PAPER

Training improvements for ultrasound beamforming with deep neural networks

View the article online for updates and enhancements.

Physics in Medicine & Biology

IPEM Institute of Physics and Engineering in Medicine

PAPER

Training improvements for ultrasound beamforming with deep neural networks

A C Luchies[1] and B C Byram

Biomedical Engineering, Vanderbilt University, Nashville, TN, United States of America
[1] Author to whom any correspondence should be addressed.

E-mail: adam.c.luchies@vanderbilt.edu

Abstract

This paper investigates practical considerations of training ultrasound deep neural network (DNN) beamformers. First, we studied training DNNs using the combination of multiple point target responses instead of single point target responses. Next, we demonstrated the effect of different hyperparameter settings on ultrasound image quality for simulated scans. This study also showed that DNN beamforming was robust to electronic noise. Next, we showed that mean squared error validation loss was not a good predictor for image quality for simulation, phantom, and *in vivo* scans. As an alternative to validation loss for selecting DNN beamformers, we studied image quality in physical phantom and *in vivo* scans and demonstrated that DNN beamformer image quality in these settings was correlated to DNN beamformer image quality in simulated images. These findings suggest that simulated image quality can be used to select DNN beamformers. Finally, we studied the effect of dataset size on DNN beamformer image quality in simulation, physical phantom, and *in vivo* scans. We interpret the results in terms of recent work on the scaling of deep learning. Overall, the results in this paper show that DNN beamforming has significant potential for improving B-mode image quality.

1. Introduction

The goal of delay and sum (DAS) beamforming is to apply delays and amplitude weights to received sensor signals such that waves coming from one direction are reinforced during summation relative to waves coming from other directions (Johnson and Dudgeon 1993). Usually these delays and amplitude weights are static and do not depend on the received data. In contrast, adaptive beamformers adjust the beamforming process depending on the received signals (Flax and O'Donnell 1988, Nikoonahad and Liv 1990, Li and Li 2003, Dahl *et al* 2006, Holfort *et al* 2009, Byram and Jakovljevic 2014, Byram *et al* 2015). In one type of adaptive imaging, a near field phase screen is used to model wavefront aberration due to sound speed inhomogeneities. After estimation, this phase screen is used to adjust the focusing delays (Flax and O'Donnell 1988, Nikoonahad and Liv 1990, Dahl *et al* 2006). In another adaptive technique, the covariance of the received signals is used to select the amplitude and phase weights to suppress waves coming from directions other than the look direction (Holfort *et al* 2009).

Recently, several approaches were investigated for using deep neural networks (DNNs) for the purposes of ultrasound image reconstruction (Gasse *et al* 2017a, Luchies and Byram 2017, Perdios *et al* 2017). One of the primary challenges of this endeavor is that DNN training requires a well labeled training dataset, which is difficult to create or obtain for ultrasound channel data. The goal of Gasse *et al* was to use neural networks to reduce the number of transmit events needed for synthetic aperture imaging (Gasse *et al* 2017b). A synthetic aperture method (i.e. coherent plane wave compounding) was used to create an ultrasound channel dataset with improved resolution and contrast (Gasse *et al* 2017b). DNNs were then trained to produce images with similar image quality improvements, but using many fewer transmit events. The disadvantage of this approach is that image quality is limited by the image quality using synthetic aperture methods.

We developed a method that relied on simulation tools to generate training data (Jensen and Svendsen 1992, Jensen 1996, Luchies and Byram 2018a, 2018b). The advantage of using linear ultrasound simulation tools is that
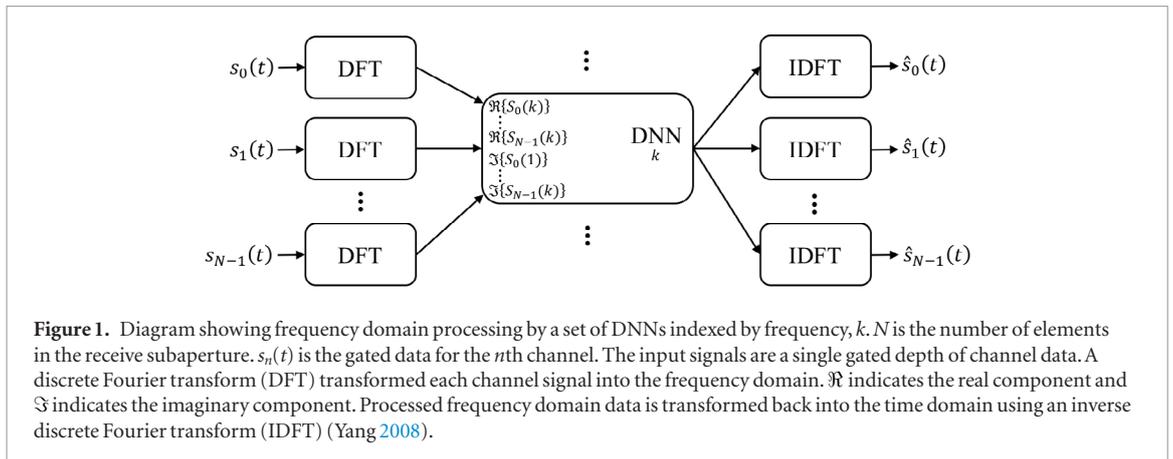
Figure 1. Diagram showing frequency domain processing by a set of DNNs indexed by frequency, k. N is the number of elements in the receive subaperture. $s_n(t)$ is the gated data for the nth channel. The input signals are a single gated depth of channel data. A discrete Fourier transform (DFT) transformed each channel signal into the frequency domain. $\Re$ indicates the real component and $\Im$ indicates the imaginary component. Processed frequency domain data is transformed back into the time domain using an inverse discrete Fourier transform (IDFT) (Yang 2008).

they are fast and make it possible to generate as much training data as desired. The disadvantage of using simulation is that the DNNs will be limited by the accuracy of the simulation tool. Our results showed that it was possible to train DNNs using simulations to produce image quality improvements in physical phantoms and *in vivo* scans (Luchies and Byram 2018a).

The DNN beamforming method that we developed was inspired by the aperture domain model image reconstruction (ADMIRE) beamforming method that Byram *et al* developed (Byram and Jakovljevic 2014, Byram *et al* 2015, Dei and Byram 2017). ADMIRE poses beamforming as a nonlinear regression problem and requires a significant amount of computation to perform beamforming. A DNN is a prime candidate for solving nonlinear regression problems and most of the computation is completed during the training phase. The inference phase for DNNs can be fast.

The goal of this work was to use empirical studies to examine practical considerations of training DNN beamformers. Previously, we trained DNNs with the responses of single point targets (Luchies and Byram 2018a). In this work, we trained DNNs with the responses from multiple point targets and studied the effect of these different training datasets on image quality. We extend our previous finding that mean squared error validation loss was not a good predictor for ultrasound image quality in simulation scans to physical phantom and *in vivo* scans (Luchies and Byram 2018b). We show that DNN beamformer image quality of simulated images was correlated with experimental and *in vivo* image quality and proposed selecting DNN beamformers based on simulated images instead of validation loss. We studied the noise suppressing ability of DNN beamformers. Finally, we studied the effect of the training dataset size on DNN beamformer image quality.

2. Methods

We briefly review the DNN beamformer that we developed in the past and also discuss several innovations (Luchies and Byram 2018a).

2.1. Frequency domain processing
The first step in the signal path was to convert the channel data to the frequency domain using a short-time Fourier transform (STFT). The STFT gate length was 16 samples (approximately one pulse length) and a 16 point DFT was used. The gate overlap was 90% and a rectangular window was used. A diagram showing the frequency domain processing by a set of DNNs is in figure 1.

2.2. Neural networks
A DNN beamformer is a set of DNNs to process data for different DFT bins. In this work, a DNN beamformer consisted of three DNNs, each trained to process a different DFT bin. The DFT bins were for the transmit center frequency and the adjacent DFT bin on either side of this center frequency. These DFT bins covered the 20 dB bandwidth for the pulse-echo characteristics specified. If the sampling frequency were increased, the number of DNNs in the beamformer would also need to be increased to maintain the filtering of the entire 20 dB bandwidth. The remaining DFT bins were zeroed out when transforming back to the time domain, which is an approach used elsewhere (Byram *et al* 2015, Dei and Byram 2017).

