



Domain Adaptation for Ultrasound Beamforming

Jaime Tierney^(✉), Adam Luchies, Christopher Khan, Brett Byram,
and Matthew Berger

Vanderbilt University, Nashville, TN 37235, USA
jaime.e.tierney@vanderbilt.edu

Abstract. Ultrasound B-Mode images are created from data obtained from each element in the transducer array in a process called beamforming. The beamforming goal is to enhance signals from specified spatial locations, while reducing signal from all other locations. On clinical systems, beamforming is accomplished with the delay-and-sum (DAS) algorithm. DAS is efficient but fails in patients with high noise levels, so various adaptive beamformers have been proposed. Recently, deep learning methods have been developed for this task. With deep learning methods, beamforming is typically framed as a regression problem, where clean, ground-truth data is known, and usually simulated. For *in vivo* data, however, it is extremely difficult to collect ground truth information, and deep networks trained on simulated data underperform when applied to *in vivo* data, due to domain shift between simulated and *in vivo* data. In this work, we show how to correct for domain shift by learning deep network beamformers that leverage both simulated data, and unlabeled *in vivo* data, via a novel domain adaption scheme. A challenge in our scenario is that domain shift exists both for noisy input, and clean output. We address this challenge by extending cycle-consistent generative adversarial networks, where we leverage maps between synthetic simulation and real *in vivo* domains to ensure that the learned beamformers capture the distribution of both noisy and clean *in vivo* data. We obtain consistent *in vivo* image quality improvements compared to existing beamforming techniques, when applying our approach to simulated anechoic cysts and *in vivo* liver data.

Keywords: Ultrasound beamforming · Domain adaptation · Deep learning

1 Introduction

Ultrasound imaging is an indispensable tool for clinicians because it is real-time, cost-effective, and portable. However, ultrasound image quality is often

Electronic supplementary material The online version of this chapter (https://doi.org/10.1007/978-3-030-59713-9_40) contains supplementary material, which is available to authorized users.

suboptimal due to several sources of image degradation that limit clinical utility. Abdominal imaging is particularly challenging because structures of interest are beneath various tissue layers which can corrupt received channel signals due to phenomena like off-axis scattering or reverberation clutter [4].

Several advanced beamforming methods have been proposed to address this problem. Compared to conventional delay-and-sum (DAS) beamforming which applies constant delays and weights to the received channel data for a given spatial location, advanced methods aim to adaptively adjust the received channel data to enhance signals of interest and suppress sources of image degradation. This adaptive beamforming has been accomplished through coherence-based techniques [21, 22], adaptive apodization schemes [15, 33], as well as through model-based approaches [1, 2, 6, 7]. Although effective, most of these advanced methods are computationally intensive and/or limited by user defined parameters. For example, despite achieving consistently superior image quality compared to DAS as well as other advanced techniques, a model-based approach called aperture domain model image reconstruction (ADMIRE) is exceedingly computationally complex, preventing real-time adjustable implementations [8].

More recently, there has been growing interest in using deep learning methods for ultrasound beamforming. These efforts generally fall under two categories, the first having the goal of using neural networks to reconstruct fully sampled data from some form of sub-sampled receive channel data [9, 19, 30, 31, 34]. The second has the goal of using neural networks to perform a form of adaptive beamforming using physical ground truth information during training [16, 23, 24, 27, 28, 36]. The former is restricted to the desired fully sampled DAS or advanced beamforming output while the latter is theoretically capable of surpassing DAS or advanced beamformer performance. Other adaptive beamforming deep learning methods have been proposed that use an advanced beamformer as ground truth [32], which despite providing improvements to DAS, are constrained by the performance of the adaptive beamformer that they mimic.

