A Generative Adversarial Neural Network for Beamforming Ultrasound Images

Invited Presentation

Arun Asokan Nair*, Trac D. Tran*, Austin Reiter†, and Muyinatu A. Lediju Bell*†‡
*Department of Electrical and Computer Engineering, Johns Hopkins University, Baltimore, MD
†Department of Computer Science, Johns Hopkins University, Baltimore, MD
‡Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD

***Abstract*—Plane wave ultrasound imaging is an ideal approach to achieve maximum real-time frame rates. However, multiple plane wave insonifications at different angles are often combined to improve image quality, reducing the throughput of the system. We are exploring deep learning-based ultrasound image formation methods as an alternative to this beamforming process by extracting critical information directly from raw radio-frequency channel data from a single plane wave insonification prior to the application of receive time delays. In this paper, we investigate a Generative Adversarial Network (GAN) architecture for the proposed task. This network was trained with over 50,000 Field-II simulations, each containing a single cyst in tissue insonified by a single plane wave. The GAN is trained to produce two outputs – a Deep Neural Network (DNN) B-mode image trained to match a Delay-and-Sum (DAS) beamformed B-mode image and a DNN segmentation trained to match the true segmentation of the cyst from surrounding tissue. We systematically investigate the benefits of feature sharing and discriminative loss during GAN training. Our overall best performing network architecture (with feature sharing and discriminative loss) obtained a PSNR score of 29.38 dB with the simulated test set and 14.86 dB with a tissue-mimicking phantom. The DSC scores were 0.908 and 0.79 for the simulated and phantom data, respectively. The successful translation of the feature representations learned by the GAN to phantom data demonstrates the promise that deep learning holds as an alternative to the traditional ultrasound information extraction pipeline.**

***Index Terms*—Deep Learning, Generative Adversarial Network, Ultrasound Image Formation, Beamforming, Image Segmentation, Machine Learning**

I. Introduction

Plane wave ultrasound imaging [1] is an ideal approach to achieve maximum real-time frame rates. However, single plane wave transmits typically yield poor quality ultrasound images. This is particularly true when the data is negatively impacted by degradations such as speckle and clutter. Coherent plane wave compounding [2] mitigates this limitation by averaging the data received from multiple plane wave insonifications at different angles. While this combination improves image quality, it results in reduced frame rates. Therefore, the steep trade-off between image quality and acquisition speed still exists.

This trade-off is an inescapable outcome of the traditional ultrasound image formation process. While beamforming is principled, interpretable and model-based, model-free data-driven machine learning techniques, especially Deep Neural Networks (DNNs), have recently proven to be very successful, outperforming traditional model-based imaging methods in diverse fields from natural image classification [3] to speech recognition [4] and game playing [5].

In the past two years there has been an explosion of interest in the ultrasound community as well in using deep learning to replace various aspects of the ultrasound image formation process. Previous work from our group [6], [7] was the first to introduce and demonstrate deep neural networks that directly segment targets of interest from the raw radiofrequency (RF) single plane wave channel data, prior to applying time delays or any other traditional beamforming steps. This work is promising for tasks that rely on increased image display rates, such as robotic ultrasound tracking tasks [7]. Our work is significantly different from other methods that apply deep learning to a subset of the beamforming process and learns parameters of a model [8]–[10]. Instead, we take the novel approach of an end-to-end transformation that learns information directly from raw RF data. Our method also differs from the method presented in [11], which operates on beamformed RF channel data rather than the unfocused RF channel data that we use in our method. Similarly, our method differs from the methods presented in [12]–[15] where deep learning tasks include using DNNs for sound speed estimation [12], reverberation noise suppression in the aperture domain [13], using DNNS to produce CT-like images from B-mode ultrasound images [14], or using focused RF channel data as inputs to the network [15]. The method presented in [16] has closest underlying goal to our approach, but the presented approach trains on real data rather than simulated data and it does not use plane wave imaging.