We trained fully connected feed-forward multilayer networks to process aperture data and performed a hyperparameter search with parameters indicated in table 1. The hyperparameters were specified on a grid and random configurations of hyperparameters were selected for creating DNN beamformers. For a single DNN beamformer, all DNNs were trained with the same hyperparameters and model architecture. A total of 100 DNN beamformers were trained. A random grid search was selected in order to aid in identifying image quality trends

Table 1. Hyperparameter search space.

Parameter	Search values
Training dataset #	1, 2, 3
Batch size	50, 100, 500, 1000
Number of hidden layers	1–5
Layer width	65, 130, 260, 520
Input Gaussian noise	True or false
Input dropout probability	0, 0.1, 0.2
Hidden node dropout probability	0, 0.1, 0.2, 0.3, 0.4, 0.5
L2 weight decay	$0, 10^{-5}, 10^{-4}$

for the hyperparameters used for training DNN beamformers. Previously, we studied the effect of hyperparameters on mean squared error validation loss (Luchies and Byram 2018a). The goal of the hyperparameter search in this study was to study the effect of hyperparameters on ultrasound image quality.

Gradient descent is by far the most popular algorithm for training neural networks (Ruder 2016). In addition, the learning rate is widely viewed as the most important hyperparameter to tune when training neural networks (Goodfellow *et al* 2016). Adaptive learning rate methods, such as Adam (adaptive moment estimation), adjust the learning rate based on parameters themselves. The learning rate is increased for parameters with infrequent or small valued gradients and decreased for parameters with frequent or large valued gradients. We used Adam with the values suggested by Kingma *et al*, including $\alpha = 10^{-3}$ (learning rate), $\beta_1 = 0.9$ and $\beta_2 = 0.999$ (coefficients used for computing running averages of the gradient and its square), and $\epsilon = 10^{-8}$ (a term to improve numerical stability of the gradient update) (Kingma and Ba 2015).

The DNNs used the rectified linear unit (ReLU) for the activation function (Glorot *et al* 2011). Mean squared error was used as the loss function during training. The weights of the network were initialized using the probability distribution given by Glorot and Bengio (2010) and He *et al* (2015)

$$W \sim N\left(0, \sqrt{\frac{2}{n}}\right) \tag{1}$$

where $N(\mu, \sigma^2)$ is a normal distribution with mean μ and variance σ^2, and n is the size of the previous layer. This probability distribution was developed based on expected activation rates when using the ReLU activation (He *et al* 2015). Initializing weights in this manner avoids reducing or magnifying input signal magnitudes as signals propagate through the network and improves network convergence. The biases for each neuron were initialized to a value of 0.01. Training was terminated if the validation loss did not improve after 20 epochs.

We examined using different levels of dropout probability for the hidden layers as indicated in table 1 (Srivastava *et al* 2014). The dropout probability used for inputs is usually smaller than that used for the hidden nodes. The input dropout probability values are in table 1. We studied using different amounts of *L2* weight decay as indicated in table 1. Input Gaussian noise was also enabled or disabled to study whether input noise improved DNN beamformers. When enabled, white Gaussian noise was added to batches of training and validation data with variable SNR in the range 0–40 dB. These methods have the effect of regularizing a DNN, can prevent a network from overfitting (i.e. memorizing the training data), and aid in network generalization.

During training, DNN inputs were normalized such that individual input vectors had maximum norm equal to one. During the inference phase, the input vector was normalized by its maximum norm and the DNN output was renormalized by the input maximum norm.

Pytorch was used to create and train all of the DNNs in this work (Paszke *et al* 2017). Training was performed on a GPU computing cluster maintained by the Advanced Computing Center for Research and Education at Vanderbilt University.

2.3. Training data

All training data were generated using Field II, which is a program for simulating pressure fields for ultrasonic sources using linear acoustics (Jensen and Svendsen 1992, Jensen 1996). The physical modeling of the acoustic wave propagation studied in this work is reviewed briefly and the interested reader is referred to existing work for a more thorough treatment (Jensen 1991, Jensen and Svendsen 1992). Jensen derived a linear inhomogeneous wave equation with a scattering term that is a function of density and speed of sound variations and given by Jensen (1991)

Figure 2. (a) Dataset #1 consisted of the responses from single scatterers. (b) One half of dataset #2 consisted of responses from a single scatterer (similar to dataset #1) and the second half consisted of the combined responses from two scatterers at different locations. (c) One third of dataset #3 consisted of responses from a single scatterer (similar to dataset #1), the second third consisted of the combined responses from two scatterers at different locations (similar to dataset #2), and the final third consisted of the combined responses from scatterers at three different locations.

$$\nabla^2 p - \frac{1}{c_0^2}\frac{\partial^2 p}{\partial t^2} = \frac{-2\Delta c}{c_0^3}\frac{\partial^2 p}{\partial t^2} + \frac{1}{\rho_0}\nabla(\Delta\rho)\cdot\nabla p, \qquad (2)$$

where p is the acoustic pressure, which is small compared to ambient pressure, Δc represents small sound speed variations about a mean value c_0, and $\Delta\rho$ represents small density variations around a mean value ρ_0. In order to model the received signals of an ultrasound imaging system, equation (2) needs to be solved by including information about the ultrasonic source and receiver (assumed here to be the one and the same) and also the scattering sites.

The emitted field for the ultrasound transducer can be modeled as Jensen and Svendsen (1992)

$$p_t(\vec{r}_1, \vec{r}_2, t) = v(t) * E_m(t) * h(\vec{r}_1, \vec{r}_2, t) \qquad (3)$$

where the convolutions are in the time domain, $\vec{r}_1$ is the field location, $\vec{r}_2$ is the transducer location, $v(t)$ is the excitation signal, $E_m(t)$ is the electromechanical impulse response of the transducer, and $h(\vec{r}_1, \vec{r}_2, t)$ is the spatial impulse response of the transducer. The spatial impulse response is the Green's function for bounded space integrated over the transducer surface (Jensen 1991). The Green's function for bounded space is used because the transducer is assumed to be on an infinite, rigid baffle.

Using acoustic reciprocity, the received pressure field at position $\vec{r}_2$ due to a small spherical source at position $\vec{r}_1$ can be modeled as Jensen (1991)

$$p_r(\vec{r}_1, \vec{r}_2, t) = E_m(t) * h(\vec{r}_2, \vec{r}_1, t) \qquad (4)$$

where the convolution is in the time domain and $E_m(t)$ is the electromechanical impulse response of the transducer. Finally, the pulse-echo response from a distribution of scatterers given by $f(\vec{r}_1)$ can be found by convolving the emitted field with the received field in the time domain and then convolving the result in the spatial domain with the scattering function.

Field II computes the spatial impulse response efficiently by splitting the transducer surface into rectangular elements that are small compared to the wavelength and summing the analytic far-field responses from these small rectangular elements (Jensen and Svendsen 1992). The DNN beamformers in this work can be thought of as learning how to recognize the pulse-echo spatial impulse response for the imaging system.

Three datasets were created and used for training DNN beamformers. The three datasets were treated as a hyperparameter as in table 1. Dataset #1, dataset #2, and dataset #3 are described in figures 2(a)–(c) respectively. Note that dataset #1 had the same configuration as our previous work on training DNN beamformers (Luchies and Byram 2018a), while dataset #2, and dataset #3 are novel. In all methods, scatterers were randomly placed in an annular sector as depicted in figure 2. The annular sector had a width of 50 pulse lengths. The responses from a particular scatterer were only kept if they appeared within the STFT gate that was centered at the transmit focus. The goal was to train a network for this STFT gate.

The scatterers were divided into two groups, including on-axis and off-axis scatterers. The division was set as the region between the nulls of the main lobe for a simulated beam at the transmit center frequency. For on-axis scatterers, the target signal was the same as the input signal. For off-axis scatterers, the target signal was a vector of zeros. When an input signal was the combination of on-axis and off-axis scatterers as in datasets #2 and #3, the target consisted of the summation of on-axis signals. If the input consisted of the responses from only off-axis scatterers, the target was a vector of zeros.

Table 2. Linear array scan parameter values.

Parameter	Value
Active elements	65
Transmit frequency	5.208 MHz
Pitch	298 μm
Kerf	48 μm
Simulation sampling frequency	520.8 MHz
Experimental sampling frequency	20.832 MHz
Speed of sound	1540 m s^{-1}
Transmit focus	70 mm

The simulation parameters are in table 2. The simulated transducer was modeled after an ATL L7-4 (38 mm) linear array transducer. For each dataset described in figure 2, a training and a validation dataset were created. The training dataset was only used to adjust weights. The validation dataset was used to stop training and studied as a method for DNN beamformer selection. The training datasets had 100 000 examples and the validation datasets had 10 000 examples.

