Training Deep Network Ultrasound Beamformers With Unlabeled *In Vivo* Data

Jaime Tierney, Adam Luchies, Christopher Khan, *Graduate Student Member, IEEE*, Jennifer Baker, Daniel Brown, Brett Byram, *Member, IEEE*, and Matthew Berger

***Abstract*— Conventional delay-and-sum (DAS) beamforming is highly efficient but also suffers from various sources of image degradation. Several adaptive beamformers have been proposed to address this problem, including more recently proposed deep learning methods. With deep learning, adaptive beamforming is typically framed as a regression problem, where clean ground-truth physical information is used for training. Because it is difficult to know ground truth information *in vivo*, training data are usually simulated. However, deep networks trained on simulations can produce suboptimal *in vivo* image quality because of a domain shift between simulated and *in vivo* data. In this work, we propose a novel domain adaptation (DA) scheme to correct for domain shift by incorporating unlabeled *in vivo* data during training. Unlike classification tasks for which both input domains map to the same target domain, a challenge in our regression-based beamforming scenario is that domain shift exists in both the input and target data. To solve this problem, we leverage cycle-consistent generative adversarial networks to map between simulated and *in vivo* data in both the input and ground truth target domains. Additionally, to account for separate as well as shared features between simulations and *in vivo* data, we use augmented feature mapping to train domain-specific beamformers. Using various types of training data, we explore the limitations and underlying functionality of the proposed DA approach. Additionally, we compare our proposed approach to several other adaptive beamformers. Using the DA DNN beamformer, consistent *in vivo* image quality improvements are achieved compared to established techniques.**

***Index Terms*— Ultrasound, beamforming, domain adaptation, GANs, deep learning.**

Manuscript received August 9, 2021; accepted August 17, 2021. Date of publication August 24, 2021; date of current version December 30, 2021. This work was supported in part by NIH under Grant S10OD016216-01 and Grant R01EB020040, in part by NSF under Grant IIS-175099, and in part by the Data Science Institute at Vanderbilt University. *(Corresponding author: Jaime Tierney.)*

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Vanderbilt University's Institutional Review Board under Application No. 171928, and performed in line with the National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research.

Jaime Tierney, Christopher Khan, and Brett Byram are with the Department of Biomedical Engineering, Vanderbilt University, Nashville, TN 37232 USA (e-mail: jaime.e.tierney@vanderbilt.edu; christopher.m.khan@vanderbilt.edu; brett.c.byram@vanderbilt.edu).

Adam Luchies was with the Department of Biomedical Engineering, Vanderbilt University, Nashville, TN 37232 USA. He is now with Siemens Healthcare, Issaqua, WA 98029 USA (e-mail: adam.c.luchies@vanderbilt.edu).

Jennifer Baker and Daniel Brown are with the Department of Radiology, Vanderbilt University Medical Center, Nashville, TN 37232 USA (e-mail: jennifer.c.baker@vumc.org; daniel.b.brown@vanderbilt.edu).

Matthew Berger is with the Department of Electrical Engineering and Computer Science, Vanderbilt University, Nashville, TN 37235 USA (e-mail: matthew.berger@vanderbilt.edu).

Digital Object Identifier 10.1109/TMI.2021.3107198

I. Introduction

Ultrasonic image formation is accomplished by a process called beamforming. Conventional delay-and-sum (DAS) beamforming is highly efficient but often produces suboptimal ultrasound B-mode image quality, limiting clinical utility. Many sources of image degradation contribute to this problem, including off-axis scattering and reverberation clutter [1].

Several advanced beamforming methods have been proposed to account and correct for image degradation to improve image quality. In contrast to applying fixed delays and weights to received channel data, as is done with conventional DAS, advanced beamforming approaches aim to adaptively enhance signals of interest and suppress sources of image degradation. Among these adaptive approaches are adaptive apodization schemes [2], [3], coherence-based techniques [4], [5], as well as model-based methods [6]–[9]. Although effective, most of these techniques require extensive computational power and/or are limited by user-defined tuning parameters, both of which prevent widespread clinical adoption.

In an effort to perform advanced beamforming more efficiently, several deep learning beamforming approaches have been recently proposed. Deep neural networks (DNNs) are a class of machine learning models that are trained to predict a target output by learning a sequence of nonlinear transformations applied to a given input. These transformations are learned during a training process which uses some variation of gradient descent to minimize the error between the transformed input and desired target data. Although several parameters are evaluated and adjusted during training, once trained, DNNs are intended to be user-independent and highly efficient. They have also been shown to be universal approximators of any continuous function [10]. Therefore, DNNs are very applicable in the context of learning efficient nonlinear regression-based adaptive ultrasound beamformers.

Generally, deep network beamforming techniques can be grouped into three classes. The first class involves training a network to beamform some form of sub-sampled channel data to produce a fully sampled output [11]–[17]. The second involves training a network to mimic an advanced beamformer [18], [19]. Lastly, the third class involves training a network to perform advanced beamforming using physical ground truth information during training [20]–[26]. The first and second classes, although effective, include methods that are hypothetically limited to the fully sampled DAS or advanced beamforming target, while the third includes techniques that could theoretically surpass DAS or advanced beamformer performance.

Although theoretically promising, one of the primary challenges of the third class is obtaining realistic training data. The overall goal of these deep network beamforming approaches is to improve clinically relevant *in vivo* image quality. However, it is arguably impossible to know *in vivo* ground truth information. Therefore, previous efforts have mainly depended on simulations to generate labeled ground truth training data [20], [24], [25]. In addition to known ground truth information, simulations can also be used to produce an unlimited amount of training data. Although generalization to *in vivo* data has been accomplished with simulation-trained networks [20], [24], [25], a shift can still exist between simulated and *in vivo* domains, ultimately limiting performance of deep network beamforming.

To solve this domain shift problem, we propose a novel domain adaptation scheme that leverages unlabeled *in vivo* data to train an *in vivo* beamformer. To do this, we use cycle-consistent generative adversarial networks (CycleGANs) to learn maps between unpaired simulated and *in vivo* data distributions [27]. Other groups have considered GANs for the purposes of ultrasound beamforming [23], [28]. However, to the best of our knowledge, GANs have never been used to train deep network beamformers with real *in vivo* data. Moreover, we expand upon previously proposed domain adaptation schemes by accounting for domain shift in both the noisy inputs and clean outputs. To be clear, in this work, we refer to four different domains: (1) labeled source domain (i.e., simulated input), (2) unlabeled source domain (i.e., *in vivo* input), (3) labeled target domain (i.e., simulated ground truth), and (4) unlabeled target domain (i.e., *in vivo* ground truth). Although irrelevant for classification tasks for which both source domains map to the same target domain [29], in our scenario, a domain shift will still exist between clean simulated target data and clean *in vivo* target data. To account for this, we compose CycleGAN maps with domain-specific regressors to effectively learn deep *in vivo* beamformers.

In this work, compared to our previous preliminary work [30], [31], we perform new, more extensive experiments to better understand the limitations and functionality of the technique. Specifically, in our previous work, we demonstrated initial feasibility using simulated anechoic cysts and *in vivo* liver data from a single healthy subject. In this work, we leverage different types of simulated training data in both the labeled and unlabeled domains, including anechoic and hypoechoic cysts with and without reverberation. The goal of these controlled simulation experiments is to provide a deeper understanding of the different types of domain shift that exist and that can be learned with our proposed approach. Additionally, we report new results on speckle point target phantoms and more clinically variable *in vivo* liver data, including data acquired on both patients with healthy and diseased livers. We compare our approach to conventional DAS, DNNs trained using simulated data only, as well as established coherence, minimum variance, and model-based advanced beamforming techniques.

II. Methods

A. Theory

Our overall goal is to simultaneously learn regressors for beamforming as well as maps that allow us to transform simulated channel data, $(x_s, y_s) \in S$, into corresponding *in vivo* data, $(x_t, y_t) \in T$, and vice versa, where x and y refer to input and ground truth target data, respectively. Our main objective is to learn a function $F_t : \mathbb{R}^d \rightarrow \mathbb{R}^d$ that beamforms *in vivo* data. This is challenging because the *in vivo* ground truth target, y_t, is unknown. Therefore, we aim to approximate (x_t, y_t) pairs that can be used to train F_t.

Previously, deep network beamforming efforts aimed to learn F_t from labeled simulated data (x_s, y_s), where x_s is a time delayed aperture domain signal with clutter and y_s is the known ground truth aperture domain signal without clutter [20]. However, F_t trained with simulations can result in domain mismatch when applied to *in vivo* inputs because the distributions of x_s and x_t differ. We hypothesize that this domain mismatch is due in part to *in vivo* physics not being completely captured in the assumptions made when modeling our simulations. However, even if the simulations precisely capture the physics of wave propagation, we still do not know the exact contributions and distributions of true signal, clutter and noise *in vivo*. Therefore, we start by addressing domain shift in the inputs. To do this, we learn a function $G_{S\rightarrow T}$ that maps a simulated input, x_s, to a corresponding *in vivo* input, x_t. Generative adversarial networks (GANs) [32], specifically for image translation tasks [33], are a commonly used tool for learning maps between different data domains. However, these methods assume labeled data, which is not true in our case because simulated and *in vivo* data do not have corresponding labels. For unlabeled data, Zhu *et al.* [27] proposed the CycleGAN approach which aims to learn maps from S to T, $G_{S\rightarrow T}$, and from T to S, $G_{T\rightarrow S}$, while enforcing cycle-consistency between maps. More concretely, for $G_{S\rightarrow T}$ we formulate the adversarial loss as follows from Zhu *et al.* [27]:

$$L_{G_{S\rightarrow T}}(G_{S\rightarrow T}, D_T) = \mathbb{E}_{x_t \sim X_T}[\log D_T(x_t)] + \mathbb{E}_{x_s \sim X_S}[1 - \log D_T(G_{S\rightarrow T}(x_s))], \tag{1}$$

where X_S and X_T are the simulated and *in vivo* data distributions, respectively, and D_T is a discriminator trained to distinguish real *in vivo* data from *in vivo* data generated by $G_{S\rightarrow T}$. A similar adversarial loss, $L_{G_{T\rightarrow S}}$, can be formulated for $G_{T\rightarrow S}$, which includes a separate discriminator, D_S, tasked with distinguishing real simulated data from simulated data generated by $G_{T\rightarrow S}$. To enforce similarity between real