In addition to using convolutional neural networks (CNNs) for image segmentation directly from raw RF data, we hypothesize that the introduction of a Generative Adversarial Network (GAN) [17] can help to improve our novel approach to ultrasound beamforming, based on the success achieved in previous work that applies GANs to ultrasound images. This previous work attempts to learn a better plane wave compounding operator [18], obtain better resolution by constructing 128 channel RF data from 32 channel RF data [19], and generate B-mode images from given echogenecity maps [20]. However, no previous work applies GANs to the

978-1-7281-1151-3/19/$31.00

beamforming process.

In this paper, we build on our pioneering approach to beamforming [6], [7] to demonstrate that it is possible to extract sufficient information to both reconstruct a DAS B-mode image and provide segmentation information. In addition, we use a GAN to learn representations from the training dataset simulated in Field-II that transfer better to data acquired from an anechoic and -6dB hypoechoic cyst in a phantom when compared with the performance of a plain CNN. We quantitatively and qualitatively assess the performance of the trained GANs as a function of multiple parameters of interest. To the authors' knowledge, this work is the first to extract useful information from single plane wave RF channel data prior to beamforming using a GAN framework trained purely on simulated data and transferred to real experimental data.

II. Methods

A dataset of 50,000 Field-II [21] simulations of a single cyst in normal tissue was used to train our GAN. A total of 32,000 simulations contained an anechoic cyst and 18,000 simulations contained a -6dB hypoechoic cyst. The other cyst parameters that were varied are listed in Table I. The training set included a random selection of 80% of these simulation cases, and the remaining 20% was used as the validation set.

In each simulation, a single plane wave insonification transmitted at 0 degrees was used. The parameters of an Alpinion L3-8 linear array transducer (see Table II) were used for the simulations to transfer the learned GAN to real data. The real data was acquired with an Alpinion L3-8 transducer connected to an Alpinion E-Cube 12R ultrasound research scanner.

An additional set of 100 Field-II simulations was generated to test the GAN. The parameters for this test set were chosen to randomly lie in the range of our training data. In addition, the circular cross section of two cylinders (one anechoic and one -6dB hypoechoic) at a depth of 40 mm in a CIRS 054GS phantom was imaged to test the transferability of our network trained exclusively on simulated data to experimental phantom data. For the simulations, the ground truth segmentation mask was known a priori, while the ground truth for the phantom data was acquired by manually annotating pixels in the beamformed ultrasound image as cyst or tissue.

TABLE I: Simulated cyst parameters

Parameter	Min	Max	Increment
Radius (r)	2 mm	8 mm	1 mm
Speed of Sound (c)	1440 m/s	1640 m/s	10 m/s
Lateral position of cyst center (x)	-15 mm	15 mm	2.5 mm
Axial position of cyst center(z)	35 mm	75 mm	2.5 mm

TABLE II: Simulated transducer parameters

Parameter	Value
Number of Elements	128
Pitch	0.30 mm
Aperture	38.4 mm
Element width	0.24 mm
Transmit Frequency	8 MHz
Sampling Frequency	40 MHz

An overview the GAN that we designed is presented in Figs. 1 and 2. The GAN has two major components – a generator and a discriminator. In the generator, raw RF channel data (prior to any delays) is fed into a fully convolutional encoder and decoder based on the U-Net [22] architecture. The encoder is a standard VGG-13 [23] encoder with BatchNorm [24] layers used to speed up convergence. The decoder has the structure of the encoder but mirrored, with up-convolutional layers. Skip connections are also used to connect neurons in the encoder and decoder at the same scale (as in U-Net [22]). The encoder-decoder block returns two outputs – one is a predicted B-mode-like image (which we term the DNN image), while the other is a predicted segmentation of the underlying cyst (which we term the DNN segmentation). The two DNN predictions are then stacked together with the RF channel data into a Fake (DNN) Stack and fed into the discriminator. The Real (Beamformed) stack consisted of the RF channel data and the corresponding DAS and true segmentation reference images.

The entire generator was trained to minimize the weighted sum of 3 losses:

1) Mean Absolute error (L1Loss) between the predicted DNN image and the beamformed DAS B-mode image. L1Loss is defined as
$$\text{L1Loss}(A, B) = \frac{||A - B||_1}{N}$$
where A and B are vectorized B-mode images with a total of N pixels each.
2) Dice Similarity Coefficient loss (DSCLoss) between the predicted DNN segmentation and the true segmentation.
$$DSCLoss(X, Y) = 1 - \frac{2|X \cap Y|}{|X| + |Y|}$$
where X and Y are vectorized binary segmentation masks.
3) A "realism" loss learned by the discriminator network. The lower the binary cross-entropy loss value is for the discriminator to predict the fake (DNN) stack as real, the more realistic the result is. The discriminator thus allows us to learn a non-trivial loss function that better preserves finer image details [25].