2.4. Generalized coherence factor

Many adaptive beamforming methods have been developed for suppressing off-axis scattering in ultrasound images. Images created using the generalized coherence factor (GCF) were included in this paper to serve as a comparison between the DNN beamformer and other adaptive beamformers (Li and Li 2003). The GCF is defined as

$$\text{GCF} = \frac{\sum_{k=-M_0}^{M_0} |S(k)|^2}{\sum_{k=-\frac{N}{2}}^{\frac{N}{2}-1} |S(k)|^2} \tag{5}$$

where N is the number of elements in the subaperture, $S(k)$ is the discrete Fourier transform across the aperture dimension, k is spatial frequency, and M_0 is the frequency cutoff for the coherent sum in the numerator. We used $M_0 = 3$ for all GCF images in this paper (Wang *et al* 2007).

2.5. Image quality metrics

We quantified image quality using contrast ratio (CR)

$$\text{CR} = -20 \log_{10} \left(\frac{\mu_{\text{lesion}}}{\mu_{\text{background}}} \right), \tag{6}$$

contrast-to-noise ratio (CNR) (Patterson and Foster 1983, Smith *et al* 1983, Smith and Wagner 1984)

$$\text{CNR} = 20 \log_{10} \left(\frac{|\mu_{\text{background}} - \mu_{\text{lesion}}|}{\sqrt{\sigma_{\text{background}}^2 + \sigma_{\text{lesion}}^2}} \right), \tag{7}$$

and speckle signal-to-noise ratio (SNRs)

$$\text{SNRs} = \frac{\mu_{\text{background}}}{\sigma_{\text{background}}}, \tag{8}$$

where μ is the mean and σ is the standard deviation of the uncompressed envelope. CR and CNR require specification of a lesion region and a background region, while SNRs requires only specification of a background region. When estimating electronic SNR between successive frames, we used (Friemel *et al* 1998)

$$\text{SNR} = 10 \log_{10} \left(\frac{\rho}{1 - \rho} \right), \tag{9}$$

where ρ is the correlation coefficient between two successive frames.

2.6. Simulation: channel SNR analysis

We studied the robustness of DNN beamformers to noise using FIELD II simulations (Jensen and Svendsen 1992, Jensen 1996). In each simulation, an anechoic cyst at a depth of 7 cm and having a diameter of 5 mm was imaged using a simulated L7-4 38 mm linear array. No scatterers were located inside the cyst and 25 scatterers per resolution cell were placed in the background region. The noiseless channel data was then corrupted using white

Gaussian noise to vary channel SNR from 10 dB to −10 dB in steps of 5 dB. A total of five simulated anechoic cysts were studied.

2.7. Physical phantom and *in vivo* scans

A linear array transducer (ATL L7-4 38 mm) was operated using a Verasonics Vantage 128 system (Verasonics, Kirkland, WA) to conduct physical phantom scans. The physical phantom was a multipurpose phantom (Model 040GSE, CIRS, Norfolk, VA) and a cylindrical anechoic cyst at a 7 cm depth with an approximately 3 mm diameter was scanned. Five scans were made at different positions along the cylindrical cyst.

To demonstrate that DNN beamformers trained using the approach that we have developed can be used with other ultrasound transducer geometries besides linear arrays, the physical phantom was also scanned using a curvilinear array transducer (ATL C5-2) using a Verasonics Vantage system. Cylindrical cysts at 1.5 cm, 4.5 cm, and 7 cm were scanned. A total of seven scans were conducted after moving the cysts to different locations in the field of view of the transducer. DNN beamformers for a C5-2 array were trained from scratch and training data was generated using Field II.

A linear array transducer (ATL L7-4 38 mm) was operated using a Verasonics Vantage 128 system (Verasonics, Kirkland, WA) to scan the liver of a 36 year old healthy male. Scanning was conducted to look at liver vasculature. The study was approved by the local Institutional Review Board.

2.8. The effect of training dataset size

When training DNNs, one of the most reliable approaches for performance improvement is to increase training dataset size (Hestness *et al* 2017). Studying model performance as a function of this quantity provides a way to predict potential performance improvements using this strategy.

After selecting a DNN beamformer based on simulated image CNR, we retrained this DNN beamformer using the same hyperparameter settings and model architecture multiple times to study the effect of training dataset size. The studied training dataset sizes include 10^2, 10^3, 10^4, and 10^5. For each training dataset size, the DNN beamformer was retrained five times using different starting weights. In each case, the size of the validation dataset was kept constant with a value of 10^4.

3. Results

3.1. Channel SNR analysis

3.1.1. Hyperparameters and image quality

Figure 3 shows image quality for simulated anechoic cysts as a function of channel SNR for different hyperparameters. Each row demonstrates the effect of a different hyperparameter on image quality. For example, figures 3(a)–(c) shows image quality for three DNN beamformers. The first DNN beamformer was selected from the group of beamformers that were trained using dataset #1 and it was selected as the beamformer that had the best performance in terms of CNR. The results for this beamformer are shown as the blue (circle) line. The second DNN beamformer was selected from the group of beamformers that were trained using dataset #2 and it was selected as the beamformer that had the best performance in terms of CNR. The results for this beamformer are shown as the orange (triangle) line. The third DNN beamformer was selected from the group of beamformers that were trained using dataset #3 and it was selected as the beamformer that had the best performance in terms of CNR. The results for this beamformer are shown as the green (square) line. Displaying results in this manner provides a way to visualize the effect of different hyperparameters on image quality, while marginalizing out the remaining hyperparameters.

Figure 3(a) shows that CR was best when training with dataset #1 (single point target responses) or dataset #2 (single point target responses and the combined responses from two point targets). Figure 3(c) shows that speckle SNR was best when training with dataset #2 or dataset #3 (single point target responses, the combined responses from two point targets, and the combined responses from three point targets). Figure 3(b) shows that CNR was comparable for all datasets when the channel SNR was sufficiently high.

Figures 3(e) and (f) show that using a batch size of 500 consistently produced the best CNR and speckle SNR. It should be noted that when disabling network regularization methods such as dropout or weight decay, using smaller batch sizes produced the best results in terms of image quality.

Figures 3(g)–(i) show how image quality improved as the number of hidden layers increased, which is similar to most deep learning applications. However, it should be noted that the network with five hidden layers produced lower image quality than did the network with four hidden layers. Figures 3(k) and (l) show similar trends in that using networks with wider layers improved image quality.