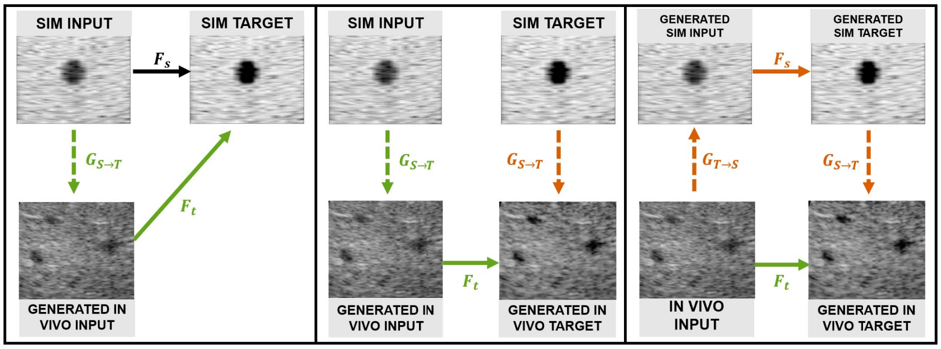

Fig. 1. Schematic of the proposed domain adaptation scheme for training an *in vivo* beamformer, F_t. *In vivo* input and target data generation are indicated by the green and orange arrows, respectively. The left diagram summarizes previous efforts for which target domain adaptation was not considered [29]. In comparison, the middle and right schematics summarize the data used to compute $L_{F_{T1}}$ and $L_{F_{T2}}$, respectively, for the proposed DA approach. The examples in this figure depict fully reconstructed images, but our networks operate on aperture domain signals (i.e., pixel-level channel data).

and generated data, a cycle-consistency regularization is also incorporated [27] as follows,

$$L_{cyc}(G_{S\to T}, G_{T\to S}) = \mathbb{E}_{x_s\sim X_S}[||G_{T\to S}(G_{S\to T}(x_s)) - x_s||_1] + \mathbb{E}_{x_t\sim X_T}[||G_{S\to T}(G_{T\to S}(x_t)) - x_t||_1]. \quad (2)$$

These discriminators and maps can be jointly optimized with F_t, for which paired *in vivo* data are generated via $(G_{S\to T}(x_s), y_s)$. This is at the core of the cycle-consistent adversarial domain adaptation (CyCADA) method [29] and is summarized in Fig. 1(left). CyCADA was proposed for recognition problems, e.g. classification and semantic segmentation, for which both source domains map to the same target domain. However, this is problematic for our scenario because domain shift still exists between y_s and y_t. Therefore, training on $(G_{S\to T}(x_s), y_s)$ necessitates F_t to simultaneously resolve domain gap and beamform.

In contrast to CyCADA [29], we instead want F_t to focus only on beamforming. To accomplish this, we leverage our domain maps, $G_{S\to T}$ and $G_{T\to S}$, in both the input and target domains. To do this, we assume that input domain shift is equivalent to target domain shift. We also introduce a learned function F_s for beamforming simulated data. Incorporating all of this, we arrive at the following *in vivo* beamforming losses:

$$L_{F_S} = \mathbb{E}_{x_s\sim X_S}[||F_s(x_s) - y_s||_l] \quad (3)$$

$$L_{F_{T1}} = \mathbb{E}_{x_s\sim X_S}[||F_t(G_{S\to T}(x_s)) - G_{S\to T}(y_s)||_l], \quad (4)$$

$$L_{F_{T2}} = \mathbb{E}_{x_t\sim X_T}[||F_t(x_t) - G_{S\to T}(F_s(G_{T\to S}(x_t)))||_l]. \quad (5)$$

where l indicates the norm used from Table II. These loss functions ensure that F_t can beamform real *in vivo* data ($L_{F_{T2}}$) as well as *in vivo* data generated from simulated data ($L_{F_{T1}}$). The middle and right panels of Fig. 1 summarize these loss functions. Example fully reconstructed simulated anechoic cyst and *in vivo* images are used for illustrative purposes in Fig. 1. However, our networks operate on aperture domain signals (i.e., pixel-level channel data), as exemplified in Fig. 2A and as described in more detail in Section II-B.

Our full loss is formulated as follows:

$$L = \underbrace{\lambda_{GAN}(\lambda_s L_{G_{S\to T}} + \lambda_t L_{G_{T\to S}} + \lambda_c L_{cyc})}_{\text{GAN}} + \underbrace{\lambda_{REG}(\lambda_{F_S} L_{F_S} + \lambda_{F_T}(L_{F_{T1}} + L_{F_{T2}}))}_{\text{Regressor}}, \quad (6)$$

where discriminators, generators, and regressors are simultaneously optimized for. Overall GAN and regressor weights, λ_{GAN} and λ_{REG}, were set to 1 unless otherwise specified. Individual GAN-related weights were set based on Hoffman *et al.* [29] (i.e., $\lambda_s = 2$, $\lambda_t = 1$,$\lambda_c = 10$), while the individual regressor weights were empirically chosen to be $\lambda_{F_S} = 1$ and $\lambda_{F_T} = 0.5$ such that equal weight is given to the simulated and *in vivo* loss terms (i.e., $\lambda_{F_S} = 2\lambda_{F_T}$). Discriminators were also regularized using the method of Mescheder *et al.* [34].

In addition to accounting for target domain shift, we also train domain-specific beamformers to ensure that the simulation and *in vivo* regressors use separate, as well as shared, features from the two domains. To do this, we learn a single regressor $F : \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$ using the augmented feature mapping method of Daumé [35], such that $F_s(x_s) = F(x_s, x_s, 0)$ and $F_t(x_t) = F(x_t, 0, x_t)$. The first argument captures shared features, while the second and third arguments capture features specific to simulated and *in vivo* data, respectively.

B. Data

Our networks work by performing a regression on time-delayed channel data to adaptively beamform each received spatial location. To generate real and imaginary signal components, a Hilbert transform was applied to all received channel data prior to network processing. Simulated cyst data as well as *in vivo* liver data were used to generate training and test data, as described in more detail below. Additional test data were generated from physical point target speckle phantoms.

1) Simulated Training Data: Several different simulated training data sets were generated for the experiments in this work and are summarized in Table I. In our preliminary work,

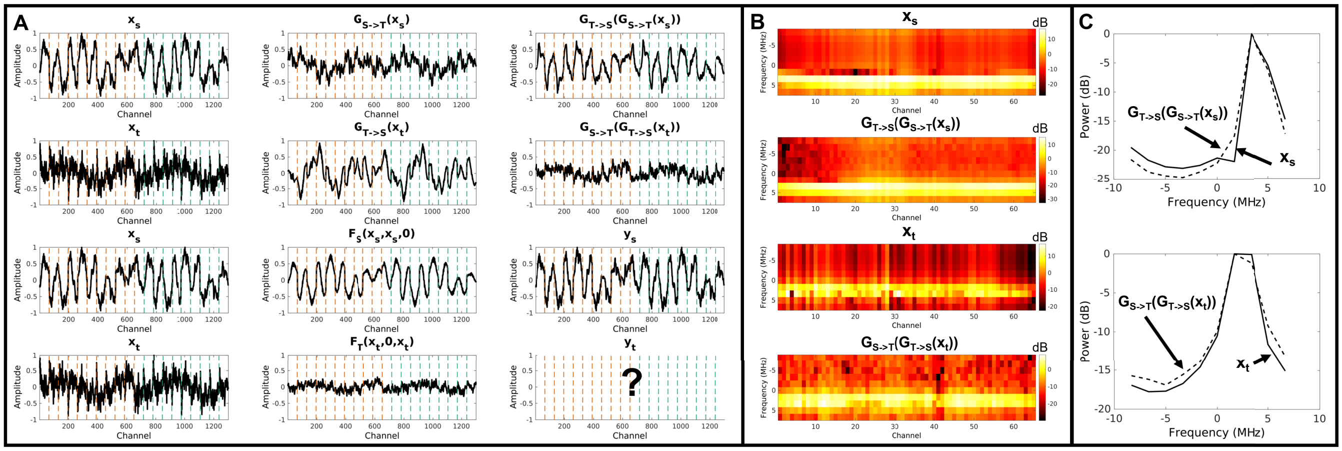

Fig. 2. (A) Example simulated and *in vivo* aperture domain signals (i.e., pixel-level channel data), x_S and x_t, respectively, are shown on the left. The first two rows demonstrate example GAN mappings, $G_{S->T}$ and $G_{T->S}$, applied to the aperture domain signals. The bottom two rows demonstrate the learned regressors, F_S and F_T, on the aperture domain signals, as compared to the labeled simulated and *in vivo* signals shown on the right, y_S and y_t, respectively, for which y_t is unknown. The channel dimension indicates 10 depths of 65 element channel signals for which the real and imaginary components are stacked (i.e., 10 × 65 × 2 reshaped to 1300 × 1). The vertical dashed lines demarcate each 65-element channel signal at a single depth. Real and imaginary components are indicated by the orange and teal lines, respectively. (B) To provide further insight about the GAN mappings, 10-sample Fourier transforms through depth for each channel are displayed for the original, x_S and x_t, and GAN-reconstructed, $G_{T->S}(G_{S->T}(x_S))$ and $G_{S->T}(G_{T->S}(x_t))$, simulated and *in vivo* signals, respectively. (C) Power spectra of the Fourier transforms through depth averaged across channels are displayed for the original (solid), x_S and x_t, and GAN-reconstructed (dashed), $G_{T->S}(G_{S->T}(x_S))$ and $G_{S->T}(G_{T->S}(x_t))$, simulated and *in vivo* signals, respectively.