The discriminator has a much simpler structure, as shown in Fig. 2. Similar to the generator, the VGG-13 encoder with batch normalization was employed here as well. If the input stack to the discriminator was a real stack (i.e., containing beamformed images), an ideal discriminator should label the stack as such. On the other hand, if the input were a fake stack (i.e., containing images created from the DNN), the discriminator should label the stack as fake. Binary cross-entropy loss used is used to train the discriminator.

We alternate between updating the weights of the two networks. First, the generator is trained to produce good output images that spoof the discriminator. Then, the discriminator is trained to make the discriminator harder to spoof. The training regimen can be viewed as a sequential game, with both

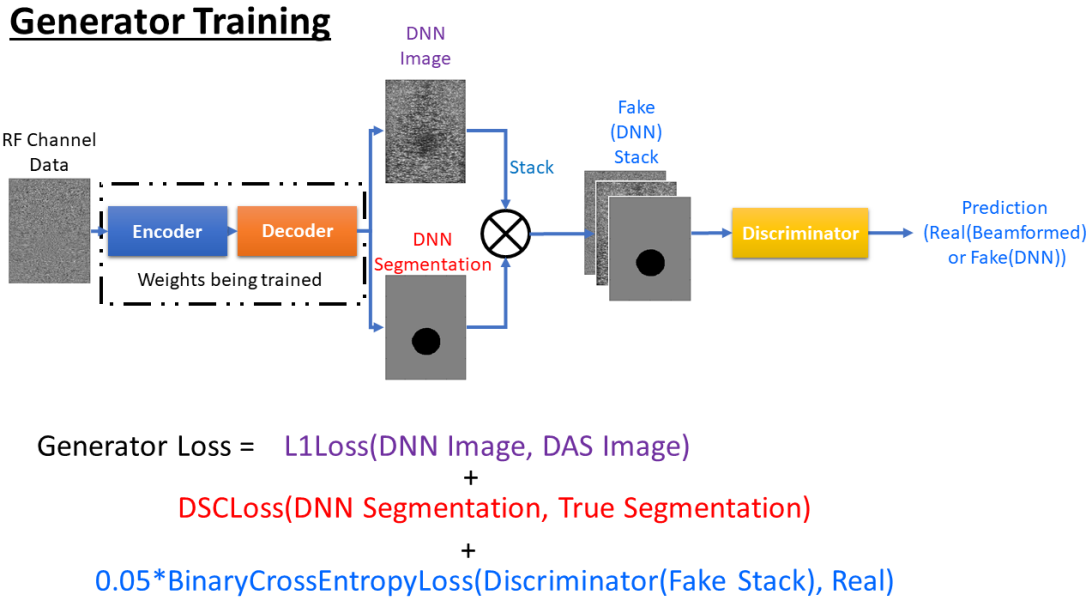

Fig. 1: An overview of our proposed pipeline relevant to the generator training update step with associated loss functions. Raw RF channel data prior to delays is transformed using a fully convolutional encoder-decoder architecture into a DNN B-mode image and a DNN segmentation. The two predicted images are stacked with the raw RF data and then fed into the discriminator, which provides a score for how "realistic" the stack is. L1 Loss guides the B-Mode prediction, DSC Loss guides the segmentation algorithm and binary cross-entropy loss is used to determine realisticness. Only the generator weights are updated during this step (discriminator weights are held fixed).

