Improving Cardiac Ultrasound with a Semi-Supervised Deep Learning Beamformer

Ying-Chun Pan*, Ryan LeFevre†, Susan Eagle†, Matthew Berger* and Brett Byram*
* Vanderbilt University School of Engineering
Vanderbilt University, Nashville, Tennessee, USA
† Department of Anesthesiology
Vanderbilt University Medical Center, Nashville, Tennessee, USA

2023 IEEE International Ultrasonics Symposium (IUS) | 979-8-3503-4645-9/23/$31.00 ©2023 IEEE | DOI: 10.1109/IUS51837.2023.10306476

***Abstract*—Deep learning beamformers have demonstrated the ability to remove a variety of artifacts from ultrasound images in recent years. Many of these algorithms operate on channel data, where *in vivo* ground truth data, free of any degradation, is unavailable. Much of the existing work estimates the ground truth distribution with synthetic data. Under this framework, the domain gap between the synthetic training data and *in vivo* test data limits the beamformer performance on inference. In this work, we introduce a multi-step, semi-supervised approach that leverages synthetic and *in vivo* data in training via cross-domain cycleGANs. We evaluate the intermediate generators with VCZ curves, and demonstrate that the beamformer trained with the proposed approach achieves a CNR gain of 2.72 ± 1.40 dB and gCNR gain of 0.284 ± 0.094 over delay-and-sum in a 32-frame test cine loop.**

***Index Terms*—semi-supervised networks, domain adaptation, beamformer**

I. Introduction

In recent years, deep learning beamformers have demonstrated potential in improving ultrasound image quality [1]–[3]. Operating at the image-formation stage, these beamformers need to contend with missing ground truth data. While easy-to-image patients certainly make better images than their difficult-to-image counterparts, it is unclear whether *in vivo* data free of any sources of image degradation is obtainable. To address this limitation, many groups rely on simulations [1]–[3]. Simulations [4]–[6] provide researchers with unparalleled control over the parameters that govern the outcome. Setting these parameters to some theoretical limit is arguably a closer approximation to the ground truth than any *in vivo* samples. Furthermore, some simulations are designed to apply to arbitrary inputs (e.g., scatter strength and position in Field II), enabling researchers to build up a diverse data set in little time. The main drawback of using simulations is the significant mismatch between training (synthetic) and test (*in vivo*) data. This paper discusses one approach to address this domain gap.

II. Methods

We propose a semi-supervised network that incorporates synthetic and *in vivo* data in training using cross-domain maps. This work is a more thorough investigation into the original domain-adaptive deep neural network proposed by Tierney et al. [7] with two important contributions. 1) This work eliminates the constraint that clean output style transfer map be identical to the noisy input style transfer map, and 2) this work evaluates the quality of these style transfer maps with van Cittert-Zernike (VCZ) curves [8]. In section II-A we describe the synthetic and *in vivo* data used in this work. In section II-B we define the four domains in our problem, recap the original work [7], and highlight the main difference in our approach. In section II-C we introduce the multi-step training approach. And lastly, in section II-D we discuss the validation of cross-domain maps with VCZ curves.

A. Data preparation

We trained our networks on delayed channel data. Each training example contained 10 axial samples (with an overlap of 90%) and one receive beam. We used analytic signal representation, and stacked the imaginary samples after the real samples [7]. The synthetic data consisted of six hypoechoic cysts (CR=10 dB) and six anechoic cysts (CR=Inf dB). The noisy synthetic data was corrupted with various levels of reverberation clutter (signal to clutter ratio of 0 or 10 dB) simulated in Field II [9]. The clean synthetic data had no clutter (SCR = ∞ dB), and off-axis scattering was removed with the same method proposed by Luchies et al. [3]. The *in vivo* training data included four consecutive frames of cardiac cineloops from three patients. Fig. 1 shows the first frame of the three cine loops.

B. Cross-domain Maps

When leveraging paired synthetic data to train a network that generalizes well to *in vivo* data, we categorize the data into four domains (Fig. 2). Noisy synthetic (x_s) and clean synthetic (y_s) domains provide the paired data required by the conventional supervised framework. Noisy *in vivo* (x_t) is the intended input domain, and the missing clean *in vivo* (y_t) marks the last domain that is only estimated.

In the original semi-supervised beamformer [7], Tierney et al. trained a pair of cross-domain maps between x_s and x_t using CycleGANs [10]. These maps enabled synthetic and *in vivo* data to contribute to the final, universal regressor (F) which used a set of domain-agnostic weights and another set of domain-specific weights when regressing synthetic data (denoted with F_s) and *in vivo* data (F_t) through the augmented

This work was supported by the NIH award R01HL156034 and S10OD016216-01, and NSF grant IIS-175099