TABLE I
Training Data Summary

Data Set	Labeled Domain	Unlabeled Domain
Baseline	simulated anechoic cysts ($N = 16,368$)	*in vivo* ($N = 19,620$)
Unlabeled hypoechoic	simulated anechoic cysts ($N = 16,368$)	simulated hypoechoic cysts ($N = 98,208$)
Unlabeled Reverb	simulated anechoic cysts ($N = 16,368$)	simulated anechoic cysts with reverb ($N = 98,208$)
Labeled Hypoechoic	baseline + simulated hypoechoic cysts ($N = 114,576$)	*in vivo* ($N = 19,620$)
Labeled Reverb	baseline + simulated anechoic cysts with reverb ($N = 114,576$)	*in vivo* ($N = 19,620$)
Combo 1	hypoechoic + reverb ($N = 212,784$)	*in vivo* ($N = 19,620$)
Combo 2	simulated hypoechoic cysts with reverb 1:1 ($N = 802,032$)	*in vivo* ($N = 19,620$)
Combo 3	simulated hypoechoic cysts with reverb ($N = 8,364,048$)	*in vivo* ($N = 19,620$)

we used simple anechoic cysts as our labeled domain and *in vivo* data as our unlabeled domain [30]. We therefore refer to this configuration as our baseline training data set. Echogenicity and the exact sources and amount of image degradation *in vivo* are unknown and cannot be controlled for. Instead, we use various combinations of simulated data in both the unlabeled and labeled domains to better understand what the domain adaptation scheme is learning. Here, we describe how we generated the simulated training data. The specific groupings of these data are described in Table I and referenced when describing specific experiments in Section II-C.

For all simulated training data, Field II [36] was used to simulate channel data of 12 10mm diameter cyst realizations centered at the transmit focus and 12 reverberation realizations. All simulations were focused at 60mm using a 4.1667MHz center frequency, 16.667MHz sampling frequency, 1540m/s sound speed, 47 transmit elements ($F/\# = 3$), and 65 active receive element channels with a pitch of 424.6μm. These parameters were used to mimic a Verasonics C5-2 probe sequence used for acquiring the *in vivo* data. For each cyst realization, channel data were simulated separately for the cyst and background scatterers (12 scatterers per resolution cell) to allow for custom scaling. The pseudononlinear approach proposed by Byram and Shu [37] was used for simulating the reverberation realizations, which used 5 scatterers per resolution cell.

The 12 simulated cyst and 12 reverberation realizations were scaled and combined to generate the various training data sets described in Table I. For all simulated hypoechoic cyst data sets, the channel data simulated from the scatterers within the cyst were scaled to achieve 6 true contrast ratios (CR) between 0 and 50dB relative to the channel data simulated from the scatterers outside of the cyst. For all simulated reverberation data sets, reverberation was scaled to achieve 6 signal-to-clutter ratios (SCR) between −5 and 20dB relative to the combined (i.e., summed inside and outside channel data) cyst realization.

For each simulated data set, ground truth target data, y_s, were generated from the clean (i.e., no reverberation) cyst realizations prior to combining the channel data for inside and outside of the cyst as follows,

$$y_s = \begin{cases} k_{CR} x_{inside}, & \text{if inside} \\ x_{outside}, & \text{if outside} \end{cases} \tag{7}$$

where k_{CR} is the scaling term used to achieve the desired CR. For anechoic cysts, $k_{CR} = 0$. Using the aperture domain signals from these separate channel data sets ensures that the true contrast ratio is preserved in our ground truth data. Input data were generated from the fully combined data sets as follows:

$$x_s = k_{CR} x_{inside} + x_{outside} + k_{SCR} x_{reverb} \tag{8}$$

where x_{inside}, $x_{outside}$, and x_{reverb} represent aperture domain signals from channel data simulated from scatterers inside the cyst, outside the cyst, and reverberation, respectively, and k_{CR} and k_{SCR} represent the scaling terms used to generate the desired CR and SCR.

Background and cyst regions were identified depending on whether the aperture signals within a 2λ axial kernel (i.e., 16 depths) originated from a location outside or inside of the cyst, respectively. The center 10 depths of the kernel were used as input and output to the network. Each aperture domain signal was concatenated through depth in addition to concatenating real and imaginary components. Example input and output signals are depicted in Fig. 2A. The number of background and cyst training examples was made equal (i.e., the full background was not used for training). A total of 1,364 paired input/target aperture domain examples were used from each simulated cyst realization.

2) Simulated Test Data: Two simulated test data sets were generated: (1) hypoechoic cysts and (2) anechoic cysts with reverberation. For both test sets, 6 5mm diameter cyst realizations centered at the transmit focus were simulated using the same acquisition parameters described in Section II-B.1. For the hypoechoic test set, for each cyst realization, the channel data simulated from the scatterers within the cyst were scaled to achieve 6 true contrast ratios (CR) between 0 and 50dB relative to the channel data simulated from the scatterers outside of the cyst. For each CR level, a total of $N = 6$ hypoechoic realizations were used for testing. For the reverberation test set, 6 additional reverberation realizations were simulated (i.e., separate from training) and were scaled to achieve 6 signal-to-clutter ratios (SCR) between -5 and 20dB relative to the combined (i.e., summed inside and outside channel data) cyst realizations. Each reverberation realization was paired with a single cyst realization, resulting in $N = 6$ realizations per SCR level. For both test sets, white gaussian noise was added to each test realization to achieve a signal-to-noise ratio of 50dB. The full field of view of each realization was used for testing. A sliding window of 1 depth was used to select 10 depth inputs, and overlapping depth outputs were averaged.

3) In Vivo Data: A Verasonics Vantage Ultrasound System (Verasonics, Inc., Kirkland, WA) and C5-2 curvilinear array transducer were used to acquire channel data of 22 different *in vivo* liver fields of view. Acquisition parameters matched those used for simulations, as described in Section II-B.1. Of the 22 data sets, 13 are of diseased livers belonging to 12 patients (one patient was scanned twice) diagnosed with either hepatocellular carcinoma or neuroendocrine tumors. The other 9 are from the same 37 year old healthy male. All included subjects gave informed written consent in accordance with the local institutional review board.

Of the 22 data sets, 6 (1 healthy and 5 diseased with no repeat patients) were used for training. Similar to what was done for the simulations, aperture domain signals originating from spatial locations within a region around the focus were extracted. A total of 3,270 unlabeled input examples were used from each *in vivo* training data set.

The remaining 16 data sets were split into equally sized validation and test data sets (i.e., $N = 8$ for both). The validation set was used for model selection only. The test set was used for all *in vivo* evaluation. The full *in vivo* field of view was used for testing. A sliding window of 1 depth was used to select 10 depth inputs, and overlapping depth outputs were averaged.

4) Phantom Data: A Verasonics Vantage Ultrasound System (Verasonics, Inc., Kirkland, WA) and C5-2 curvilinear array transducer were used to acquire channel data of 4 different speckle realizations with point targets from a tissue mimicking phantom (CIRS Model 040GSE, Norfolk, Virginia). Acquisition parameters matched those used for simulations, as described in Section II-B.1.

C. Experiments

1) Methodology Evaluation: To demonstrate the importance of the novel aspects of our approach, we evaluate the effectiveness of the DA approach with and without the target domain adaptation and domain-specific regressors. Specifically, for this study, 4 different training schemes were evaluated: (1) the previously proposed CyCADA approach for which target DA is not accounted for [29], (2) CyCADA expansion with target DA but without augmented feature mapping, (3) CyCADA expansion with augmented feature mapping but without target DA, and (4) the proposed DA DNN beamforming approach that combines the target DA and augmented feature mapping. The baseline training data indicated in Table I (i.e., labeled simulated anechoic cyst and unlabeled *in vivo* data) were used to train each DA DNN implementation.

2) Baseline Comparison to Established Beamformers: In our previous preliminary work [30], we demonstrated that our DA DNN approach outperformed other established beamformers. In that work, the DA DNN was trained with *in vivo* data acquired from a single subject with a healthy liver. In this work, we include data from several subjects with varying liver health. We hypothesize that our DA DNN approach can translate to more variable *in vivo* data and produce similar image quality improvements compared to established beamformers.

As a direct baseline comparison to the proposed DA DNN approach, a conventional DNN trained only on simulated data, but with otherwise similar network parameters, was also evaluated. Additionally, an established frequency-domain DNN approach [20] was also evaluated. This method differs from the conventional DNN approach in that it uses short-time Fourier transformed (STFT) data and trains separate networks for individual frequencies. For the data in this work, 3 networks were trained separately for the 3 most prominent frequencies within a 16 sample axial window (i.e., 2λ) of channel data. For this approach, model training and selection were performed as in [20] to highlight a best case scenario.

For the DA DNN beamformer, the baseline data set indicated in Table I (i.e., labeled simulated anechoic cyst and unlabeled *in vivo* data) was used for training, while the conventional DNN and STFT DNN were trained with labeled simulated anechoic cysts only. All beamformers were evaluated on the *in vivo* test data ($N = 8$). Additionally, all beamformers were tested on the speckle phantom data with point targets.

In addition to comparing our approach to other deep learning methods, performance was also evaluated in comparison to other established beamformers, including conventional DAS, the generalized coherence factor (GCF) [4], aperture domain model image reconstruction (ADMIRE) [7], a robust capon minimum variance (MV) beamformer [2], [38], and an eigen-based minimum variance (EIBMV) beamformer [39]. For the GCF approach, as suggested by Li *et al.* [4], a cutoff spatial frequency of 3 frequency bins (i.e., $M_0 = 1$) was used to compute the weighting mask. For the MV approach, tuning parameters were chosen as suggested in Synnevag *et al.* [2]: number of elements used to compute covariance matrix (L) = 32 (i.e., half the number of elements) and diagonal loading constant $= \frac{1}{100L}$. For the EIBMV approach, tuning parameters were chosen based on the standard values described in Heidari *et al.* [40]: temporal averaging window (K) = 8 samples (i.e., pulse length), number of elements used to compute covariance matrix (L) = 32 (i.e., half of the number of elements), diagonal loading constant $= \frac{1}{100L}$, and $\gamma = 0.5$. The EIBMV approach was implemented using code available from the Ultrasound Toolbox [41].

3) Loss Function Regularization Evaluation: To evaluate the effects of varying constraints on the loss function terms described in Eq. 6 with respect to DA DNN performance, two experiments were performed. For the first experiment, the overall regression weight, λ_{REG}, was varied between 0.25 and 1.75 spaced by 0.25 while the overall GAN weight, λ_{GAN}, remained fixed at 1. For the second experiment, the overall regression weight, λ_{REG}, was varied with respect to the overall GAN weight, λ_{GAN}, such that $\lambda_{REG} + \lambda_{GAN} = 2$. The same range of λ_{REG} values was used for both studies. Each experiment involved training 7 additional DA DNN networks for which we ensured that the same random initialization was used for each training run. The hyperparameters indicated in Table II and the baseline training data indicated in Table I were used when training each network. These additional DA DNNs were tested on the *in vivo* data and compared to the other evaluated beamformers.