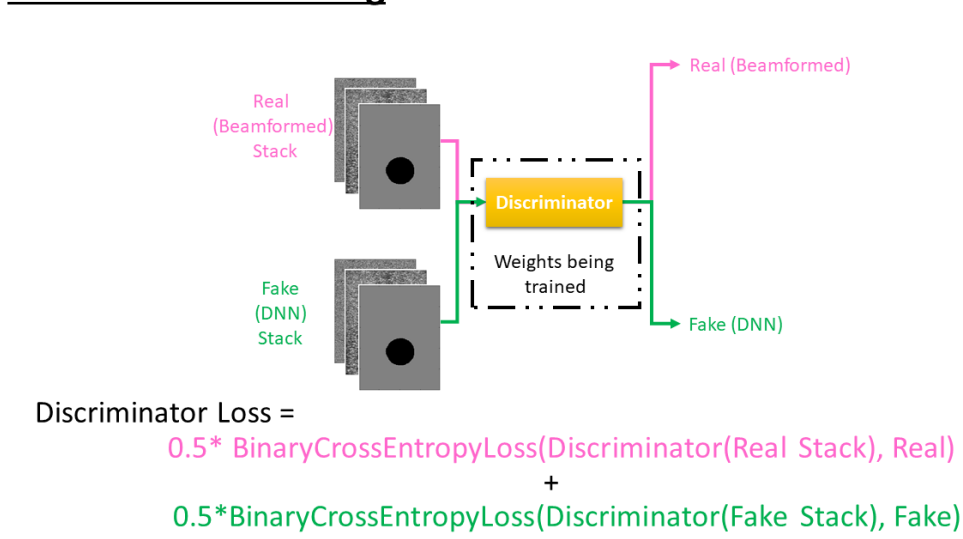

Fig. 2: An overview of our proposed pipeline relevant to the discriminator training update step with associated loss functions. A copy of the Raw RF channel data is stacked with the DAS B-Mode image obtained from beamforming and the true segmentation mask to form the "real (beamformed)" stack. Another copy of the raw RF data is stacked with the DNN B-mode image and the DNN segmentation mask to form the "fake (DNN)" stack. The discriminator is a classification network trained to label the real stack as real and the fake stack as fake. Doing so, the discriminator learns a loss function which can be used to evaluate the "realisticness" of a generated stack, which the generator can use to guide its training. Only the discriminator weights are updated during this step (generator weights are held fixed).

networks seeking to minimize their respective loss functions – hence the term "adversarial" in the name of the network. Note however that the discriminator is not updated following every generator update. The discriminator is only updated when the discriminator loss is greater than some threshold, which allows the generator (which is harder to train) to catch up to the discriminator and otherwise resulted in a failure to converge. The entire network was trained end-to-end with the Adam optimization algorithm for 100 epochs using a learning rate parameter of 10^{-5} on 4 NVIDIA P40 GPUs in parallel. The weights from the epoch with least validation loss were selected to evaluate on the test set.

To evaluate the outputs of the network, the PSNR (in dB) metric was used to evaluate the DNN images produced while Dice Similarity Coefficient (DSC) metric was used to evaluate the DNN segmentations.

A. Network Architecture Variations

Neural networks with various architectures based on Figs. 1 and 2 were trained to answer the following questions:

1) Does feature sharing help? Are better results obtained with a single encoder-decoder being used to produce both the DNN image and the DNN segmentation (exactly as in Fig. 1), or should two separate encoder-decoder blocks be used – one for each output? To answer these questions, networks were trained with both a single encoder-decoder block, as well as with two encoder-decoder blocks.
2) How much does the presence of the discriminator help? To answer this, we also trained the architecture giving no contribution to the discriminator's loss term when training the generator (i.e., when changing the weight for the discriminator's loss term from 0.05, as in Fig. 2, to 0).

Answering each question yields two variations. Therefore, four types of neural network architectures were trained in total: (1) an architecture with feature sharing and with discriminative loss, (2) an architecture without feature sharing and with discriminative loss, (3) an architecture with feature sharing and without discriminative loss, and (4) an architecture without feature sharing and without discriminative loss.

III. Results & Discussion

A. Simulation Results

The four trained architectures (single encoder-decoder with discriminator, two encoder-decoders with discriminator, single encoder-decoder without discriminator, two encoder-decoders without discriminator) were tested on the simulated test set, with mean $\pm$ one standard deviation (PSNR (dB), DSC) scores of (29.38 ± 0.92, 0.908 ± 0.076), ($28.86.\pm1.10$, 0.906 ± 0.086), (29.42 ± 1.14, 0.901 ± 0.138) and (28.35 ± 1.53, 0.895 ± 0.143), respectively.