Figure 3(m) show that there was little difference for CNR and speckle SNR when enabling or disabling input Gaussian noise. In contrast, figures 3(p)–(r) show that image quality improved when using input dropout.
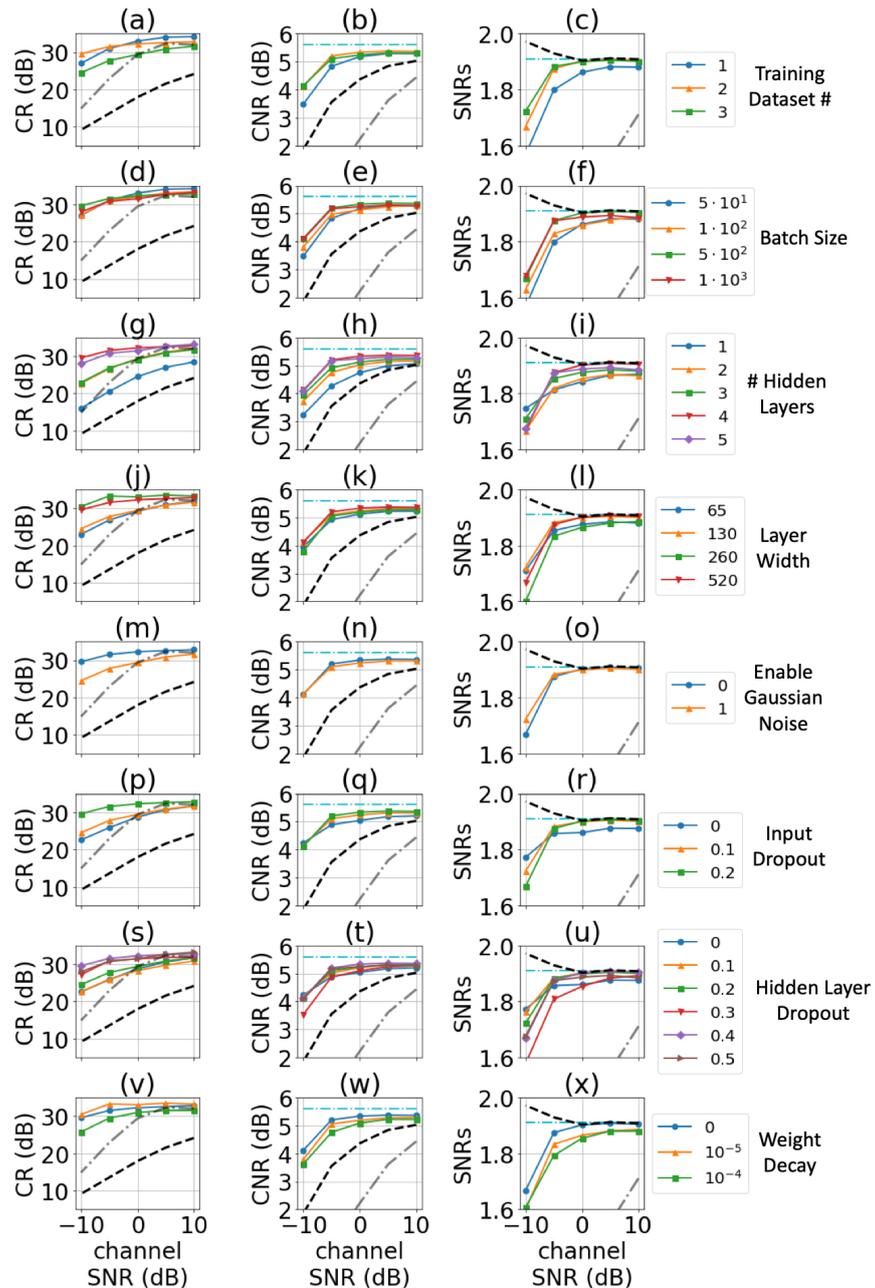
Figure 3. CR, CNR, and speckle SNR for DNN beamformers as a function of channel SNR. (a)–(c) Out of all DNN beamformers that used training dataset #1, we selected the best DNN beamformer in terms of CNR, and display CR, CNR, and SNRs for this beamformer as the blue (circle) line. Out of all DNN beamformers that used training dataset #2, we selected the best DNN beamformer in terms of CNR, and display CR, CNR, and SNRs for this beamformer as the orange (triangle) line. Out of all DNN beamformers that used training dataset #3, we selected the best DNN beamformer in terms of CNR, and display CR, CNR, and SNRs for this beamformer as the green (square) line. Performance of the best DNN beamformer as a function of (d)–(f) batch size, (g)–(i) number of hidden layers, (j)–(l) layer width, (m)–(o) input Gaussian noise, (p)–(r) input dropout, (s)–(u) hidden layer dropout, and (v)–(x) weight decay. For comparison, the dashed black line shows the performance of DAS and the dark gray dashed dotted line shows the performance for GCF. The cyan dashed dotted line shows the theoretical limit for CNR.

Figures 3(s)–(u) show that using some dropout for the hidden layers improved image quality. This finding is in contrast to our previous work which suggested that dropout was not advantageous for training networks for DNN beamforming (Luchies and Byram 2018a). However, in that work we showed DNN beamformer training results in terms of validation loss and not image quality. Figures 3(v)–(x) show that enabling weight decay did not provide any benefit for image quality compared to disabling weight decay.

Overall, figure 3 shows that the trained DNN beamformers were robust to noise. The improvement in CNR of DNN beamformers compared to DAS actually increased for higher levels of noise. Finally, DNN beamformers produced better CNR than DAS and GCF, irrespective of the examined noise levels.

Figure 4. CNR as a function of mean squared error validation loss for DNN beamformers trained with (a) dataset #1, (b) dataset #2, and (c) dataset #3. The validation loss is the average validation loss for all DNNs in a DNN beamformer. The CNR values are average values across the channel SNR range studied in the anechoic cyst simulation. The black dashed line indicates CNR for DAS and the gray dashed dotted line indicates CNR for GCF. (d) Each blue line represents the channel SNR after processing by a specific a DNN beamformer. The dashed black line is the channel SNR for the best DNN beamformer selected based on CNR. The dashed-dotted gray line is the input channel SNR.

3.1.2. Validation loss and image quality

Typically, DNNs are selected based on the validation loss. However, figure 4 shows that network mean squared error validation loss was not a good predictor of image quality (Luchies and Byram 2018b). The R^2 for figures 4(a)–(c) were -0.05, -0.62, and -0.12, respectively. The best DNN beamformer in terms of CNR was trained with dataset #2 and the validation loss was an order of magnitude higher than the DNN beamformer with the lowest validation loss. The DNN beamformer with lowest validation loss produced CNR that was 15%–30% lower than the DNN beamformer with best CNR. Figure 4(a) also shows how the CNR variance amongst DNN beamformers that used dataset #1 was larger than the CNR variance for DNN beamformers that used datasets #2 and #3.

3.1.3. Channel SNR improvements

Figure 4(d) shows that the trained DNN beamformers increased channel SNR by at least 10 dB and up to 20 dB compared to the input channel SNR. These results provide evidence of the denoising ability of DNN beamformers.

3.1.4. DNN beamformer selection and performance

Instead of selecting a DNN beamformer based on validation loss, we selected the best performing DNN beamformer by averaging the measured CNR values for the studied channel SNR values and picking the DNN beamformer with the best average CNR. The hyperparameter values for the DNNs in this beamformer were as follows: the DNNs were trained using dataset #2, 500 batch size, 4 hidden layers, no Gaussian noise was added to the inputs, 0.2 input dropout probability, 0.4 hidden layer dropout probability, and no weight decay was used. Simulated images of anechoic cysts using this DNN beamformer are in figure 5, which demonstrate the robustness of DNN beamforming to electronic noise along with the improvements to image quality offered by DNN beamforming.

3.2. Physical phantom

3.2.1. Validation loss and image quality

Figure 6 shows that mean squared error validation loss was also not a good predictor for image quality in the physical phantom scans. The R^2 for figures 6(a)–(c) were -0.17, -0.02, and 0.05, respectively.
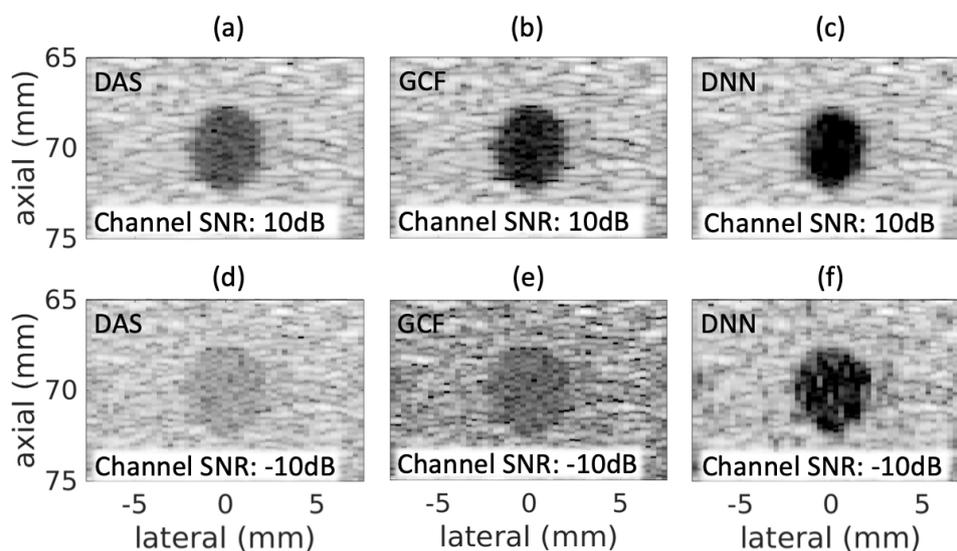
Figure 5. ((a), (d)) DAS, ((b), (e)) GCF, and ((c), (f)) DNN images for (a–c) high and (d–f) low channel SNR. Images shown with 60 dB dynamic range.
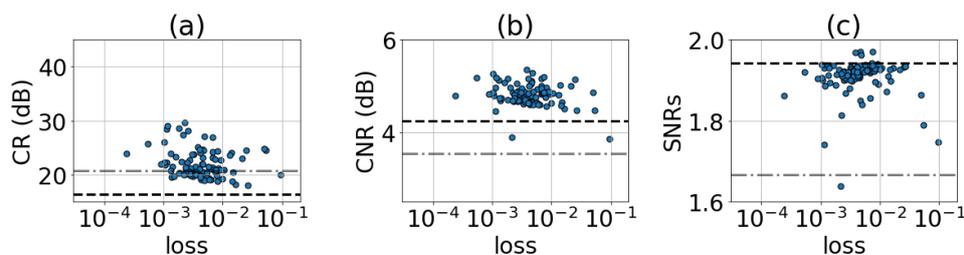
Figure 6. (a) CR, (b) CNR, and (c) SNRs for an anechoic cyst inside a physical phantom as a function of mean squared error validation loss. The validation loss is the average validation loss for all DNNs in a DNN beamformer. The black dashed line indicates values for DAS and the gray dashed dotted line indicates values for GCF.
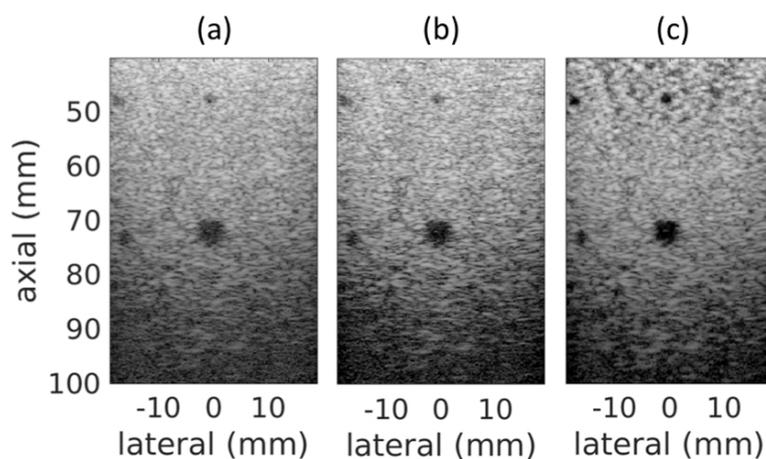
Figure 7. (a) DAS and (b) GCF, and (c) DNN images for anechoic cyst in a physical phantom. Images shown with 60 dB dynamic range.