4) Unlabeled Domain Evaluation: We expect *in vivo* data to contain both hypoechoic cysts and reverberation clutter. Therefore, we hypothesize that the DA DNN beamformer can learn domain shifts between anechoic and hypoechoic data as well as clean and cluttered data. To test this, we performed two unlabeled domain experiments. First, we trained a DA DNN beamformer using labeled simulated anechoic cysts and unlabeled simulated hypoechoic cysts (unlabeled hypoechoic data set in Table I). Additionally, we trained a DA DNN beamformer using labeled simulated anechoic cysts and unlabeled simulated anechoic cysts with reverberation (unlabeled reverb data set in Table I).

TABLE II

Selected DA DNN and Conventional DNN Hyperparameters Are Indicated in Bold and Italicized Font, Respectively

Parameter	Value
GAN Optimization	$\alpha = 0.0002$, $\beta_1 = 0.5$, and $\beta_2 = 0.5$
Regression Optimization	$\alpha = 0.001$, $\beta_1 = 0.9$, and $\beta_2 = 0.999$
Layer Width	*100*,**200**,300,400,500,900,1300
Number of Hidden Layers	*5*,**6**,7
Loss Function	**Smooth L1** or *MSE*
Dropout	0.2
Batch size	561

The DA DNN beamformers trained with these two different domain combinations were compared to conventional DNN beamformers trained with only the labeled simulated anechoic cyst data. Additionally, conventional DNNs were trained using labeled anechoic and hypoechoic cysts (labeled hypoechoic data set in Table I) as well as labeled anechoic cysts with and without reverberation (labeled reverberation data set in Table I). These conventional DNNs trained on the labeled hypoechoic and reverberation data serve as a best case scenario comparison.

Ground truth image quality metrics were computed using the separately beamformed cyst and background simulations. As described in section II-B.2, the channel data and full field of view were used to compute the scaling terms for achieving the desired CR and SCR values when generating the test data. Therefore, because beamformed envelope data and smaller regions of interest (ROIs) were used to compute image quality metrics, the computed ground truth image quality metrics are not equivalent to the CR and SCR values reported in Section II-B.2.

5) Labeled Domain Evaluation: As demonstrated in our previous work [30], the DA DNN beamformer performs well when there exists a substantial domain gap between trivial anechoic cysts and complex *in vivo* data. To determine how the performance changes when the domain gap decreases, we created increasingly complex simulated training data sets to use as the labeled domain. We hypothesize that as the domain shift decreases, domain adaptation becomes less necessary.

These varied training data sets are described in Table I. In addition to the baseline, labeled hypoechoic, and labeled reverberation data sets, 3 additional combinations of these data were also evaluated: (1) combo 1 includes labeled hypoechoic cysts without reverberation and anechoic cysts with reverberation (2) combo 2 includes labeled hypoechoic cysts with reverberation for which each cyst realization is paired with a single reverberation realization (i.e., 1:1), and (3) combo 3 includes labeled hypoechoic cysts with reverberation for which every combination of cyst and reverberation realizations were accounted for.

For all DA DNN beamformers trained with these varying labeled simulation domains, the unlabeled *in vivo* domain remained the same. The conventional DNN beamformer was also evaluated with these different labeled simulation domains. Furthermore, the DA DNN and conventional DNN beamformers for this study were evaluated in comparison to the other established beamformers described in Section II-C.2.

D. Performance Metrics

Contrast-to-noise ratio (CNR), contrast ratio (CR), generalized contrast-to-noise ratio (GCNR), and speckle signal-to-noise ratio (SNRs) were used to evaluate beamformer performance as follows,

$$CNR = 20\log_{10}\frac{|\mu_{background} - \mu_{lesion}|}{\sqrt{\sigma^2_{background} + \sigma^2_{lesion}}} \tag{9}$$

$$CR = -20\log_{10}\frac{\mu_{lesion}}{\mu_{background}} \tag{10}$$

$$GCNR = 1 - \int min\{p_{lesion}(x), p_{background}(x)\}dx \tag{11}$$

$$SNRs = \frac{\mu_{background}}{\sigma^2_{background}} \tag{12}$$

where μ, σ, and p are the mean, standard deviation, and empirical density function [42] of the uncompressed envelope. Prior to computing image quality metrics, envelope data were log compressed, histogram-matched to the corresponding DAS data, and then uncompressed back to envelope data. For the *in vivo* and speckle phantom data, a full histogram matching approach was used. For the simulated data, an ROI-based partial histogram matching approach was used. These approaches were chosen and implemented as described in Bottenus *et al.* [43]. Images were made for qualitative comparison by log compressing the histogram-matched envelope data, scan converting, and scaling to a 60dB dynamic range. For the phantom data, axial and lateral resolution were measured on the upsampled (x10) log compressed envelope data as the -6dB width of the main lobe.

E. Network Details

All networks were trained using Pytorch [44]. A rectified linear unit activation function [45] and Adam optimization [46] were used. All input signals were normalized to a maximum absolute value of 1 prior to network processing. Network weights were initialized using a zero mean Gaussian random variable with variance equal to $\sqrt{\frac{2}{n}}$, where n is the size of the previous layer [47], [48]. All DA DNN and conventional DNN networks were trained to achieve 30,000 training iterations.

DA DNN (including GANs, discriminators, and regressors) and conventional DNN hyperparameters were selected using a small grid search. Hyperparameters corresponding to layer width (e.g. 100-1300), number of hidden layers (e.g. 5-7), and regression losses (e.g. mean squared error, smooth L1) were varied. The input and output to all DA DNN and conventional DNN networks was a 1D vector of aperture domain data, as depicted in Fig. 2A and described in Section II-B. DA DNN and conventional DNN models were trained using the largest labeled simulated training data set (i.e., combo 3 data set in Table I). These models were evaluated on the *in vivo* validation data withheld from training and testing.

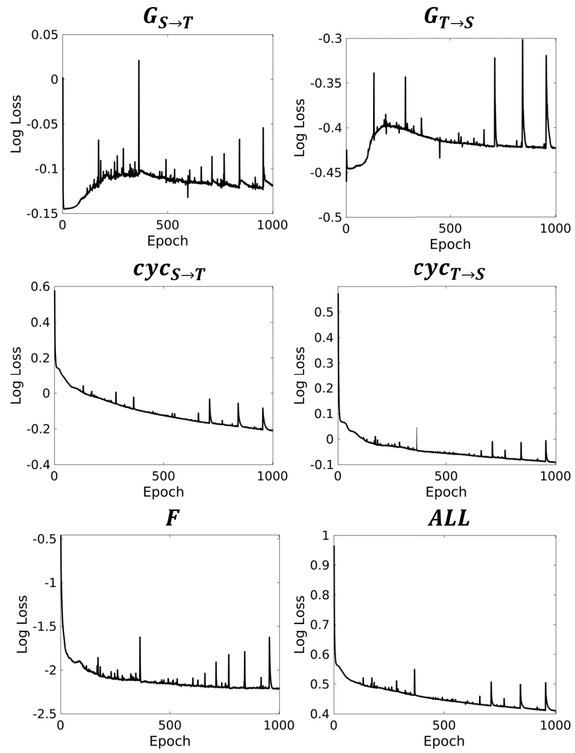

Fig. 3. Log compressed losses for each component of the overall objective function (Eq. 6) are plotted for each training epoch for the DA DNN beamformer trained with the baseline training data set. Each network was trained to achieve 30,000 iterations, which, given $N = 16{,}368$ training examples and a batch size of 561, equates to 1001 epochs.

The model hyperparameters that produced the highest CNR on the validation *in vivo* data were used for all DA DNN and conventional DNN beamformers in this work and are indicated in Table II. GAN and discriminator networks did not use dropout but had otherwise identical hyperparameters to the regressors. Fig. 3 shows the training loss curves for the DA DNN beamformer trained with the baseline training data set indicated in Table I.

Both fully connected [20], [26] and convolutional [23]–[25] architectures have been considered in the context of ultrasound beamforming, and it was demonstrated previously that minimal performance differences exist between the two approaches [49]. To be consistent with the network approach used for comparison in this work [20] as well as the known signal coherence patterns of ultrasound channel data [50], our networks are fully connected across the aperture. Therefore, all networks, including generators, discriminators, and regressors, were fully connected. However, because we apply the same network weights to all depths, our networks are implicitly convolutional through depth.

III. Results

A. Methodology Evaluation

Incorporating both the target domain adaptation and augmented feature mapping into our final training algorithm were necessary for realizing the full potential of our proposed DA DNN beamfomer, as demonstrated in Fig. 4. The example in

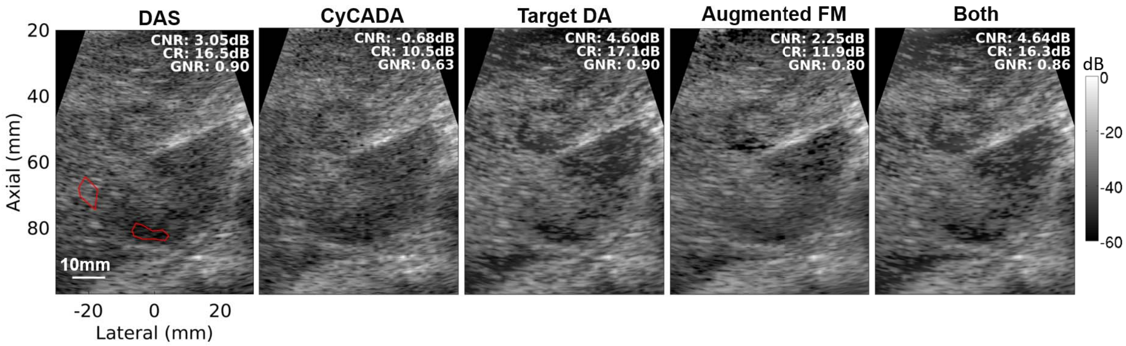

Fig. 4. Example *in vivo* B-mode images are shown to demonstrate the novel and necessary expansions from the previously proposed CyCADA scheme [29]. The example here shows a liver tumor sitting above the kidney. The baseline training data (i.e., labeled simulated anechoic cyst and unlabeled *in vivo* data) were used for all DA DNNs. The regions of interest used to compute image quality metrics for the displayed example are shown in red on the DAS B-mode image. All images are histogram matched to the DAS B-mode image and scaled to individual maximums and a 60dB dynamic range.