Fig. 3 shows example images from different stages of the neural network pipeline for a simulated test example. Specifically, the results presented in Fig. 3 correspond to an architecture with a single encoder-decoder and a discriminator. The raw simulated RF channel data is input into the encoder-decoder part of the generator, producing a DNN image trained to match the reference DAS image provided, and a DNN segmentation trained to match the true segmentation. The specific outputs presented here yield a (PSNR (dB), DSC) score of (29.65, 0.962).

Fig. 4 shows DSC variation as a function of radius (r), and a clear trend is shared by the four architectures with performance increasing monotonically from an overall average (averaged over the four architectures) DSC score of 0.595 for 2 mm radii cysts to 0.975 for 8 mm radii cysts. Between the four networks, the presence of the discriminator appears to be the major distinguishing factor, greatly aiding the segmentation of small cysts – the 2 networks with a discriminator have a mean DSC score of 0.697 for 2mm radii cysts, as compared to a mean DSC score of 0.492 for the networks without a discriminator.

There is no clear trend for network performance variation with respect to the speed of sound (c), nor with respect to axial position of cyst center (z) and lateral position of cyst center (x). There is a subtle trend with respect to intrinsic contrasts, as the DSC scores of the four networks appear to degrade as intrinsic contrast increases from anechoic ($-\infty$) to -6 dB.

Studying PSNR variation as a function of radius (r) in Fig. 4, it is observed that the networks with feature sharing i.e. a single encoder-decoder block perform better than the networks without feature sharing. In particular, higher PSNR scores of 29.38 dB and 29.42 dB were obtained for the network with a single encoder-decoder and a discriminator and for the network with a single encoder-decoder and no discriminator, respectively, when compared to PSNR scores of 28.86 dB and 28.35 dB for the network with two encoder-decoders and a discriminator and for the network with two encoder-decoders and no discriminator, respectively. There does not appear to be a clear trend for network performance variation with respect to the other parameters of the simulation.

B. Phantom Results

Fig. 5 shows data from an elevational slice of an anechoic and a -6dB hypoechoic cylinder at a depth of 40 mm in a CIRS 054GS phantom imaged to test if our networks trained exclusively on simulated data generalized to experimental phantom data. Specifically, the outputs returned by the single encoder-decoder with discriminator architecture are shown in Fig. 5. The (PSNR (dB), DSC) scores of the four networks for the phantom slice were (14.86, 0.79), (14.95, 0.75), (15.06, 0.62) and (15.05, 0.49), respectively. While the PSNR scores are similar across the four network architectures, it is observed that DSC score is greatly improved by the presence of the discriminator and to a lesser extent feature sharing, showcasing the importance of the GAN framework in learning representations that generalize well.

C. Training with discriminator loss alone

As an experiment, the weights for the L1Loss and DSC Loss terms were set to 0 (instead of 1 as in Fig. 1) and the network

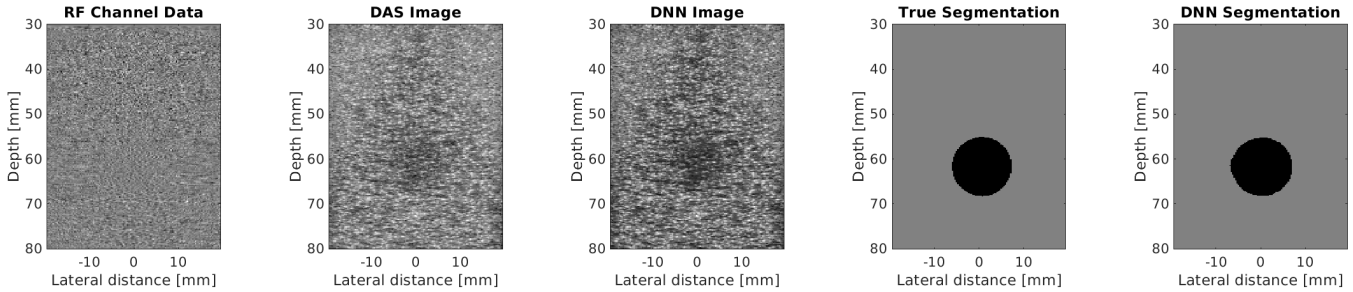

Fig. 3: Results from a simulated example. Simulated RF channel data is beamformed to obtain a DAS B-mode ultrasound image, presented alongside the DNN B-mode image produced by our network. In addition, the true segmentation of the cyst from surrounding tissue is presented with the DNN segmentation predicted by our network. Both DNN outputs are close to their reference images, yielding a (PSNR (dB), DSC) score of (29.65, 0.962).