One of the main limitations of deep learning for performing adaptive ultrasound beamforming is the lack of ground truth information *in vivo*. Previous efforts have primarily relied on simulations to generate labeled training data sets [16, 23, 27]. Unlike *in vivo* data, which is also costly to obtain, simulations can be controlled to generate unlimited amounts of realistic training data for which ground truth information is known. Network beamformers trained with simulated data have shown some success at generalizing to *in vivo* data [16, 23, 24]. However, despite sophisticated ultrasound simulation tools, a domain shift still exists between simulated and *in vivo* data, which ultimately limits network beamformer performance.

To address these limitations, we propose a novel domain adaptation scheme to incorporate *in vivo* data during training. Our approach uses cycle-consistent generative adversarial networks (CycleGANs) which learn maps between two data distributions absent of paired data [35]. GANs have been proposed previously in the context of ultrasound beamforming [28] but, to the best of our knowledge, have never been considered for performing domain adaption with real

unlabeled *in vivo* data. Further, although CycleGANs have been leveraged to address domain shift in inputs for recognition tasks [14], in our scenario domain shift exists for *both* inputs and outputs. We mitigate both sources of domain shift by composing CycleGAN maps with domain-specific regressors to effectively learn deep *in vivo* beamformers. We develop and evaluate our approach using simulated anechoic cyst and *in vivo* liver data, and compare our approach to conventional DAS, deep neural networks (DNNs) trained using simulated data only, as well as established coherence and model-based advanced beamforming techniques.

2 Methods

2.1 Domain Adaptation

The basic intuition behind our approach is to, simultaneously, learn both *regressors* for beamforming, as well as *maps* that allow us to transform simulated channel data into corresponding *in vivo* data, and vice versa. More concretely, we denote S as our simulated domain, T as our *in vivo* domain, x_s and x_t refer to input channel data for simulated and *in vivo* data, respectively, and y_s and y_t refer to output channel data for their respective S and T domains. All channel data are d -length signals that we treat as d -dimensional vectors. We are provided a set of simulated input and output pairs, each denoted (x_s, y_s) , as well as a set of *in vivo* inputs, but no corresponding outputs, denoted x_t . Our main goal is to learn a function $F_t : \mathbb{R}^d \rightarrow \mathbb{R}^d$ that serves as a beamformer for *in vivo* data. The challenge we address in this work is how to learn F_t absent of any *in vivo* outputs y_t , wherein our goal is to produce (x_t, y_t) pairs that approximate actual input/output *in vivo* pairs, and can be used to train F_t .

There are several ways to address this problem. Previous deep beamformer approaches [23] learn F_t from simulated pairs (x_s, y_s) . However, applying F_t to *in vivo* inputs leads to *domain mismatch*, as the data distributions of x_s and x_t differ, due to simulation modeling assumptions that do not always capture *in vivo* physics. Thus, the starting point for our approach is to mitigate domain shift in the inputs. Specifically, our aim is to learn a mapping $G_{S \rightarrow T}$ that takes a given x_s and maps it into a corresponding *in vivo* input x_t . A common method for learning maps between domains is the use of generative adversarial networks [12], specifically for image translation tasks [17]. Such methods, however, assume paired data, where in our case simulated and *in vivo* data are not paired. For unpaired data, the CycleGAN approach of Zhu et al. [35] proposes to learn maps from S to T , $G_{S \rightarrow T}$, and from T to S , $G_{T \rightarrow S}$, enforcing cycle-consistency between maps. Specifically, for $G_{S \rightarrow T}$ we formulate the adversarial loss as follows:

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

where D_T is the discriminator, tasked with distinguishing real *in vivo* data from fake *in vivo* data produced by $G_{S \rightarrow T}$, while X_S and X_T are the distributions for simulated and *in vivo* data respectively. We may define a similar adversarial