3.2.2. DNN beamformer performance

Figure 7 provides a comparison of DAS and DNN beamformer images when using the DNN beamformer with the best CNR on simulated images. Contrast improvements are noticeable in the DNN beamformer image. After a depth of about 55 mm, the speckle pattern of the DNN beamformer image is identical to that of DAS. At shallower depths, the speckle variance of the DNN beamformer image increased. However, contrast for the small anechoic cysts at 50 mm depth also improved in the DNN beamformer image compared to the DAS image.

Figure 8. Scatter plots for (a) CR, (b) CNR and (c) SNRs measured on simulated data and on physical phantom data. Each circle represents an individual DNN beamformer.

Figure 9. (a) DAS and (b) DNN images for anechoic cysts in a physical phantom using a curvilinear array. Images shown with 60 dB dynamic range.
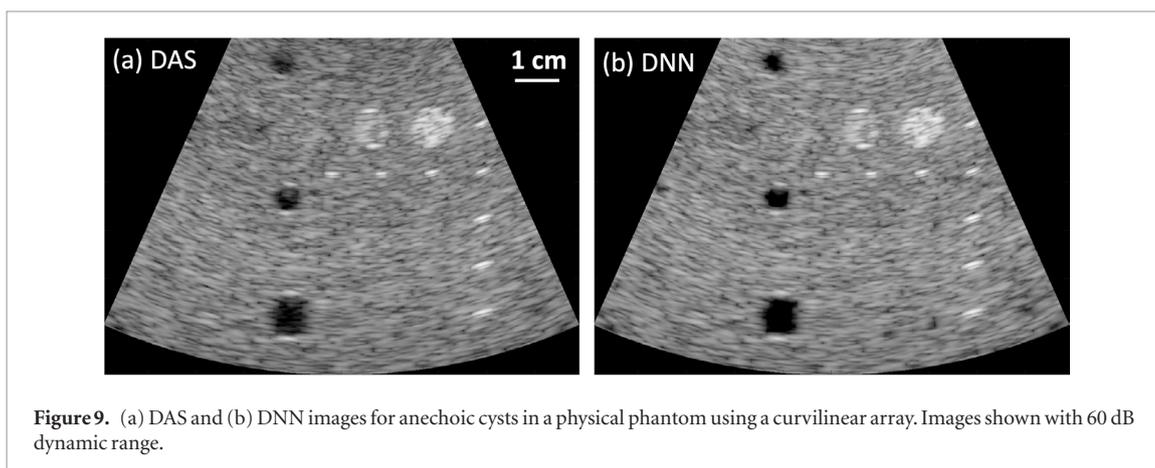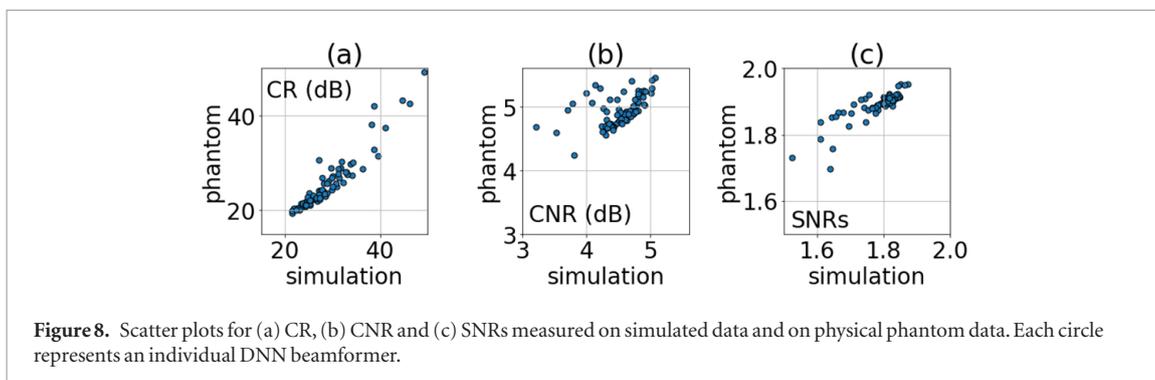
It should also be noted that the DNN beamformer was trained assuming a beam centered in the active subaperture. In the lateral regions of the scan in figure 7 (i.e. less than -1 cm and greater than 1 cm), the beams were not centered in the active subaperture. For the DNN beamformer, speckle variance appeared to increase slightly in these regions compared to DAS; however, training DNN beamformers to account for non-centered beams should reduce this artifact.

3.2.3. Image quality in simulation and phantom scans

Figure 8 shows scatter plots for image quality metrics measured in simulation compared to image quality metrics measured in a physical phantom. The R^2 coefficient for CR, CNR, and speckle SNR was 0.95, 0.85 and 0.94, respectively. In general these results show that for the studied DNN beamformers, the image quality measured in simulation correlated with the image quality measured in the physical phantom. This finding suggests that selecting DNN beamformers based on simulated images is a better strategy than selecting based on validation loss.

3.2.4. Phantom scans using a curvilinear array

The phantom scans using a curvilinear array are in figure 9. For DAS and DNN beamforming, the CR for the 4.5 cm deep cyst was 19.4 ± 2.3 dB and 22.4 ± 5.2 dB, respectively. For DAS and DNN beamforming, the CNR for the 4.5 cm deep cyst was 4.4 ± 0.6 dB and 5.1 ± 0.6 dB, respectively. These reported speckle statistics demonstrate quantitative improvements to image contrast using DNN beamforming that support the qualitative improvements that are visible in figure 9.

3.3. *In vivo*

3.3.1. Validation loss and image quality

Figure 10 shows that mean squared error validation loss was not a good predictor for image quality in *in vivo* scans, which is consistent with simulation and physical phantom scans. The R^2 for figures 10(a)–(c) were -0.07, 0.03 and 0.18, respectively. Each point on the scatter plot represented the average across seven images and a single pair of background and inside regions were selected in each image.