TABLE III

AVERAGE CNR, CR, AND GCNR (± STANDARD DEVIATION) ACROSS THE 8 *IN VIVO* TEST EXAMPLES FOR THE DIFFERENT IMPLEMENTATIONS OF DOMAIN ADAPTATION: BASELINE CYCADA, CYCADA WITH TARGET DOMAIN ADAPTATION (DA), CYCADA WITH AUGMENTED FEATURE MAPPING (FM), AND CYCADA WITH BOTH TARGET DOMAIN ADAPTATION AND AUGMENTED FEATURE MAPPING

Method	CNR (dB)	CR (dB)	GCNR
DAS	1.94 (±1.63)	14.5 (±3.94)	0.78 (±0.13)
CyCADA	0.18 (±2.43)	11.3 (±4.24)	0.61 (±0.16)
Target DA	3.89 (±1.28)	19.9 (±5.98)	0.88 (±0.07)
Augmented FM	3.10 (±1.61)	14.1 (±3.78)	0.79 (±0.10)
Both	4.17 (±1.15)	19.7 (±6.09)	0.87 (±0.08)

TABLE IV

AVERAGE CNR, CR, AND GCNR (± STANDARD DEVIATION) ACROSS THE 8 *IN VIVO* TEST EXAMPLES FOR EACH BEAMFORMER EVALUATED. THE BASELINE TRAINING DATA SET INDICATED IN TABLE I WAS USED FOR THE DA DNN APPROACH

Method	CNR (dB)	CR (dB)	GCNR
DAS	1.94 (±1.63)	14.5 (±3.94)	0.78 (±0.13)
GCF	2.24 (±1.48)	19.5 (±5.66)	0.82 (±0.08)
ADMIRE	2.12 (±1.45)	18.0 (±4.39)	0.82 (±0.10)
EIBMV	2.06 (±2.25)	11.6 (±3.74)	0.66 (±0.15)
MV	-0.61 (±2.36)	10.5 (±3.82)	0.60 (±0.17)
STFT DNN	3.12 (±1.41)	18.2 (±5.47)	0.84 (±0.09)
DNN	2.78 (±2.05)	18.3 (±3.68)	0.78 (±0.11)
DA DNN	4.17 (±1.15)	19.7 (±6.09)	0.87 (±0.08)

Fig. 4 shows a liver tumor sitting above the kidney. Using the previously proposed CyCADA approach produces overall worse image quality compared to DAS. Incorporating target domain adaptation visibly improves image quality compared to DAS. Incorporating the augmented feature mapping without target domain adaptation improves image quality compared to the baseline CyCADA approach but does not improve image quality compared to DAS. Incorporating both target domain adaptation and augmented feature mapping produces the most qualitatively compelling image as well as improvements in CNR, CR, and GCNR on average across the full *in vivo* test set compared to conventional DAS, as indicated in Table III.

B. Baseline Comparison to Established Beamformers

Using the baseline training data (i.e., labeled simulated anechoic cysts and unlabled *in vivo* data), DA DNN produced qualitative and quantitative improvements in image quality compared to the other evaluated beamformers, as shown in Fig. 5 and Table IV. The example in the top row of Fig. 5 shows a tumor directly to the left of an anechoic gallbladder. The example in the bottom row of Fig. 5 shows vessels in a healthy liver. The conventional DNN beamformer for both examples produces images with noticeably better contrast than DAS, but they also have more drop out regions compared to the other beamformers, resulting in lower CNR. For the example in the top row of Fig. 5, apart from preserving the speckle background, the image made using the DA DNN beamformer shows the clearest delineation of the tumor boundary in the near field. These overall trends are described quantitatively in Table IV, for which DA DNN produced the highest average CNR and GCNR overall while still maintaining higher CR than DAS.

All beamformers resulted in improved speckle SNR compared to DAS when computed on the physical phantom data. Additionally, all beamformers except for ADMIRE resulted in improved lateral resolution. Apart from ADMIRE and EIBMV, all of the beamformers resulted in slightly worse axial resolution compared to DAS. These results are supported qualitatively in Fig. 6 and quantitatively in Table V.

C. Loss Function Regularization Evaluation

Fig. 7 demonstrates that the DA DNN approach is generally robust to varying amounts of regularization on the overall objective function. When varying just the regressor weights (left column in Fig. 7), larger weights result in overall better performance compared to the baseline value of $\lambda_{REG} = 1$. However, even with smaller weights ($\lambda_{REG} < 1$), the DA DNN still performs well relative to the other beamformers. When varying the regressor weights with respect to the GAN weights (right column in Fig. 7), performance begins to

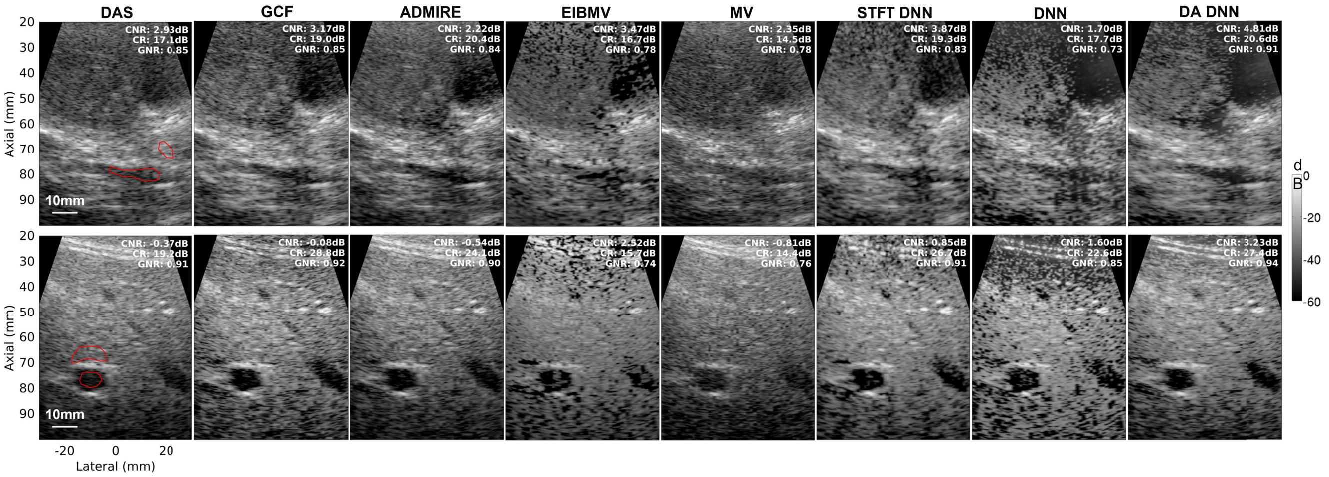

Fig. 5. Example *in vivo* B-mode images are shown for each beamformer. The top example shows a liver tumor to the left of the anechoic gallbladder. The bottom example shows vessels in a healthy liver. The baseline training data (i.e., labeled simulated anechoic cyst and unlabeled *in vivo* data) were used to train the DA DNN. The regions of interest used to compute image quality metrics for the displayed example are shown in red on the DAS B-mode image. All images are histogram matched to the DAS B-mode image and scaled to individual maximums and a 60dB dynamic range.

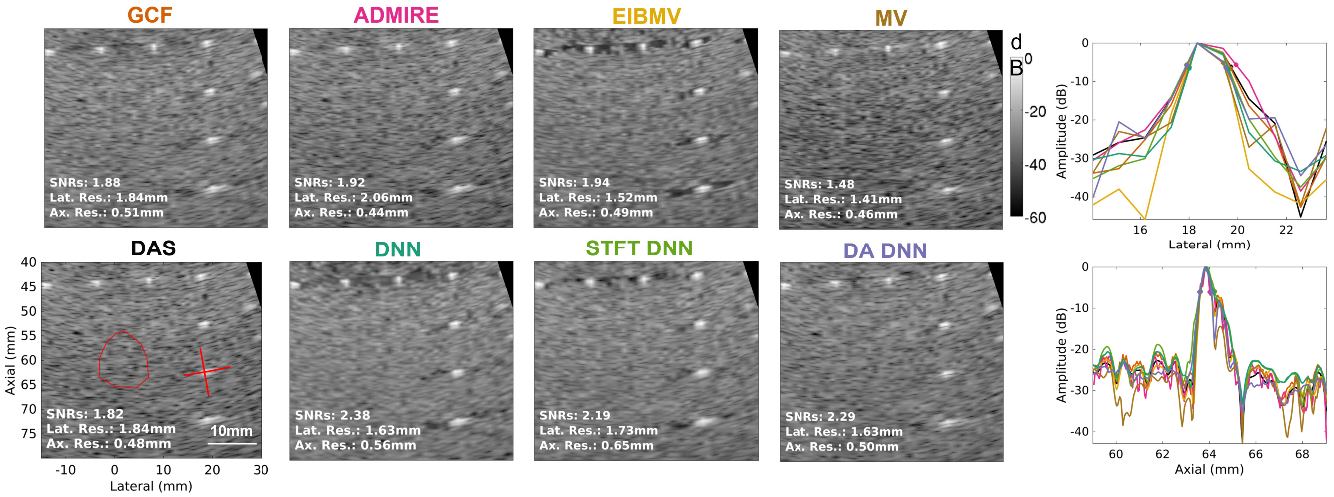

Fig. 6. Example physical phantom B-mode images are shown for each beamformer. The baseline training data (i.e., labeled simulated anechoic cyst and unlabeled *in vivo* data) were used to train the DA DNN. The region of interest used to compute SNRs for the displayed example is shown in red on the DAS B-mode image. The lateral and axial plots corresponding to the red lines displayed in the DAS image were used to compute resolution for each case. The red lines are slanted in the DAS image because resolution was computed on the log compressed envelope data prior to scan conversion. All images are histogram matched to the DAS B-mode image and scaled to individual maximums and a 60dB dynamic range.