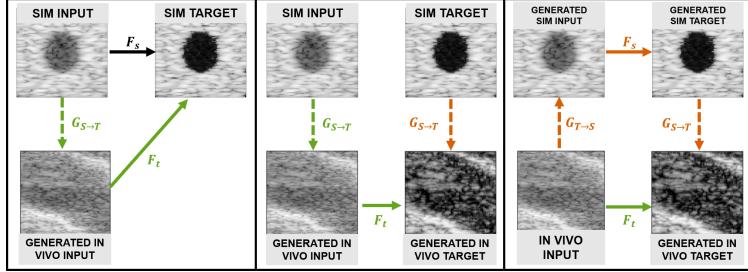


Fig. 1. Summary of input and output domain adaptation for training an *in vivo* beamformer, F_t . Green and orange arrows indicate *in vivo* input and output data generation, respectively. The left schematic summarizes previous efforts for which output domain adaptation was not considered [14]. In comparison, data used to compute $L_{F_{T1}}$ and $L_{F_{T2}}$ for the proposed method are indicated by the middle and right schematics, respectively. Data used to compute L_{F_S} is indicated by the black arrow on the left.

loss $L_{G_{T \rightarrow S}}$ for $G_{T \rightarrow S}$, with its own discriminator D_S . A cycle consistency regularization is also incorporated to ensure similarity between reconstructed signals and original data [35], defined as

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

The above discriminators and maps can be jointly optimized with F_t , where we may provide generated, paired *in vivo* data via $(G_{S \rightarrow T}(x_s), y_s)$, highlighted in Fig. 1(left). This is at the core of the CyCADA method [14]. CyCADA is focused on recognition problems, e.g. classification and semantic segmentation, and thus the target output used (class label) can be easily leveraged from the source domain. However, for our scenario this is problematic, as domain shift still exists between y_s and y_t , and thus training on $(G_{S \rightarrow T}(x_s), y_s)$ necessitates F_t to *both* resolve domain gap, and beamform.

In contrast to CyCADA [14], we would rather have F_t focus *only* on beamforming. To address this, we leverage our domain maps for which we make the assumption that the domain shift in the inputs is identical to domain shift in the outputs, and also introduce a learned function F_s for beamforming simulated data, to arrive at the following *in vivo* beamforming losses, as illustrated in Fig. 1:

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

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

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

Intuitively, $L_{F_{T1}}$ ensures F_t can beamform generated *in vivo* data produced from paired simulated data. The term $L_{F_{T2}}$ ensures that F_t can beamform real *in vivo* data. In Fig. 1, example fully reconstructed simulated anechoic cyst and *in vivo*

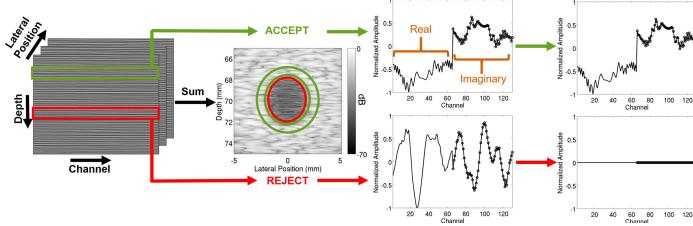


Fig. 2. Example labeled simulated anechoic cyst training data. An example channel data set and corresponding B-mode image is shown on the left to indicate accept and reject regions. A total of 10 depth locations were used as input and output to each network as illustrated by the red and green boxes on the channel data. These boxes represent 10 pixels within the corresponding red and green regions in the B-mode image. Example input and output aperture domain signals are shown on the right for a single depth. Real and imaginary components are stacked to form a 1D vector as the input and output to the DNN beamformer.

images are used, however our networks operate on aperture domain signals, as depicted in Fig. 2 and described in more detail in the following section.

Our full loss may be formulated as follows:

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

where we simultaneously optimize for discriminators, generators, and regressors. We set GAN-related weights based on Hoffman et al. [14] (i.e., $\lambda_s = 2$, $\lambda_t = 1, \lambda_c = 10$), while the regressor loss weights were empirically chosen and both set to 1. We also regularize discriminators based on the approach of Mescheder et al. [26]. Furthermore, in order to ensure that the regressors utilize distinct, and shared, features from the two domains, we learn a single regressor $F : \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}^d$ using the augmentation method of Daumé [5], such that $F_s(x_s) = F(x_s, x_s, 0)$ and $F_t(x_t) = F(x_t, 0, x_t)$, e.g. the first argument captures domain-invariant features, while the second and third arguments capture simulated and *in vivo* dependent features, respectively.