3.3.2. DNN beamformer performance

Figure 11 provide comparison images for DAS and DNN beamforming. The DNN beamformer used was selected as the DNN beamformer with best CNR in simulated images. These results show the increase in contrast when using DNN beamforming. Figure 11(a) also shows the speckle preserving ability of DNN beamformers.
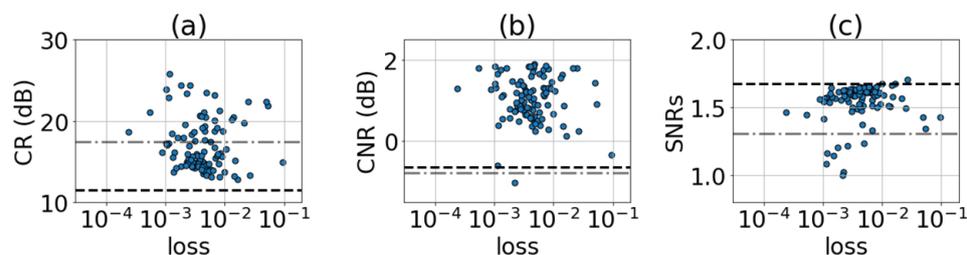
Figure 10. (a) CR, (b) CNR, and (c) SNRs for *in vivo* scans as a function of DNN beamformer mean squared error validation loss. The validation loss is the average validation loss for all DNNs in a DNN beamformer. The black dashed lines indicate values for DAS and the gray dashed dotted lines indicate values for GCF.
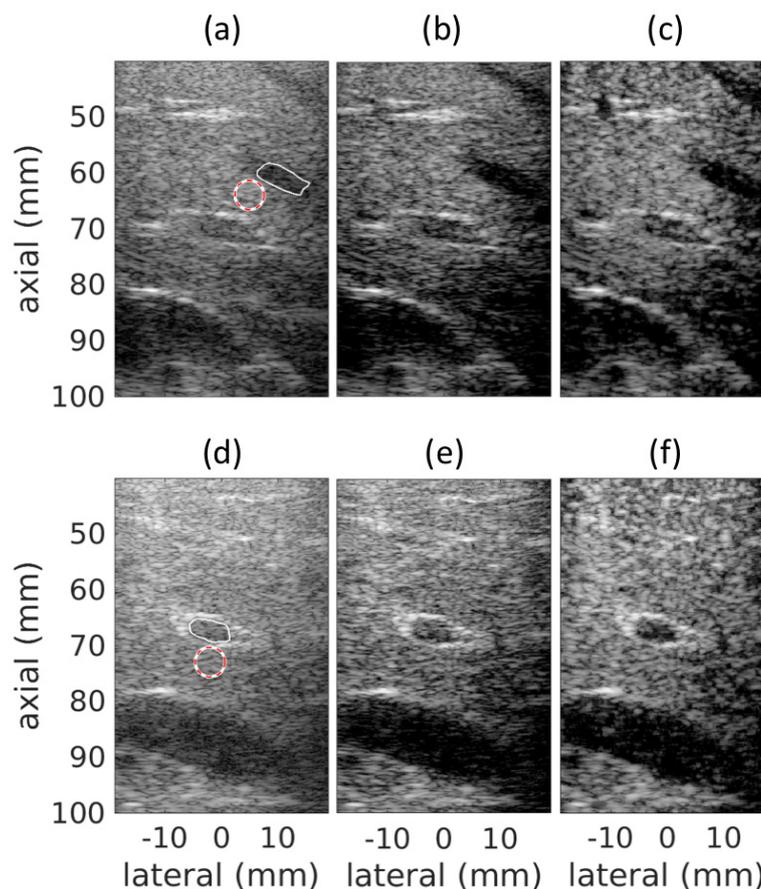
Figure 11. *In vivo* scans of human liver using ((a), (d)) DAS, ((b), (e)) GCF and ((c), (f)) DNNs. The solid white lines shows the inside region and the dashed red line shows the background region used to calculate image quality metrics. Images shown with 60 dB dynamic range.

Similar depths of field are also apparent for the DNN beamformer in physical phantom and *in vivo* images when comparing figures 7 and 11.

3.3.3. Image quality in simulation and in vivo scans

Figure 12 shows scatter plots for image quality metrics measured in simulation compared to image quality metrics measured in *in vivo* scans. The R^2 coefficient for CR, CNR and speckle SNR was 0.84, 0.84 and 0.78, respectively. In general these results show that for the studied DNN beamformers, the image quality measured in simulation correlated with the image quality measured in the *in vivo* scans. This finding provides further evidence that selecting DNN beamformers based on simulated images is a better strategy than selecting based on validation loss.

Figure 12. Scatter plots for (a) CR, (b), CNR, and (c) SNRs measured on simulated data and *in vivo*. Each circle represents an individual DNN beamformer.
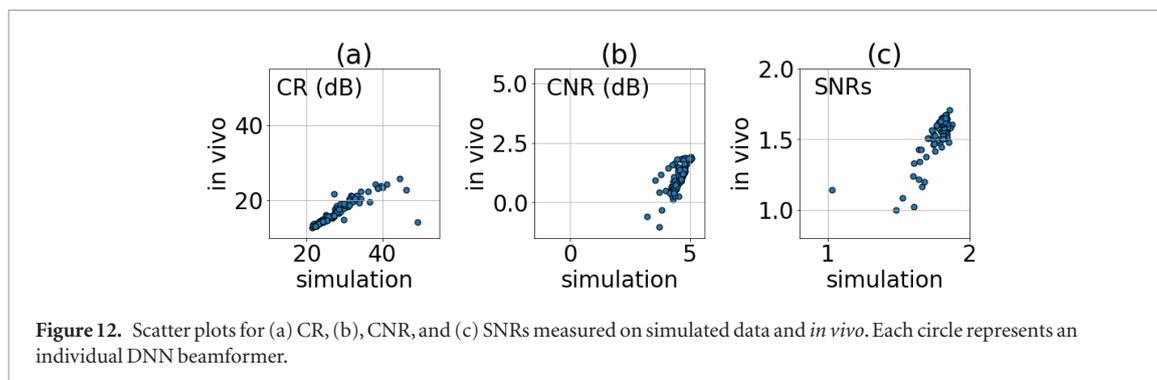
3.4. Training dataset size

Figure 13 shows the effect of training dataset size on image quality for simulation, physical phantom, and *in vivo* scans. In particular, there were large improvements in CR and CNR between dataset sizes of 10^3 and 10^4 for simulation, physical phantom, and *in vivo* scans. In addition, the variance amongst DNN beamformers trained at the same dataset size decreased noticeably for training dataset sizes equal to or greater than 10^4 in physical phantom and *in vivo* scans. Finally, the findings in figure 13 show that DNN beamformers trained with the same settings but with different weight initialization produced consistent image quality as long as the training dataset size was large enough.

4. Discussion

Figures 4, 6 and 10 demonstrate the wide range of image quality achieved by the different DNN beamformers for simulated, physical phantom, and *in vivo*. These results are significant because they show that almost all of the DNN beamformers trained in this work improved image contrast in simulation, experimental, and *in vivo* scans. Previously, we showed this result, but only for simulated scans (Luchies and Byram 2018b). It should be noted that some of the DNN beamformers also caused severe speckle degradation. The DNN beamformers that performed well in terms of image quality probably learned the best hidden representation for suppressing off-axis scattering.

Although CNR in figure 3(b) was comparable for all of the studied training datasets, we would argue that it is preferable to train using dataset #2 or dataset #3 because these datasets produced speckle SNR values that were comparable to DAS. We expect that it will be more feasible in the future to improve the CR results when training with dataset #2 or #3 than it will be to improve the speckle SNR results with dataset #1.

Our goal for this work was to train DNNs that produced better quality images but still possessed the fundamental features of B-Mode ultrasound images. The primary metric for determining image quality is CNR, but our objective of preserving fundamental features of B-Mode places an upper limit of 5.6 dB on CNR (Dahl *et al* 2006). With this in mind, figures 4, 6 and 10 show that mean squared error validation loss was not a good predictor for ultrasound image quality. As an initial solution to the problem of DNN beamformer selection, we proposed using image quality from simulated images to evaluate and select DNN beamformers. Figures 8 and 12 demonstrated that using simulated image quality to select a DNN beamformer was effective for selecting the DNN beamformer with best image quality for physical phantom and *in vivo* scans.

The results in figure 3 show that DNN beamformers trained with input white Gaussian noise did not improve image quality compared to the DNN beamformers that were trained without this type of input noise. We suspect these results can be explained because the DNN beamformers were trained to remove a perturbation (i.e. off-axis scattering) from a signal of interest (i.e. on-axis scattering) and this removal process encouraged the DNNs to be able to handle other sorts of perturbations such as white Gaussian noise. However, it is curious that using input dropout did offer image quality improvements. These results suggest that DNN beamformers may already be robust to missing or blocked elements (Li *et al* 1993, Jakovljevic *et al* 2017).