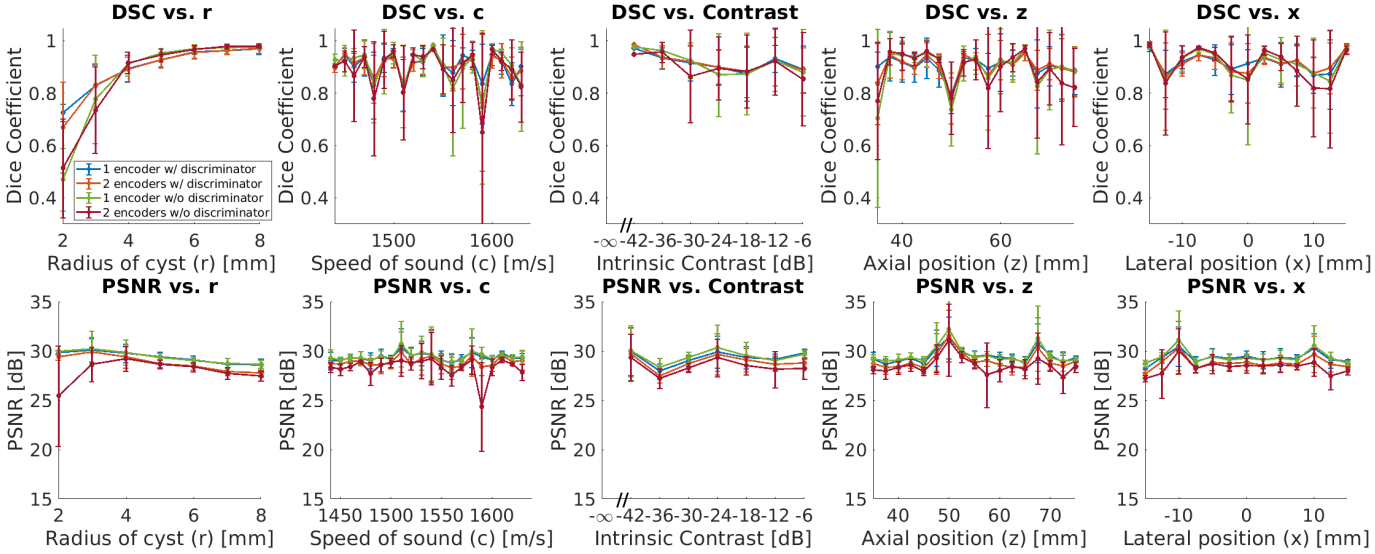

Fig. 4: Network performance with the simulated test dataset. Cyst radius (r), speed of sound (c), intrinsic cyst contrast, axial position of cyst center (z), and lateral position of cyst center (x) were varied in turn aggregating over the other parameters. Results show the mean Dice similarity coefficient (DSC) and Peak Signal-to-Noise Ratio (PSNR) with error bars showing $\pm$ one standard deviation.