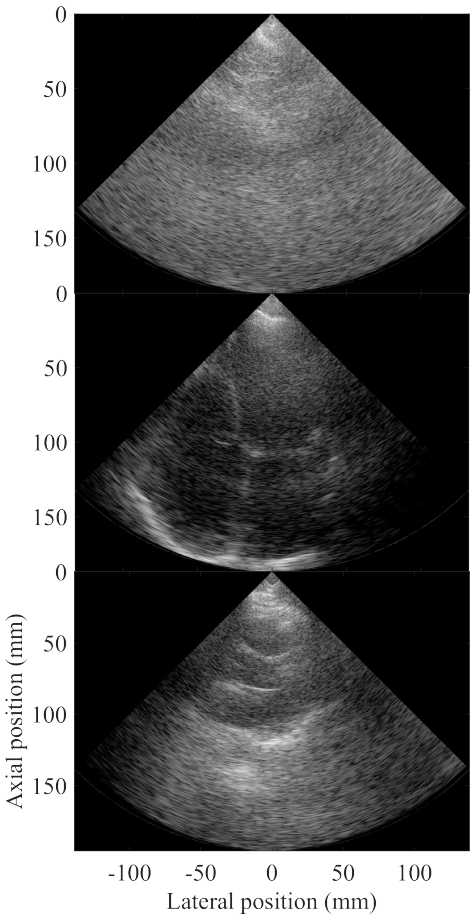


Fig. 1: *In vivo* training data consists of echocardiographs of different quality and views. All images are normalized by the maximum absolute value and shown on a 60 dB dynamic range.

feature mapping technique [11]. Altogether, these components composed the following regressor loss:

$$\begin{aligned}\mathcal{L}_{gen1}(F; G_{st}, G_{ts}) = &||F_t(x_t) - G_{st}(F_s(G_{ts}(x_t)))|| \\ &+ ||F_t(G_{st}(x_s)) - G_{st}(y_s)|| \\ &+ ||F_s(x_s) - y_s||\end{aligned}$$

This approach assumed that the cross-domain map on the clean side is identical to the map trained between the noisy domains. In the rest of this work we eliminate this assumption by training two additional cross-domain maps on the clean side.

C. Multi-step Approach

Training the cross-domain maps between the clean domains (y_s and y_t) requires an estimate of y_t. To achieve this, we adopt a three-step approach. In the first step, we train an intermediate regressor (F_1) with Tierney et al.'s approach. We label the synthetic and *in vivo* components of this network with F_{s1} and F_{t1} respectively. Along the way, we obtain some initial estimate of G_{st} and G_{ts}. In the second step, we initialize the new generators on the clean output side (G'_{st}, G'_{ts}) with the noisy generators from step one (G_{st}, G_{ts}) respectively, and estimate the clean *in vivo* distribution $\hat{y}_t = F_{t1}(x_t)$. Using $\hat{y}_t$ and y_s, we fine-tune G'_{st} and G'_{ts}. In the third and final step, after the losses for all four generators have stabilized, we reinitialize a new regressor $F_2 = F_1$ and fine-tune it with the following losses:

$$\begin{aligned}\mathcal{L}_{gen2}(F_2; G_{st}, G_{ts}, G'_{st}) = &||F_{t2}(x_t) - G'_{st}(F_{s2}(G_{ts}(x_t)))|| \\ &+ ||F_{t2}(G_{st}(x_s)) - G'_{st}(y_s)|| \\ &+ ||F_{s2}(x_s) - y_s||\end{aligned}$$

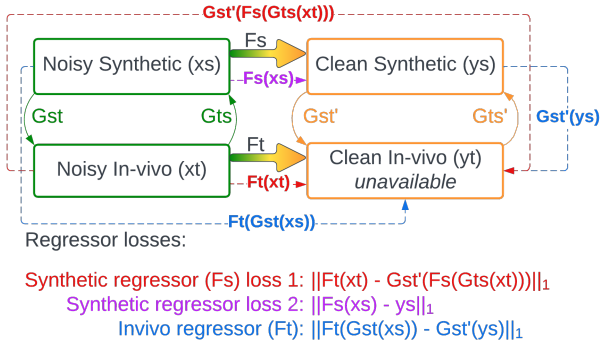


Fig. 2: Proposed framework. Curved arrows indicate generators trained to map between domains. Each generator has an associated discriminator that is omitted for notational simplicity.

D. van Cittert-Zernike Evaluation

The proposed approach involves a total of ten networks (four cross-domain generators each with their own discriminator, the intermediate regressor F_1, and the final regressor F_2. Each of these regressors also has domain-specific weights for synthetic F_s and *in vivo* F_t data). While regressors can be evaluated with image quality metrics (e.g., CNR, gCNR), the intermediate generators are harder to evaluate. Here we propose to evaluate the synthetic-to-*in vivo* maps G_{st} and G'_{st} with VCZ curves of diffuse scatterers. The VCZ theorem states that "in the case of a focused illumination, the spatial covariance of the backscattered pressure field is proportional to the autocorrelation of the transmitting aperture function" [8]. In our simulation of phased array with no transmit apodization, this implies the coherence function should be a triangular function whose base is twice the aperture size. In this analysis, we simulated 12 realizations of diffuse scatterers, mapped them to the style of noisy *in vivo* data (with G_{st}) and clean *in vivo* data (with G'_{st}), and computed their VCZ curves. To our knowledge, this is the first time VCZ curves are used to evaluate any component of a deep learning beamformer.