The results in figure 10 are significant because they demonstrate CNR improvements using DNN beamformers across three times as many *in vivo* scans than our previous work (Luchies and Byram 2018a). In addition, almost all DNN beamformers studied in this work produced better CNR than DAS and GCF. Figure 11 also shows how DNN beamformers improved image quality over a much larger field of view compared to that suggested by our previous work (Luchies and Byram 2018a).

It should be noted that DNN beamforming adds additional computational cost above and beyond that which is required by DAS. For the images in this work, processing by a DNN beamformer usually took 30 s or less. We are currently working on implementing a GPU based beamformer capable of DNN beamforming. We expect to be able to achieve real-time or close to real-time imaging with DNN beamformers. Previously, Hyun *et al* imple-

Figure 13. Image quality as a function of training sample size for ((a)–(c)) simulation SNR study, ((d)–(f)) physical phantom, and ((g)–(i)) *in vivo* scans. The dashed lines indicate performance achieved by DAS.
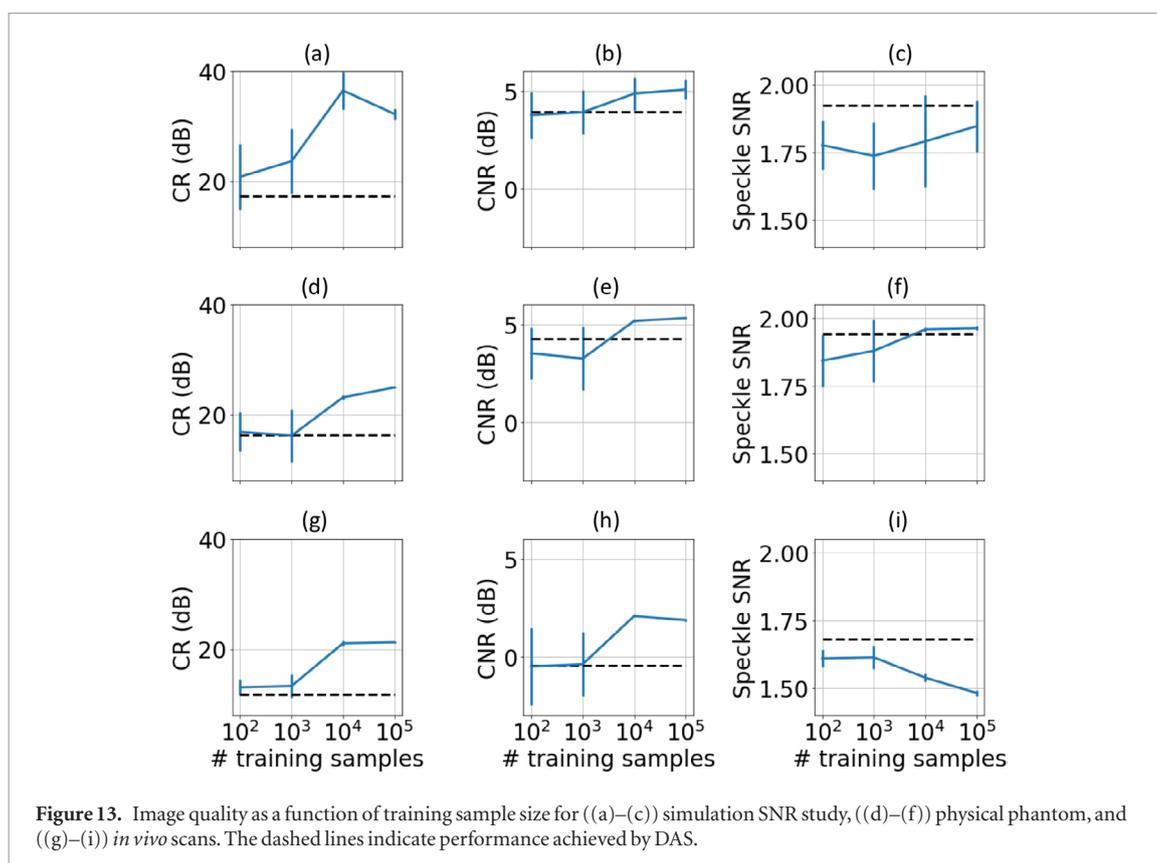
mented short-lag spatial coherence (SLSC) on GPUs and were able to achieve frame rates of about 20 frames per second (Hyun *et al* 2015).

A major challenge that faces the development of DNN beamforming is to verify that anatomical details uncovered by a DNN beamformer correspond to the actual anatomy of the patient and were not mistakes made by the beamformer. Future work is required to develop methods to perform this verification and we describe several methods here.

One strategy to verify the fidelity of DNN beamforming is to compare the results to existing advanced beamformers (e.g. GCF, ADMIRE, etc). For example, there is a blood vessel that is visible on the right side of the image at a depth of 60 mm in figures 11(a)–(c). For the DAS image in figure 11(a), image contrast and also confidence that this region was actually a blood vessel were poor. For the GCF image in figure 11(b), image contrast for this region was high as well as confidence that this region was a blood vessel. For the DNN beamformer image in figure 11(c), image contrast was also high. The agreement between the DNN beamformer image and GCF image for this region of the image serve as verification that the DNN beamformer produced a correct image in this case.

A strategy that correlates anatomical details revealed by a DNN beamformer with those revealed by a different advanced beamformer could be used to systematically build confidence in a DNN beamformer. However, it should be noted that this type of strategy is not useful for assessing anatomical features revealed by a DNN beamformer that were not revealed by other advanced beamformers. We hypothesize that DNN beamformers will reveal details that other beamformers will not reveal.

Developing methods to verify image details in these cases will be the subject of future investigation. For example, advanced ultrasound simulation tools could be used to simulate anatomical details that are masked by reverberation clutter and wavefront aberration (Pinton *et al* 2009). Because the ground truth would be known, it becomes possible to verify that uncovered features are actually present. Similarly, experimental studies could be conducted by imaging phantoms through animal abdominal wall tissue to mask features of interest through wavefront aberration and reverberation. Using a well characterized phantom would allow for verification that all revealed features are present.

If on-axis scattering is considered the signal of interest and off-axis scattering a perturbation that needs to be removed, the DNNs in this work can be viewed as being similar to denoising autoencoders (Vincent *et al* 2008). Autoencoders have several uses including dimensionality reduction, representation learning, and deep neural network pre-traning (Goodfellow *et al* 2016). Interpreting DNN beamformers as denoising autoencoders also allows us to hypothesize that the DNN beamformers operate by learning an encoding stage, perturbation suppression stage, and decoding stage. Compared to the input domain, the hidden representation provides a better space to suppress the unwanted perturbation (i.e. off-axis scattering and noise) while preserving the signal

of interest (i.e. on-axis scattering). It should be noted that methods like ADMIRE can also be interpreted as an encoder, perturbation suppression, and decoder.

Hestness *et al* note that the deep learning community has relied on three recipes for advancing deep learning performance, including searching for improved model architectures, creating larger training datasets, and scaling computation (Hestness *et al* 2017). They note that model architecture advances tend to be serendipitous, whereas the other recipes more reliably improve performance. In addition, they also postulate a power law learning curve as a function of training dataset size and note that in many applications, model architectures merely shift the intercept of this power law and do not change the exponent. In other words, the rate of performance improvement as a function of model size and training dataset size appears to be the same for differing model architectures.

Hestness *et al* argue that DNN learning curves can be divided into three regions (Hestness *et al* 2017). First, there is a flat region called the small data region where DNN performance is poor and constant as a function of training dataset size. Second, is the power law region where performance increases linearly as a function of training dataset size. Finally, there is the irreducible error region, where the performance curve is flat again with increasing training dataset size.

Focusing on figure 13(d), a small data region is apparent for training dataset sizes between 10^2 and 10^3 and CR is comparable to DAS. Next, between 10^3 and 10^5, a power law region is apparent as CR improves dramatically compared to DAS. The findings suggest that further improvements in CR might be achieved using larger training datasets. Improving CR is necessary to expand the contrast ratio dynamic range of DNN beamformers (Dei *et al* 2017). Doing so could enable DNN beamformers to avoid or reduce the dark artifact region that affects many adaptive beamforming methods (Rindal *et al* 2017).