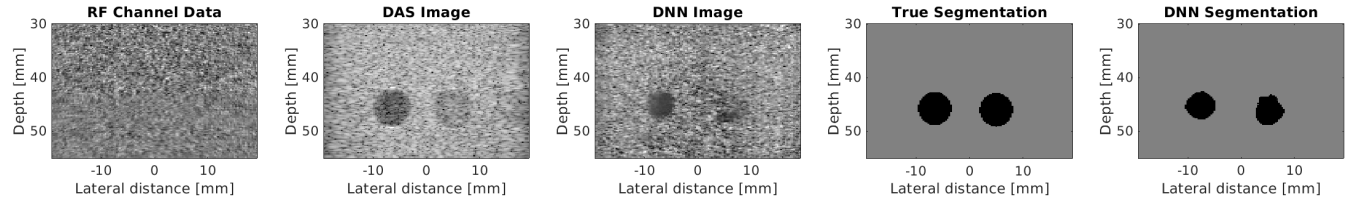

Fig. 5: Results from a phantom example. RF channel data of an elevational slice of an anechoic and -6 dB hypoechoic cylinder in a CIRS 054GS phantom is beamformed to obtain a DAS B-mode ultrasound image, presented alongside the DNN B-mode image produced by our network. In addition, the true segmentation of the cyst from surrounding tissue is presented with the DNN segmentation predicted by our network. Both DNN outputs are close to their reference images, yielding a (PSNR (dB), DSC) score of (14.86, 0.79).

was trained using only the "realism" loss provided by the discriminator. Using only the "realism" loss provided by the discriminator proved insufficient – a mean (PSNR (dB),DSC) score of just (21.75, 0.45) was obtained on the validation set. This result highlights the need for stronger supervision, provided by the L1Loss and the DSC Loss, when training a GAN to guide the weights into a good initial region by ensuring the output DNN image and DNN segmentation are close to the DAS B-mode and true segmentation.

D. Training with discriminator loss alone and without appending RF Channel data to the input stacks

We were interested in studying the representations learned if RF channel data was excluded from the stacks fed into the discriminator, and through this glean insight into how the discriminator uses the RF channel data to decide if a stack is real (beamformed) or fake (DNN). We hypothesized that failing to provide RF channel data would result in the discriminator not learning the correspondences between a given RF channel data and its B-mode image and segmentation, instead only learning if a B-mode image and its associated segmentation are a realistic pair or not. The easiest way to spoof such a discriminator would be to generate a single DNN B-Mode image and its corresponding segmentation, regardless of the input. This is a problem that often plagues GANs and is called mode collapse [26]. Experimentally, this is exactly what occurs. The generator produced a single DNN image and a single DNN segmentation, regardless of the RF channel data fed into the generator.

IV. Conclusion

A generative adversarial neural network (GAN) was trained to extract information directly from raw RF channel data created from a single plane wave transmission with no receive delays applied. The GAN produced two outputs: (1) a DNN B-mode image trained to mimic DAS B-mode imaging and (2) a DNN segmentation that matched the true segmentation of a cyst from surrounding tissue. In addition, the efficacy of feature sharing and the discriminative loss were studied in great detail. The trained networks were successfully transferred to experimental data after exclusive training on simulated cysts. This work demonstrates the promise that deep learning holds as an alternative to the traditional ultrasound image formation (i.e., beamforming) and image segmentation process.

Acknowledgments

This work is supported by NIH Trailblazer Award R21 EB025621.

References

[1] Mickael Tanter and Mathias Fink, "Ultrafast imaging in biomedical ultrasound," *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, vol. 61, no. 1, pp. 102–119, 2014.

[2] Robert R Entrekin, Bruce A Porter, Henrik H Sillesen, Anthony D Wong, Peter L Cooperberg, and Cathy H Fix, "Real-time spatial compound imaging: application to breast, vascular, and musculoskeletal ultrasound," in *Seminars in Ultrasound, CT and MRI*. Elsevier, 2001, vol. 22, pp. 50–64.

[3] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[4] Geoffrey Hinton et al., "Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups," *IEEE Signal processing magazine*, vol. 29, no. 6, pp. 82–97, 2012.

[5] David Silver, Julian Schrittwieser, Karen Simonyan, Ioannis Antonoglou, Aja Huang, Arthur Guez, Thomas Hubert, Lucas Baker, Matthew Lai, Adrian Bolton, et al., "Mastering the game of go without human knowledge," *Nature*, vol. 550, no. 7676, pp. 354, 2017.

[6] Arun Asokan Nair, Trac D Tran, Austin Reiter, and Muyinatu A Lediju Bell, "A deep learning based alternative to beamforming ultrasound images," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2018, pp. 3359–3363.

[7] Arun Asokan Nair, Mardava Rajugopal Gubbi, Trac Duy Tran, Austin Reiter, and Muyinatu A Lediju Bell, "A fully convolutional neural network for beamforming ultrasound images," in *2018 IEEE International Ultrasonics Symposium (IUS)*. IEEE, 2018, pp. 1–4.