2.2 Data Summary

Our networks operate on time delayed aperture domain signals to perform a regression-based beamforming for each received spatial location. A Hilbert transform was applied to all received channel data prior to network processing to generate real and imaginary components. Training and test examples were generated from simulated anechoic cyst data as well as *in vivo* liver data. Although trivial, anechoic cysts provide clean and intuitive ground truth information and ensure an obvious domain shift between simulated and *in vivo* data.

Training Data. Field II [18] was used to simulate channel data of 12 5 mm diameter anechoic cyst realizations focused at 70mm using a 5.208 MHz center frequency, 20.832 MHz sampling frequency, 1540 m/s sound speed, and 65 active element channels with a pitch of 298 μm . These parameters were used to mimic the probe used for acquiring the *in vivo* data, as described below. Simulated training data were split into accept and reject regions depending on whether the aperture signals within a 0.5λ axial kernel (i.e., 10 depths) originated from a location outside or inside of the cyst, respectively. An output y_s is taken to be the input x_s if the signal is in the accept region, whereas the output y_s is a vector of zeros if x_s is in the reject region. Example simulated training data are shown in Fig. 2. Each aperture domain signal was concatenated through depth in addition to concatenating real and imaginary components. The number of accept and reject training examples was made equal (i.e., the full background was not used for training). For each simulated data set, 2,782 aperture domain examples (i.e., pre-reconstructed pixels) were used during training, resulting in a total of 33,384 total paired simulated examples.

A 36 year old healthy male gave informed written consent in accordance with the local institutional review board to acquire free-hand ultrasound channel data of his liver. A Verasonics Vantage Ultrasound System (Verasonics, Inc., Kirkland, WA) and ATL L7-4 (38 mm) linear array transducer were used to acquire channel data of 15 different fields of view of the liver, 6 of which were used for training. Acquisition parameters matched those used for simulations. For each of the 6 data sets 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. The same total number of examples were used from *in vivo* data as were used from simulations (i.e., 33,384 unpaired *in vivo* examples).

Test Data. For testing, 21 separate anechoic cyst realizations were simulated using the same parameters as above. White gaussian noise was added to the test realizations to achieve a signal-to-noise ratio of 50 dB. The remaining 9 *in vivo* examples not used during training were used for testing. For both the simulations and *in vivo* data, the full 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.

2.3 Evaluation

Network Details. Network hyperparameters corresponding to layer width, number of hidden layers, and regression losses (e.g. mean squared error, l_1 , Huber loss) were varied. Models were tested on *in vivo* validation data, withheld from training and testing. The model that produced the highest CNR on the validation *in vivo* data was selected. Additional details are included in supplementary materials.

Both convolutional [16, 27, 28] and fully connected [23, 36] neural networks have been investigated for the purposes of ultrasound beamforming, and it was

shown previously that there was minimal difference between the two architectures [3]. Our networks are implicitly convolutional through depth, but fully connected across the transducer elements, which is consistent with the known signal coherence patterns of ultrasound channel data [25]. For this reason, and based on the network approach used for comparison in this work [23], all networks in this work – generators, discriminators, and regressors – are implemented as fully connected.

Comparison to Established Beamformers. Both frequency [23, 36] and time [16, 27] domain approaches have been considered for ultrasound deep learning methods. Given the added complexity of our proposed training architecture and the success of other time domain implementations [16, 27], we use time domain data in this work. Therefore, as a direct baseline comparison to the proposed DA-DNN approach, a conventional DNN trained only on time domain simulated data, but with otherwise similar network parameters, was also evaluated. For this approach, the only loss used for optimization is summarized in Eq. 3. Additionally, for completeness, an established frequency-domain DNN approach [23] was also evaluated. This approach differs from the aforementioned conventional DNN approach in that it uses short time Fourier transformed (STFT) data and trains separate networks for individual frequencies. For this approach, model selection was performed as in [23] to highlight a best case scenario.