It should be noted that speckle SNR actually decreased for *in vivo* scans in figure 13(i), suggesting that DNN beamformers trained with smaller datasets performed better than those trained with larger datasets. However, we postulate that DNN beamformers trained with larger datasets revealed small blood vessels that were not revealed by DNN beamformers trained using smaller datasets or by DAS. The revelation of these small blood vessels increased speckle variance and decreased speckle SNR. Assuming this hypothesis is true, speckle SNR values measured from *in vivo* liver scans were not a good metric for selecting training dataset size. This phenomena of decreasing speckle SNR as a function of increasing dataset size was not observed in the physical phantom scans, which had uniform scattering background regions.

5. Conclusion

In this work, we report on the noise robustness of DNN ultrasound beamforming. We used simulations to demonstrate that DNN beamformers improved image quality in low and high noise situations and that DNN beamformers increased channel SNR by 10–20 dB. In addition, we studied the effect of different hyperparameters on ultrasound image quality. We also demonstrated that the mean squared error validation loss of DNN beamformers was not a good predictor for image quality for simulation, phantom, and *in vivo* scans, which motivates trying to identify a new loss function in the future. Based on these findings, we studied image quality for all of the trained DNN beamformers in physical phantom and *in vivo* scans and showed that the image quality in physical phantom scans and *in vivo* scans correlated with simulation image quality. These results suggest that selecting a DNN beamformer based on simulated image quality is a better strategy than selecting based on mean squared error validation loss. Overall, the results in this report demonstrate the potential of using DNN beamformers to improve B-mode ultrasound image quality.

Acknowledgment

ORCID iDs

A C Luchies ⬤ https://orcid.org/0000-0002-8211-2422

References

Byram B, Dei K, Tierney J and Dumont D 2015 A model and regularization scheme for ultrasonic beamforming clutter reduction *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **62** 1913–27
Byram B and Jakovljevic M 2014 Ultrasonic multipath and beamforming clutter reduction: a chirp model approach *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **61** 428–40
Dahl J J, McAleavey S A, Pinton G F, Soo M S and Trahey G E 2006 Adaptive imaging on a diagnostic ultrasound scanner at quasi real-time rates *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **53** 1832–43

Dei K and Byram B 2017 The impact of model-based clutter suppression on cluttered, aberrated wavefronts *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **64** 1450–64

Dei K, Luchies A and Byram B 2017 Contrast ratio dynamic range: a new beamformer performance metric *Proc. IEEE Ultrasonics Symp. (Washington, DC, 6–9 September 2017)* (https://doi.org/10.1109/ULTSYM.2017.8091763)

Flax S W and O'Donnell M 1988 Phase-aberration correction using signals from point reflectors and diffuse scatterers: basic principles *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **35** 758–67

Friemel B H, Bohs L N, Nightingale K R and Trahey G E 1998 Speckle decorrelation due to two-dimensional flow gradients *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **45** 317–27

Gasse M, Millioz F, Roux E, Garcia D, Liebgott H and Friboulet D 2017a Accelerating plane wave imaging through deep learning-based reconstruction: an experimental study *Proc. of IEEE Ultrasonics Symp.*

Gasse M, Millioz F, Roux E, Garcia D, Liebgott H and Friboulet D 2017b High-quality plane wave compounding using convolutional neural networks *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **64** 1637–9

Glorot X and Bengio Y 2010 Understanding the difficulty of training deep feedforward neural networks *Proc. AISTATS* pp 249–56

Glorot X, Bordes A and Bengio Y 2011 Deep sparse rectifier neural networks *Proc. AISTATS* pp 315–23

Goodfellow I, Bengio Y and Courville A 2016 *Deep Learning* (Cambridge, MA: MIT Press)

He K, Zhang X, Ren S and Sun J 2015 Delving deep into rectifiers: surpassing human-level performance on imagenet classification *Proc. 2015 IEEE Int. Conf. on Computer Vision (Santiago, Chile, 7–13 December 2015)* pp 1026–34

Hestness J, Narang S, Ardalani N, Diamos G, Jun H, Kianinejad H, Patwary M M A, Yang Y and Zhou Y 2017 Deep learning scaling is predictable, empirically (arXiv:1712.00409v1)

Holfort I K, Gran F and Jensen J A 2009 Broadband minimum variance beamforming for ultrasound imaging *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **56** 314–25

Hyun D, Trahey G E and Dahl J J 2015 Real-time high-frame rate *in vivo* cardiac SLSC imaging with a GPU-based beamformer *Proc. of IEEE Ultrasonics Symp. (Taipei, Taiwan, 21–24 October 2015)* (https://doi.org/10.1109/ULTSYM.2015.0077)

Jakovljevic M, Pinton G F, Dahl J J and Trahey G E 2017 Blocked elements in 1D and 2D arrays part I: detection and basic compensation on simulated and *in vivo* targets *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **64** 910–21

Jensen J A 1991 A model for the propagation and scattering of ultrasound in tissue *J. Acoust. Soc. Am.* **89** 182–90

Jensen J A 1996 Field: a program for simulating ultrasound systems *Proc. Med. Biol. Eng. Comput.* **34** 351–3

Jensen J A and Svendsen N B 1992 Calculation of pressure fields from arbitrarily shaped, apodized, and excited ultrasound transducers *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **39** 262–7

Johnson D H and Dudgeon D E 1993 *Array Signal Processing: Concepts and Techniques* (Englewood Cliffs, NJ: Prentice-Hall)

Kingma D and Ba J 2015 Adam: a method for stochastic optimization *3rd Int. Conf. for Learning Representations (San Diego, 2015)* (arXiv:1412.6980)

Li P C, Flax S W, Ebbini E S and O'Donnell M 1993 Blocked element compensation in phased array imaging *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **40** 283–92

Li P C and Li M L 2003 Adaptive imaging using the generalized coherence factor *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **50** 128–41

Luchies A and Byram B 2017 Deep neural networks for ultrasound beamforming *Proc. of IEEE Ultrasonics Symp. (Washington, DC, 6–9 September 2017)* (https://doi.org/10.1109/ULTSYM.2017.8092159)

Luchies A C and Byram B C 2018a Deep neural networks for ultrasound beamforming *IEEE Trans. Med. Imag.* **37** 2010–21

Luchies A and Byram B 2018b Suppressing off-axis scattering using deep neural networks *Proc. SPIE* **10580** 105800G

Nikoonahad M and Liv D C 1990 Medical ultrasound imaging using neural networks *IEEE Electron. Lett.* **26** 545–6

Paszke A, Gross S, Chintala S, Chanan G, Yang E, DeVito Z, Lin Z, Desmaison A, Antiga L and Lerer A 2017 Automatic differentiation in PyTorch *NIPS*

Patterson M S and Foster F S 1983 The improvement and quantitative assessment of B-mode images produced by an annular array/cone hybrid *Ultrason. Imaging* **5** 195–213

Perdios D, Besson A, Arditi M and Thiran J P 2017 A deep learning approach to ultrasound image recovery *Proc. of IEEE Ultrasonics Symp. (Washington, DC, 6–9 September 2017)* (https://doi.org/10.1109/ULTSYM.2017.8092746)

Pinton G F, Dahl J, Rosenzweig S and Trahey G E 2009 A heterogeneous nonlinear attenuating full-wave model of ultrasound *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **56** 474–88

Rindal O, Rodriguez-Molares A and Austeng A 2017 The dark region artifact in adaptive ultrasound beamforming *Proc. IEEE Ultrasonics Symp. (Washington, DC, 6–9 September 2017)* (https://doi.org/10.1109/ULTSYM.2017.8092255)

Ruder S 2016 An overview of gradient descent optimization algorithms (arXiv:1609.04747v2)

Smith S W and Wagner R F 1984 Ultrasound speckle size and lesion signal to noise ratio: verification of theory *Ultrason. Imaging* **6** 174–80

Smith S W, Wagner R F, Sandrik J M and Lopez H 1983 Low contrast detectability and contrast/detail analysis in medical ultrasound *IEEE Trans. Sonics Ultrason.* **30** 164–73

Srivastava N, Hinton G E, Krizhevsky A, Sutskever I and Salakhutdinov R 2014 Dropout: a simple way to prevent neural networks from overfitting *J. Mach. Learn. Res.* **15** 1929–58

Vincent P, Larochelle H, Bengio Y and Manzagol P A 2008 Extracting and composing robust features with denoising autoencoders *Proc. ICML* pp 1096–103

Wang S L *et al* 2007 Performance evaluation of coherence-based adaptive imaging using clinical breast data *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **54** 1669–79

Yang B 2008 A study of inverse short-time Fourier transform *IEEE Int. Conf. Acoustics, Speech and Signal Processing* pp 3541–4