TABLE V
AVERAGE SNRs, LATERAL AND AXIAL RESOLUTION (± STANDARD DEVIATION) ACROSS THE 4 PHANTOM TEST EXAMPLES FOR EACH BEAMFORMER EVALUATED. THE BASELINE TRAINING DATA SET INDICATED IN TABLE I WAS USED FOR THE DA DNN APPROACH

Method	SNRs	Lateral (mm)	Axial (mm)
DAS	1.85 (±0.08)	1.71 (±0.10)	0.47 (±0.01)
GCF	1.94 (±0.04)	1.65 (±0.18)	0.51 (±0.03)
ADMIRE	1.96 (±0.07)	1.92 (±0.41)	0.37 (±0.08)
EIBMV	2.12 (±0.13)	1.25 (±0.22)	0.46 (±0.04)
MV	1.63 (±0.11)	1.14 (±0.29)	0.47 (±0.04)
STFT DNN	2.17 (±0.06)	1.68 (±0.11)	0.78 (±0.11)
DNN	2.23 (±0.09)	1.68 (±0.28)	0.57 (±0.06)
DA DNN	2.27 (±0.02)	1.63 (±0.26)	0.50 (±0.04)

degrade when $\lambda_{GAN} < 1$ and $\lambda_{REG} > 1$, indicating that the GAN regularization is more influential than the regularization for the regression. Overall, these results confirm that DA DNN performance is relatively stable when considering extra constraints in the loss function.

D. Unlabeled Domain Evaluation

The proposed domain adaptation scheme was able to successfully account for domain shift between anechoic and hypoechoic cysts. As shown on the right of Fig. 8, the DA DNN beamformer produced an image that has an estimated contrast ratio of 26.7dB, which, compared to the other beamformers, is closest to the true contrast ratio of 26dB. The conventional DNN approach overestimates contrast ratio overall, which is consistent with what we see *in vivo*, as demonstrated in Fig. 5. As hypothesized, the DNN trained with ground truth hypoechoic cysts performs best across the full contrast ratio test set (i.e., the pink line stays closest to the true contrast line overall), but the DA DNN is substantially closer to the

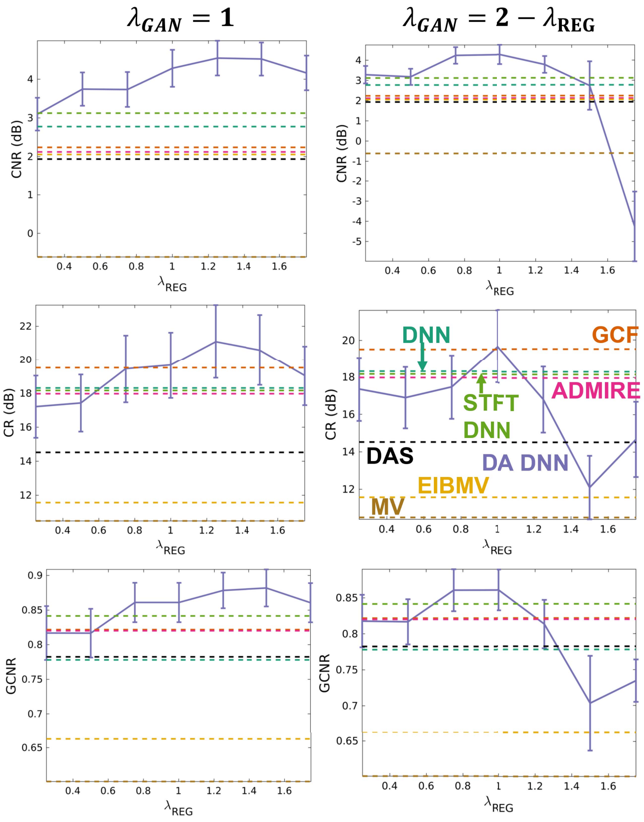

Fig. 7. Average (± standard error) DA DNN CNR (top), CR (middle), and GCNR (bottom) as a function of λ_{REG} from Eq. 6 is shown in purple in each plot. The left column shows results for when λ_{GAN} is fixed at 1 while the right shows results for when λ_{GAN} varies as a function of λ_{REG} such that $\lambda_{GAN} = 2 - \lambda_{REG}$. Average values for all other beamformers are indicated by the dashed horizontal lines in each plot.

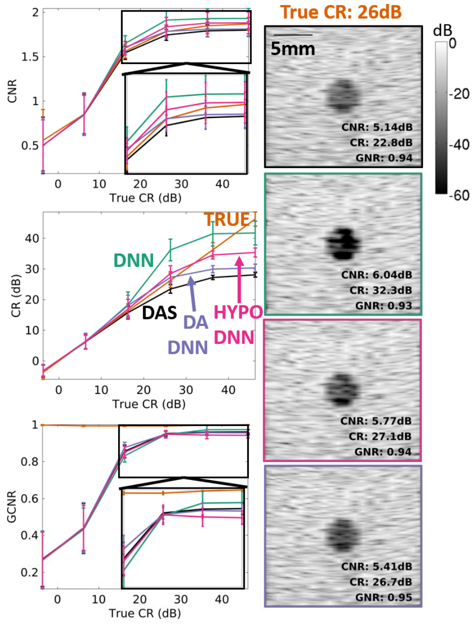

Fig. 8. Average CNR, CR, and GCNR (± standard deviation) computed on simulated hypoechoic cyts are shown for DAS (black), DNN trained with labeled simulated anechoic cysts (teal), DNN trained with labeled hypoechoic cysts (pink), and DA DNN trained with labeled simulated anechoic cysts and unlabeled hypoechoic cysts (purple). The CNR plot displays the uncompressed values for better visualization of differences between curves. Additionally, the CNR and GCNR plots show zoomed in versions of the curves between 20 and 50dB true CR values. Example 5mm diameter simulated anechoic cyst B-mode images are shown on the right with a true CR of 26dB. Each image is outlined in corresponding colors from the plots. All images are histogram matched to the DAS B-mode image and scaled to individual maximums and a 60dB dynamic range.

true contrast compared to the conventional DNN and DAS approaches, as shown in the plots on the left of Fig. 8.

The domain adaptation approach is also able to account for domain shift between anechoic cysts without and with reverberation that would otherwise cause degraded speckle. The qualitative example shown on the right of Fig. 9 shows how the DA DNN approach is able to preserve the speckle background better than the conventional DNN. This results in CNR that is closer to the true CNR with the DA DNN approach for all SCR levels, as seen in the plots on the left of Fig. 9. The DNN trained with ground truth simulated anechoic cysts with reverberation provides the best clutter suppression overall, as expected.

E. Labeled Domain Evaluation

Using more complex, *in vivo*-like simulations for training resulted in consistent image quality improvements for the conventional DNN approach. This conclusion is supported qualitatively in the top row of Fig. 11 for which the DNN trained with the labeled combo 3 simulated data produced the best image compared to the DNNs trained with the other evaluated labeled data sets. Quantitatively, Fig. 10 shows how DNN CNR improves consistently as the labeled domain complexity increases. In contrast, the DA DNN approach produces qualitatively similar images across the different labeled simulated training data sets, as shown in the bottom row of Fig. 11. This observation is supported quantitatively in Fig. 10 for which the DA DNN CNR, CR, and GCNR remain fairly constant among the different labeled training data sets. Although the DA DNN approach does not produce the highest CR overall, it produces consistently higher CR compared to DAS and also consistently produces the highest CNR and GCNR across all of the evaluated beamformers.

IV. Discussion

Domain adaptation for ultrasound beamforming required two main contributions from previous domain adaptation efforts [29]: (1) accounting for target domain shift and (2) learning distinct regressors for simulations and *in vivo* data using augmented feature mapping. The importance of these contributions are qualitatively and quantitatively supported in Fig. 4 and Table III. An important fundamental assumption of this approach is that the domain shift between simulated and *in vivo* data is the same for the inputs and the outputs. Based on our results, this seems to be a reasonable baseline assumption. However, unlike the target domain adaptation, the input domain adaptation needs to account for sources of clutter, a discrepancy that could potentially invalidate our assumption.

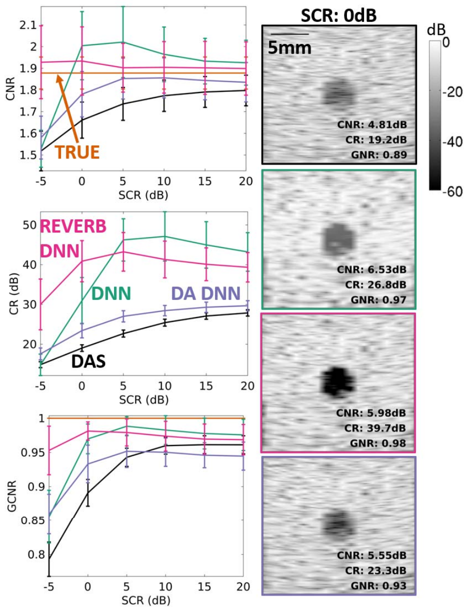

Fig. 9. Average CNR, CR, and GCNR (± standard deviation) computed on simulated anechoic cysts with varied levels of reverberation are shown for DAS (black), DNN trained with labeled anechoic cysts (teal), DNN trained with labeled anechoic cysts with reverberation (pink), and DA DNN trained with labeled simulated anechoic cysts and unlabeled simulated anechoic cysts with reverberation (purple). The CNR plot displays the uncompressed values for better visualization of differences between curves. Example 5mm diameter simulated anechoic cyst B-mode images are shown on the right for a signal-to-clutter ratio (SCR) of 0dB. Each image is outlined in corresponding colors from the plots. All images are histogram matched to the DAS B-mode image and scaled to individual maximums and a 60dB dynamic range.

In our initial work [30], we developed and tested our DA DNN approach using data acquired with a linear array and higher center frequency than that used with the curvilinear array in this work. Additionally, in our initial work, we used *in vivo* data from a single healthy subject for training and testing, whereas in this work, we include data from multiple subjects with varying liver health. In other words, the *in vivo* data set used in this work is substantially more variable than that used in our original work. Despite these differences, we observe similar results and trends with respect to the DA DNN and how it compares to other established beamformers. This is noteworthy because it suggests that the technique is robust and reproducible for different acquisition sequences as well as different data types. That said, it is worth investigating further the extent of *in vivo* data variability, including, for example, *in vivo* data acquired with different scanners.