In addition to comparing the proposed DA-DNN approach to conventional and STFT DNN beamforming, performance was also evaluated in comparison to other established beamformers, including conventional DAS, the generalized coherence factor (GCF) [22], and aperture domain model image reconstruction (ADMIRE) [1]. For the GCF approach, as suggested by Li et al. [22], a cutoff spatial frequency of 3 frequency bins (i.e., $M_0 = 1$) was used to compute the weighting mask.

Performance Metrics. Contrast-to-noise ratio (CNR) and contrast ratio (CR) were used to evaluate beamformer performance as follows,

$$CNR = 20 \log_{10} \frac{|\mu_{background} - \mu_{lesion}|}{\sqrt{\sigma_{background}^2 + \sigma_{lesion}^2}} ; CR = -20 \log_{10} \frac{\mu_{lesion}}{\mu_{background}} \quad (7)$$

where μ and σ are the mean and standard deviation of the uncompressed envelope. Images were made for qualitative comparison by log compressing the envelope data and scaling to a 60dB dynamic range.

3 Results

DNN beamformers trained using only simulated data work well on simulations but often fail to generalize to *in vivo* data, as demonstrated in Fig. 3. Minimal difference was observed between the conventional DNN and DA-DNN approach

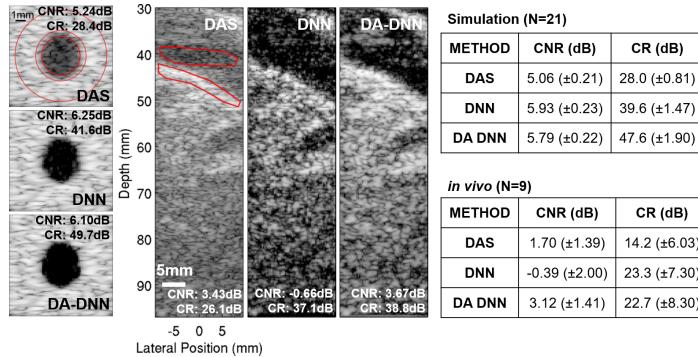


Fig. 3. Example DAS, DNN, and DA-DNN results for simulated anechoic cysts and *in vivo* data. All images are scaled to individual maximums and displayed with a 60dB dynamic range. Example regions of interest used to compute CNR and CR are indicated on the DAS images in red. Corresponding CNR and CR values are displayed on each image. Tables on the right indicate average CNR and CR (\pm standard deviation) for each method for simulations and *in vivo* data.

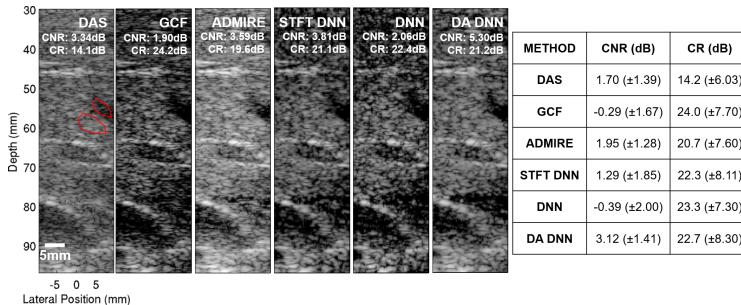


Fig. 4. Example *in vivo* B-mode images are shown for each beamformer. The regions of interest used to compute image quality metrics for the displayed example are shown in red on the DAS B-mode image. All images are scaled to individual maximums and a 60 dB dynamic range. Average CNR and CR (\pm standard deviation) across all 9 *in vivo* test examples are indicated in the table on the right.

when tested on simulated anechoic cysts in terms of CNR. In contrast, substantial improvements were observed both qualitatively and quantitatively when using the DA-DNN beamformer on *in vivo* data. Despite producing a high CR *in vivo*, the conventional DNN approach resulted in substantial drop out (i.e., extreme low amplitude pixels) in the background, resulting in a lower CNR than DAS. DA-DNN beamforming was able to maintain a higher CR compared to DAS while also improving CNR.