[8] Adam C Luchies and Brett C Byram, "Deep neural networks for ultrasound beamforming," *IEEE transactions on medical imaging*, vol. 37, no. 9, pp. 2010–2021, 2018.

[9] Adam C Luchies and Brett C Byram, "Training improvements for ultrasound beamforming with deep neural networks," *Physics in medicine and biology*, 2019.

[10] Sanketh Vedula, Ortal Senouf, Grigoriy Zurakhov, Alex Bronstein, Oleg Michailovich, and Michael Zibulevsky, "Learning beamforming in ultrasound imaging," *arXiv preprint arXiv:1812.08043*, 2018.

[11] Maxime Gasse, Fabien Millioz, Emmanuel Roux, Damien Garcia, Hervé Liebgott, and Denis Friboulet, "High-quality plane wave compounding using convolutional neural networks," *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, vol. 64, no. 10, pp. 1637–1639, 2017.

[12] Micha Feigin, Daniel Freedman, and Brian W Anthony, "A deep learning framework for single sided sound speed inversion in medical ultrasound," *arXiv preprint arXiv:1810.00322*, 2018.

[13] Leandra L Brickson, Dongwoon Hyun, and Jeremy J Dahl, "Reverberation noise suppression in the aperture domain using 3d fully convolutional neural networks," in *2018 IEEE International Ultrasonics Symposium (IUS)*. IEEE, 2018, pp. 1–4.

[14] Sanketh Vedula, Ortal Senouf, Alex M Bronstein, Oleg V Michailovich, and Michael Zibulevsky, "Towards ct-quality ultrasound imaging using deep learning," *arXiv preprint arXiv:1710.06304*, 2017.

[15] Yeo Hun Yoon, Shujaat Khan, Jaeyoung Huh, and Jong Chul Ye, "Deep learning in rf sub-sampled b-mode ultrasound imaging," *arXiv preprint arXiv:1712.06096*, 2017.

[16] Walter Simson, Magdalini Paschali, Nassir Navab, and Guillaume Zahnd, "Deep learning beamforming for sub-sampled ultrasound data," in *2018 IEEE International Ultrasonics Symposium (IUS)*. IEEE, 2018, pp. 1–4.

[17] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio, "Generative adversarial nets," in *Advances in neural information processing systems*, 2014, pp. 2672–2680.

[18] Xi Zhang, Jing Li, Qiong He, Heye Zhang, and Jianwen Luo, "High-quality reconstruction of plane-wave imaging using generative adversarial network," in *2018 IEEE International Ultrasonics Symposium (IUS)*. IEEE, 2018, pp. 1–4.

[19] Chao-Yi Huang, Oscal Tzyh-Chiang Chen, Guo-Zua Wu, Chih-Chi Chang, and Chang-Lin Hu, "Ultrasound imaging improved by the context encoder reconstruction generative adversarial network," in *2018 IEEE International Ultrasonics Symposium (IUS)*. IEEE, 2018, pp. 1–4.

[20] Francis Tom and Debdoot Sheet, "Simulating patho-realistic ultrasound images using deep generative networks with adversarial learning," in *Biomedical Imaging (ISBI 2018), 2018 IEEE 15th International Symposium on*. IEEE, 2018, pp. 1174–1177.

[21] Jørgen Arendt Jensen, "Field: A program for simulating ultrasound systems," in *10TH NORDICBALTIC CONFERENCE ON BIOMEDICAL IMAGING, VOL. 4, SUPPLEMENT 1, PART 1: 351–353*. Citeseer, 1996.

[22] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.

[23] Karen Simonyan and Andrew Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[24] Sergey Ioffe and Christian Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *arXiv preprint arXiv:1502.03167*, 2015.

[25] Ashish Shrivastava, Tomas Pfister, Oncel Tuzel, Joshua Susskind, Wenda Wang, and Russell Webb, "Learning from simulated and unsupervised images through adversarial training," in *CVPR*, 2017, pp. 2107–2116.

[26] Martin Arjovsky, Soumith Chintala, and Léon Bottou, "Wasserstein gan," *arXiv preprint arXiv:1701.07875*, 2017.