The physical phantom results confirm that the proposed DA DNN beamformer improves or maintains speckle SNR and resolution in phantoms for which we know what to expect. However, because the DA DNN is trained with *in vivo* data to account for domain shift between ground truth simulations and *in vivo* data, it is optimized to perform best on *in vivo* data. Therefore, we do not expect the DA DNN beamformer to produce substantial improvements on physical phantom data. We could train with unlabeled physical phantom data instead of *in vivo* data to account for domain shift between simulations and physical phantom data, but this has not been an apparent issue, as demonstrated in Fig. 6 and Table V for which both the conventional DNN and STFT DNN also resulted in improvements in speckle SNR and resolution compared to DAS. In other words, DNN beamformers trained with ground truth simulations tend to generalize well to physical phantom data but not always to the *in vivo* data that we care about. We could also train with labeled physical phantom data instead of (or in addition to) simulated data to account for domain shift between physical phantoms and *in vivo* data. However, given the similar performance on simulations and phantoms, we think that in addressing the domain gap between simulations and *in vivo* data we are also accounting for a lot of the domain gap potentially observed between physical phantoms and *in vivo* data. Additionally, we do not definitively know ground truth information in physical phantoms. We can approximate cyst and speckle regions of interest, but these will not be as precise as simulations. Furthermore, even if we had accurate regions of interest, we believe the domain gap between simulations or phantom data and *in vivo* data is primarily a result of the presence and unique distributions of various sources of clutter. Therefore, even if we were able to accurately synthesize and control for different types of clutter in our physical phantom data, it is still impossible to synthesize the exact contributions and combinations of each type of clutter that are found *in vivo*. Our approach aims to solve this problem, which is equally applicable to simulations and phantom data

By using controlled simulations for the unlabeled domain studies, the domain shift was known and allowed for better understanding of what the DA DNN beamformer, including the regressor and GAN mappings, is actually learning. Although Fig. 2 demonstrates that the approach can learn to map between clean simulations and noisy *in vivo* aperture domain signals, it is impossible to know exactly what types of noise are being accounted for and what the ground truth regressed output should be for the *in vivo* data. The unlabeled domain study using controlled simulations provided meaningful insight to these otherwise unknown questions. Although these studies were informative, a potential limitation of the reverberation study is that the labeled simulated anechoic cyst and unlabeled anechoic cyst with reverberation source domains both map to the same target domain (i.e., clean anechoic cysts). It is possible that the original CyCADA approach with augmented feature mapping would perform better in this scenario. Augmented feature mapping would still be beneficial to differentiate between beamforming that requires reverberation suppression or not. In other words, the augmented feature mapping would ensure that the regressor itself was still invariant to domain shift. In contrast, the original CyCADA approach aims to address domain shift only in the input domain data without also accounting for domain shift in the regressor itself.

The DA DNN approach was robust across varying and increasingly more *in vivo*-like labeled simulation domains. This is noteworthy because it suggests that simple labeled anechoic cysts are sufficient for training a successful DA DNN *in vivo* beamformer. Additionally, this result suggests that the DA DNN approach is less prone to overfitting to the simulated

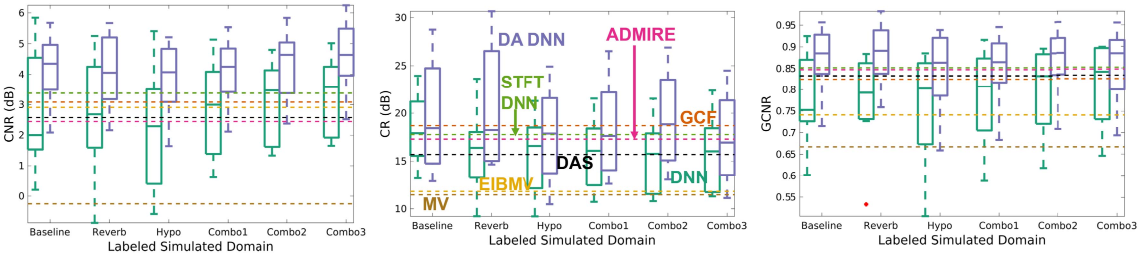

Fig. 10. Median CNR (left), CR (middle) and GCNR (right) across the 8 *in vivo* test examples for DNN (teal) and DA DNN (purple) trained with different labeled simulated data sets. The median value for each method is the central mark in each box. The 25th and 75th percentiles are the bottom and top edges of each box, respectively. The bars extending from each box indicate the minimums and maximums, and outliers are marked in red. Reference CNR, CR, and GCNR median values are shown as the dashed curves for DAS (black), GCF (orange), ADMIRE (pink), and STFT DNN (green).

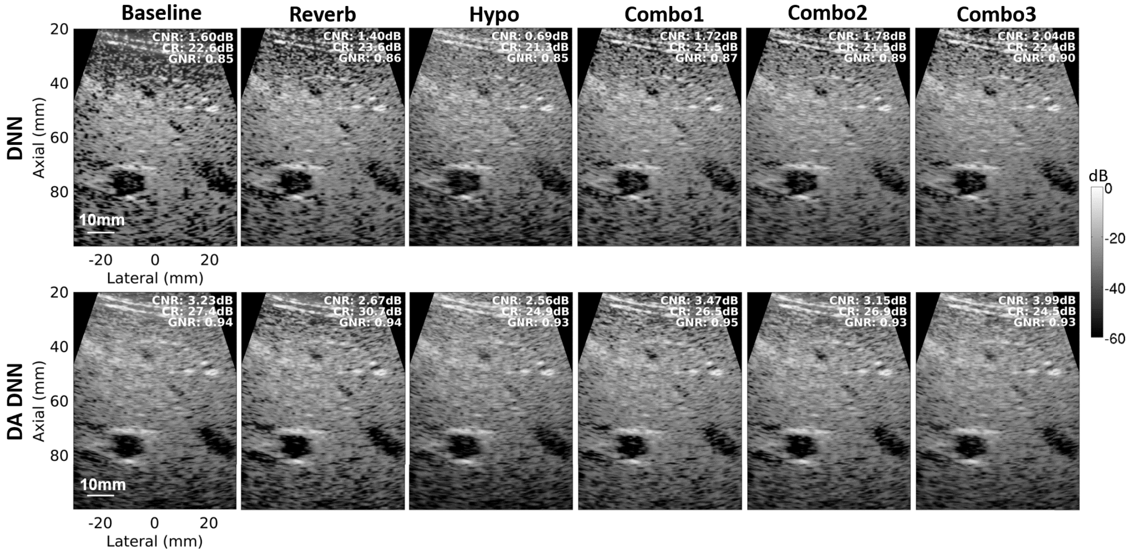

Fig. 11. Example *in vivo* B-mode images are shown for each labeled simulated training data set evaluated for DNN and DA DNN. The example here shows vessels in a healthy liver. The regions of interest used to compute image quality metrics for the displayed example are shown in red on the DAS B-mode image in the bottom row of Fig. 5. All images are histogram matched to the corresponding DAS B-mode image and scaled to individual maximums and a 60dB dynamic range.

data compared to the conventional DNN approach, which did not generalize as well to *in vivo* data when trained with the least *in vivo*-like simulations (i.e., simple anechoic cysts). Moreover, the results of the labeled domain evaluation suggest that the magnitude of the domain shift (i.e., combo 3 labeled simulations to unlabeled *in vivo* being the smallest domain shift evaluated) does not affect the performance of the DA DNN. This is also noteworthy because it suggests that our most complex simulations are potentially still not capturing the full ultrasound physics that occur *in vivo*. It is worth investigating if more sophisticated phase aberration or full wave simulations decrease the domain shift between simulations and *in vivo* data. However, all simulation models make assumptions that are not always met *in vivo*, which motivates the persistent need for domain adaptation when performing *in vivo* DNN beamforming.

Despite the evident robustness of the DA DNN approach to varying labeled simulation domains, there are also subtle differences between the DA DNN results that could indicate *in vivo* characteristics. For example, the DA DNN beamformer trained with labeled simulated anechoic cysts with reverberation produces higher median contrast overall and darker vessel regions in the example in Fig. 11 compared to the DA DNN trained with labeled hypoechoic cysts. This result might suggest that reverberation contributes more to domain shift between the simulated and *in vivo* data evaluated in this work. However, it could also mean that the regions of interest used to compute quantitative metrics are not all anechoic in a ground truth sense and therefore the contrast is over estimated with the labeled reverberation case and more accurate with the hypoechoic case. It is also worth noting that we did not incorporate hyperechoic cysts in the simulated training data which could potentially introduce a bias towards anechoic and hypoechoic structures. Therefore, although the observed differences are subtle, they suggest the potential benefit of combining different types of labeled simulated data to produce more realistic representations of *in vivo* data.

All simulated cysts used for training in this work were generated to be centered about the transmit focus. Previous work [20] demonstrated that this will cause a limited depth of field for the STFT DNN, as seen by the degraded speckle in the shallow depths of the STFT DNN images in Fig. 5. This trend was also prevalent for the conventional DNN approach. However, the DA DNN approach seems to increase the depth of field compared to the STFT and conventional DNN approaches, despite also only being trained with aperture domain examples that originated from within the focal zone. It possible to improve the depth of field further by including training examples from a larger spatial range or by training separate networks for shallow depths.

In addition to the demonstrated image quality improvements compared to the other evaluated beamformers, once trained, the DA DNN approach is extremely efficient, especially in comparison to the non-DNN advanced beamformers. Additionally, although the proposed DA DNN approach involves a complex training scheme in terms of the number of models being trained (2 generators, 2 discriminators, and 1 regressor), because all of these models are trained simultaneously, only a single training run is required to train a DA DNN beamformer. Similarly, a single network is used at test time. This is in contrast to the STFT DNN beamformer used in this work which requires 3 separate networks to train and test a single STFT DNN beamformer (i.e., a separate network for each frequency). Additionally, because the DA DNN approach operates on time domain data, no immediate pre- or post-network processing is required. In contrast, the STFT approach requires both an STFT and inverse STFT during both training and testing.

Although our proposed implementation is meant to be used for ultrasound beamforming in the aperture domain, we hypothesize that it is also broadly applicable to other regression-based tasks for which labeled *in vivo* training data are lacking but some form of other relevant labeled data is available. For example, the proposed framework could be applied to other deep learning beamforming efforts that similarly use simulations to obtain physical ground truth training data [24], [25], albeit with different architectures and overall training objectives. Additionally, it is plausible that the proposed approach could be used to perform a form of image processing on DAS images to improve image quality. Although image processing can improve image quality, removing sources of image degradation like off-axis scattering and reverberation is challenging without the channel information, which is why we focus on beamforming in the present work. Finally, although these examples are specific to ultrasound, we believe that the overall framework could also be applied to other imaging modalities that aim to perform reconstruction or processing tasks for which obtaining ground truth *in vivo* data is challenging or impossible.