Qualitative and quantitative improvements in image quality with the DA-DNN beamformer were observed compared to the evaluated beamformers, as shown in Fig. 4. GCF and conventional DNN beamformers produce noticeably

better contrast than DAS, but they also result in more drop out regions compared to ADMIRE and DA-DNN. These trends are described quantitatively by the corresponding tables, for which DA-DNN produced the highest average CNR overall while still maintaining higher CR than DAS.

4 Conclusion

Conventional DNN adaptive beamforming relies on ground truth training data which is difficult to acquire *in vivo*. To address this challenge, we propose a domain adaptation scheme to train an *in vivo* beamformer absent of any labeled *in vivo* data. We demonstrated substantial image quality improvements using our proposed approach compared to conventional DNN beamforming and to other established beamformers, including DAS, GCF, and ADMIRE. We show that DA-DNN beamforming achieved image quality consistent with or higher than state of the art ADMIRE without the same computational limitations. As stated throughout, an important fundamental assumption of our 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, but it's worth investigating this further in future work.

References

1. Byram, B., Dei, K., Tierney, J., Dumont, D.: A model and regularization scheme for ultrasonic beamforming clutter reduction. *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **62**(11), 1913–1927 (2015)
2. Byram, B., Jakovljevic, M.: Ultrasonic multipath and beamforming clutter reduction: a chirp model approach. *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **61**(3), 428–440 (2014)
3. Chen, Z., Luchies, A., Byram, B.: Compact convolutional neural networks for ultrasound beamforming. In: 2019 IEEE International Ultrasonics Symposium (IUS), pp. 560–562. IEEE (2019)
4. Dahl, J.J., Sheth, N.M.: Reverberation clutter from subcutaneous tissue layers: simulation and *in vivo* demonstrations. *Ultrasound Med. Biol.* **40**(4), 714–726 (2014)
5. Daumé III, H.: Frustratingly easy domain adaptation. In: Proceedings of ACL (2007)
6. Dei, K., Byram, B.: The impact of model-based clutter suppression on cluttered, aberrated wavefronts. *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **64**(10), 1450–1464 (2017)
7. Dei, K., Byram, B.: A robust method for ultrasound beamforming in the presence of off-axis clutter and sound speed variation. *Ultrasonics* **89**, 34–45 (2018)
8. Dei, K., Schlunk, S., Byram, B.: Computationally efficient implementation of aperture domain model image reconstruction. *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **66**(10), 1546–1559 (2019)
9. Gasse, M., Millioz, F., Roux, E., Garcia, D., Liebgott, H., Friboulet, D.: High-quality plane wave compounding using convolutional neural networks. *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **64**(10), 1637–1639 (2017)

