served as the Chair, Research Committee of Engineering Deans Canada. Asif is a member of the Professional Engineering Society of Ontario and a senior member of IEEE.



Hassan Rivaz received BSc degree from Sharif University of Technology, MASc from the University of British Columbia, and PhD from Johns Hopkins University.

He directs the IMPACT lab: IMage Processing And Characterization of Tissue, and is a Concordia University Research Chair in Medical Image Analysis. His research interests are medical image analysis, machine learning, deep learning, elastography and quantitative ultrasound.

He is an Associate Editor of IEEE Transactions on Medical Imaging (TMI), and IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control (TUFFC). He is a member of the Organizing Committee of IEEE EMBC 2020 (Montreal, Canada), IEEE ISBI 2021 (Nice, France), and IEEE IUS 2023 (Montreal, Canada).

He has served as an Area Chair of MICCAI since 2017. He co-organized two tutorials on advances in ultrasound imaging at IEEE ISBI 2019 and 2018. He also co-organized the CURIOUS 2018 and CURIOUS 2019 Challenges on registration of ultrasound and MRI in neurosurgery, and the CereVis 2018 Workshop on Cerebral Data Visualization, all in conjunction with MICCAI. He was a Petro-Canada Young Innovator from 2016 to 2018.