V. Conclusion

Conventional deep learning adaptive beamforming techniques rely on ground truth training data which is arguably impossible to obtain *in vivo*. To solve this problem, we developed a novel domain adaptation scheme to incorporate unlabeled *in vivo* data during training. We show that the proposed DA DNN beamforming is robust in the presence of several different types of domain shift. Additionally, we compared our approach to conventional DNN beamforming and to other established beamformers, including DAS, GCF, ADMIRE, MV, and STFT DNNs, and we demonstrated consistent image quality improvements with the DA DNN beamformer. Notably, we show that our approach can achieve image quality consistent with or higher than state-of-the-art ADMIRE and STFT DNN beamforming without the same computational limitations.

References

[1] J. J. Dahl and N. M. Sheth, “Reverberation clutter from subcutaneous tissue layers: Simulation and *in vivo* demonstrations,” *Ultrasound Med. Biol.*, vol. 40, no. 4, pp. 714–726, Apr. 2014.

[2] J.-F. Synnevåg, A. Austeng, and S. Holm, “Adaptive beamforming applied to medical ultrasound imaging,” *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 54, no. 8, pp. 1606–1613, Aug. 2007.

[3] I. K. Holfort, F. Gran, and J. A. Jensen, “Broadband minimum variance beamforming for ultrasound imaging,” *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 56, no. 2, pp. 314–325, Feb. 2009.

[4] P.-C. Li and M.-L. Li, “Adaptive imaging using the generalized coherence factor,” *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 50, no. 2, pp. 128–141, Feb. 2003.

[5] M. A. Lediju, G. E. Trahey, B. C. Byram, and J. J. Dahl, “Short-lag spatial coherence of backscattered echoes: Imaging characteristics,” *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 58, no. 7, pp. 1377–1388, Jul. 2011.

[6] B. Byram and M. Jakovljevic, “Ultrasonic multipath and beamforming clutter reduction: A chirp model approach,” *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 61, no. 3, pp. 428–440, Mar. 2014.

[7] B. Byram, K. Dei, J. Tierney, and D. Dumont, “A model and regularization scheme for ultrasonic beamforming clutter reduction,” *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 62, no. 11, pp. 1913–1927, Nov. 2015.

[8] K. Dei and B. C. Byram, “The impact of model-based clutter suppression on cluttered, aberrated wavefronts,” *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 64, no. 10, pp. 1450–1464, Oct. 2017.

[9] K. Dei and B. Byram, “A robust method for ultrasound beamforming in the presence of off-axis clutter and sound speed variation,” *Ultrasonics*, vol. 89, pp. 34–45, Sep. 2018.

[10] K. Hornik, M. Stinchcombe, and H. White, “Multilayer feedforward networks are universal approximators,” *Neural Netw.*, vol. 2, no. 5, pp. 359–366, 1989.

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

[12] M. Gasse, F. Millioz, E. Roux, D. Garcia, H. Liebgott, and D. Friboulet, “High-quality plane wave compounding using convolutional neural networks,” *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 64, no. 10, pp. 1637–1639, Oct. 2017.

[13] 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 Trans. Med. Imag.*, vol. 38, no. 2, pp. 325–336, Feb. 2018.

[14] O. Senouf *et al.*, “High frame-rate cardiac ultrasound imaging with deep learning,” in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2018, pp. 126–134.

[15] 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 Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 65, no. 11, pp. 1983–1996, Nov. 2018.

[16] S. Khan, J. Huh, and J. C. Ye, “Universal deep beamformer for variable rate ultrasound imaging,” 2019, *arXiv:1901.01706*. [Online]. Available: http://arxiv.org/abs/1901.01706

[17] S. Khan, J. Huh, and J. C. Ye, “Adaptive and compressive beamforming using deep learning for medical ultrasound,” *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 67, no. 8, pp. 1558–1572, Aug. 2020.

[18] W. Simson *et al.*, “End-to-end learning-based ultrasound reconstruction,” 2019, *arXiv:1904.04696*. [Online]. Available: http://arxiv.org/abs/1904.04696

[19] B. Luijten *et al.*, “Adaptive ultrasound beamforming using deep learning,” *IEEE Trans. Med. Imag.*, vol. 39, no. 12, pp. 3967–3978, Dec. 2020.

[20] A. C. Luchies and B. C. Byram, "Deep neural networks for ultrasound beamforming," *IEEE Trans. Med. Imag.*, vol. 37, no. 9, pp. 2010–2021, Sep. 2018.

[21] A. C. Luchies and B. C. Byram, "Training improvements for ultrasound beamforming with deep neural networks," *Phys. Med. Biol.*, vol. 64, no. 4, Feb. 2019, Art. no. 045018.

[22] A. C. Luchies and B. C. Byram, "Assessing the robustness of frequency-domain ultrasound beamforming using deep neural networks," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 67, no. 11, pp. 2321–2335, Nov. 2020.

[23] A. A. Nair, T. D. Tran, A. Reiter, and M. A. L. Bell, "A generative adversarial neural network for beamforming ultrasound images: Invited presentation," in *Proc. 53rd Annu. Conf. Inf. Sci. Syst. (CISS)*, Mar. 2019, pp. 1–6.

[24] A. A. Nair, K. N. Washington, T. D. Tran, A. Reiter, and M. A. L. Bell, "Deep learning to obtain simultaneous image and segmentation outputs from a single input of raw ultrasound channel data," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 67, no. 12, pp. 2493–2509, Dec. 2020.

[25] D. Hyun, L. L. Brickson, K. T. Looby, and J. J. Dahl, "Beamforming and speckle reduction using neural networks," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 66, no. 5, pp. 898–910, May 2019.

[26] R. Zhuang and J. Chen, "Deep learning based minimum variance beamforming for ultrasound imaging," in *Smart Ultrasound Imaging and Perinatal, Preterm and Paediatric Image Analysis*. Cham, Switzerland: Springer, 2019, pp. 83–91.

[27] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2223–2232.

[28] O. Huang *et al.*, "MimickNet, mimicking clinical image post-processing under black-box constraints," *IEEE Trans. Med. Imag.*, vol. 39, no. 6, pp. 2277–2286, Jun. 2020.

[29] J. Hoffman *et al.*, "CyCADA: Cycle-consistent adversarial domain adaptation," 2017, *arXiv:1711.03213*. [Online]. Available: http://arxiv.org/abs/1711.03213

[30] J. Tierney, A. Luchies, C. Khan, B. Byram, and M. Berger, "Domain adaptation for ultrasound beamforming," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2020, pp. 410–420.

[31] J. Tierney, A. Luchies, C. Khan, B. Byram, and M. Berger, "Accounting for domain shift in neural network ultrasound beamforming," in *Proc. IEEE Int. Ultrason. Symp. (IUS)*, Sep. 2020, pp. 1–3.

[32] I. Goodfellow *et al.*, "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 2672–2680.

[33] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1125–1134.

[34] L. Mescheder, A. Geiger, and S. Nowozin, "Which training methods for GANs do actually converge?" 2018, *arXiv:1801.04406*. [Online]. Available: http://arxiv.org/abs/1801.04406

[35] T. Tommasi and B. Caputo, "Frustratingly easy NBNN domain adaptation," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 897–904.

[36] J. A. Jensen, "Field: A program for simulating ultrasound systems," in *Proc. 10th Nordicbaltic Conf. Biomed. Imag.*, vol. 4, 1996, pp. 351–353.

[37] B. Byram and J. Shu, "Pseudononlinear ultrasound simulation approach for reverberation clutter," *J. Med. Imag.*, vol. 3, no. 4, 2016, Art. no. 046005.

[38] J.-F. Synnevåg, A. Austeng, and S. Holm, "Benefits of minimum-variance beamforming in medical ultrasound imaging," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 56, no. 9, pp. 1868–1879, Sep. 2009.

[39] B. M. Asl and A. Mahloojifar, "Eigenspace-based minimum variance beamforming applied to medical ultrasound imaging," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 57, no. 11, pp. 2381–2390, Nov. 2010.

[40] M. H. Heidari, M. Mozaffarzadeh, R. Manwar, and M. Nasiriavanaki, "Effects of important parameters variations on computing eigenspace-based minimum variance weights for ultrasound tissue harmonic imaging," *Photons Plus Ultrasound, Imag. Sens.*, vol. 10494, Feb. 2018, Art. no. 104946R.

[41] A. Rodriguez-Molares, O. M. H. Rindal, O. Bernard, H. Liebgott, A. Austeng, and L. Lovstakken, "The ultrasound toolbox," in *Proc. IEEE Int. Ultrason. Symp. (IUS)*, Sep. 2017, pp. 1–4.

[42] A. Rodriguez-Molares *et al.*, "The generalized contrast-to-noise ratio: A formal definition for lesion detectability," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 67, no. 4, pp. 745–759, Apr. 2020.

[43] N. Bottenus, B. C. Byram, and D. Hyun, "Histogram matching for visual ultrasound image comparison," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 68, no. 5, pp. 1487–1495, May 2021.

[44] A. Paszke *et al.*, "Automatic differentiation in PyTorch," in *Proc. NIPS*, 2017, pp. 1–4.

[45] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks," in *Proc. 14th Int. Conf. Artif. Intell. Statist.*, 2011, pp. 315–323.

[46] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*. [Online]. Available: http://arxiv.org/abs/1412.6980

[47] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proc. 13th Int. Conf. Artif. Intell. Statist.*, 2010, pp. 249–256.

[48] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1026–1034.

[49] Z. Chen, A. Luchies, and B. Byram, "Compact convolutional neural networks for ultrasound beamforming," in *Proc. IEEE Int. Ultrason. Symp. (IUS)*, Oct. 2019, pp. 560–562.

[50] R. Mallart and M. Fink, "The van Cittert–Zernike theorem in pulse echo measurements," *J. Acoust. Soc. Amer.*, vol. 90, no. 5, pp. 2718–2727, 1991.