10. Glorot, X., Bengio, Y.: Understanding the difficulty of training deep feedforward neural networks. In: Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, pp. 249–256 (2010)
11. Glorot, X., Bordes, A., Bengio, Y.: Deep sparse rectifier neural networks. In: Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics, pp. 315–323 (2011)
12. Goodfellow, I., et al.: Generative adversarial nets. In: Advances in Neural Information Processing Systems, pp. 2672–2680 (2014)
13. He, K., Zhang, X., Ren, S., Sun, J.: Delving deep into rectifiers: surpassing human-level performance on ImageNet classification. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 1026–1034 (2015)
14. Hoffman, J., et al.: CyCADA: cycle-consistent adversarial domain adaptation. arXiv preprint [arXiv:1711.03213](https://arxiv.org/abs/1711.03213) (2017)
15. Holfort, I.K., Gran, F., Jensen, J.A.: Broadband minimum variance beamforming for ultrasound imaging. *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **56**(2), 314–325 (2009)
16. Hyun, D., Brickson, L.L., Looby, K.T., Dahl, J.J.: Beamforming and speckle reduction using neural networks. *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **66**(5), 898–910 (2019)
17. Isola, P., Zhu, J.Y., Zhou, T., Efros, A.A.: Image-to-image translation with conditional adversarial networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1125–1134 (2017)
18. Jensen, J.A.: Field: a program for simulating ultrasound systems. *Med. Biol. Eng. Comput.* **34**, 351–353 (1996)
19. Khan, S., Huh, J., Ye, J.C.: Universal deep beamformer for variable rate ultrasound imaging. arXiv preprint [arXiv:1901.01706](https://arxiv.org/abs/1901.01706) (2019)
20. Kingma, D.P., Ba, J.: Adam: a method for stochastic optimization. arXiv preprint [arXiv:1412.6980](https://arxiv.org/abs/1412.6980) (2014)
21. Lediju, M.A., Trahey, G.E., Byram, B.C., Dahl, J.J.: Short-lag spatial coherence of backscattered echoes: imaging characteristics. *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **58**(7), 1377–1388 (2011)
22. Li, P.C., Li, M.L.: Adaptive imaging using the generalized coherence factor. *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **50**(2), 128–141 (2003)
23. Luchies, A.C., Byram, B.C.: Deep neural networks for ultrasound beamforming. *IEEE Trans. Med. Imaging* **37**(9), 2010–2021 (2018)
24. Luchies, A.C., Byram, B.C.: Training improvements for ultrasound beamforming with deep neural networks. *Phys. Med. Biol.* **64**, 045018 (2019)
25. Mallart, R., Fink, M.: The van Cittert-Zernike theorem in pulse echo measurements. *J. Acoust. Soc. Am.* **90**(5), 2718–2727 (1991)
26. Mescheder, L., Geiger, A., Nowozin, S.: Which training methods for GANs do actually converge? arXiv preprint [arXiv:1801.04406](https://arxiv.org/abs/1801.04406) (2018)
27. Nair, A.A., Tran, T.D., Reiter, A., Bell, M.A.L.: A deep learning based alternative to beamforming ultrasound images. In: 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 3359–3363. IEEE (2018)
28. Nair, A.A., Tran, T.D., Reiter, A., Bell, M.A.L.: A generative adversarial neural network for beamforming ultrasound images: invited presentation. In: 2019 53rd Annual Conference on Information Sciences and Systems (CISS), pp. 1–6. IEEE (2019)
29. Paszke, A., et al.: Automatic differentiation in PyTorch (2017)

30. Perdios, D., Besson, A., Ardit, M., Thiran, J.P.: A deep learning approach to ultrasound image recovery. In: 2017 IEEE International Ultrasonics Symposium (IUS), pp. 1–4. IEEE (2017)
31. Senouf, O., et al.: High frame-rate cardiac ultrasound imaging with deep learning. In: Frangi, A.F., Schnabel, J.A., Davatzikos, C., Alberola-López, C., Fichtinger, G. (eds.) MICCAI 2018. LNCS, vol. 11070, pp. 126–134. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00928-1_15
32. Simson, W., et al.: End-to-end learning-based ultrasound reconstruction. arXiv preprint [arXiv:1904.04696](https://arxiv.org/abs/1904.04696) (2019)
33. Synnevag, J.F., Austeng, A., Holm, S.: Adaptive beamforming applied to medical ultrasound imaging. IEEE Trans. Ultrason. Ferroelectr. Freq. Control **54**(8), 1606–1613 (2007)
34. Yoon, Y.H., Khan, S., Huh, J., Ye, J.C.: Efficient B-mode ultrasound image reconstruction from sub-sampled RF data using deep learning. IEEE Trans. Med. Imaging **38**(2), 325–336 (2018)
35. Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2223–2232 (2017)
36. Zhuang, R., Chen, J.: Deep learning based minimum variance beamforming for ultrasound imaging. In: Wang, Q., et al. (eds.) PIPPI/SUSI -2019. LNCS, vol. 11798, pp. 83–91. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-32875-7_10