III. Results

Cross-domain maps G_{st} and G'_{st} learned distinct styles. When diffuse scatterers were passed into G_{st}, the output contained several sources of degradation including phase aberration and reverberation (Fig. 3). Recall in section II-A that the synthetic data only contains off-axis scattering and reverberation clutter. This implies that the synthetic-to-*in vivo* map on the noisy side learned to recreate phase aberration from *in vivo* data. This suggests that cross-domain maps are capable of bridging the domain gap. In contrast, diffuse scatterers mapped with G'_{st} retained the gradual decorrelation across the aperture that is expected of an incoherent source.

We also saw this difference in coherence with the VCZ curves (Fig. 4). The coherence of synthetic diffuse target followed the triangular function predicted by the VCZ theorem. G_{st} introduced *in vivo* noise sources into the diffuse scatterer data and caused the coherence to roll off quickly [12]. In contrast, G'_{st} learned to transform diffuse scatterers without significantly changing the coherence curve expected of an incoherent source.

Testing the final regressor F_2 on a fourth patient's cineloop of 32 frames, we found that our method outperformed conventional delay-and-sum (DAS) in several metrics (Table. I). *In vivo* comparison are shown in Fig. 5a and Fig. 5b.

TABLE I: Image quality metrics evaluated on a 32-frame cine loop

	CNR (dB)	**gCNR**
DAS	1.05 $\pm$ 0.97	0.698 $\pm$ 0.09
Proposed	3.77 $\pm$ 0.99	0.982 $\pm$ 0.03

IV. Discussions

The proposed method is highly flexible and imposes few constraints on data distribution. This advantage comes at the cost of a large number of networks and optimization challenges. The multi-step optimization approach was designed to mitigate training instability. And thankfully, the VCZ theorem defines some expected outcome that our cross-domain maps satisfied in a post hoc analysis. A future step is to incorporate the VCZ curves into the loss function to guide the cross-domain maps in training.

Another aspect to investigate is a lack of output diversity. The narrow error bars in Fig. 4 show that outputs mapped by the noisy generator G_{st} are similar. This does not fully represent the diversity of training data quality shown in Fig. 1 and suggests mode collapse in the network. Future iterations of this work should consider alternative GAN losses that are more robust against mode collapse (e.g., Wasserstein GAN [13]).

V. Conclusion

In this work we proposed a multi-step, semi-supervised training approach for training a beamformer, and showed that cross-domain maps between the noisy input domains (G_{st}, G_{ts}) behave differently from the maps between the clean output domains (G'_{st}, G'_{ts}). These distinct maps enabled us to more accurately incorporate synthetic and *in vivo* data when training a final, universal regressor (F_2).

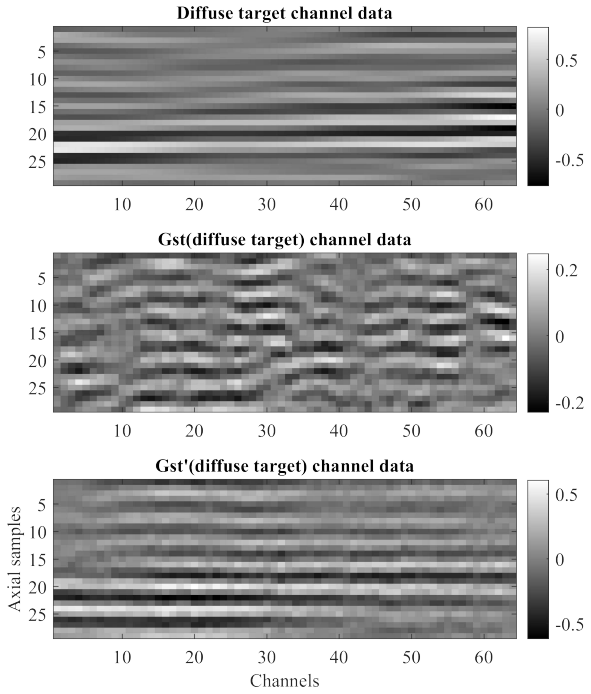


Fig. 3: Diffuse scatterers mapped with two different generators. G_{st} introduces sources of image degradation to channel data of diffuse scatterers (middle panel), whereas G'_{st} transforms the input data while preserving the spatial coherence (bottom panel).

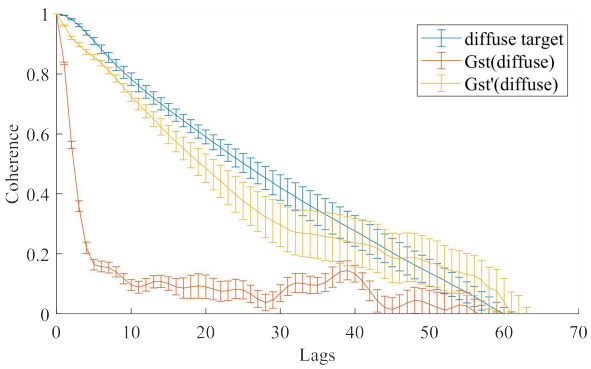
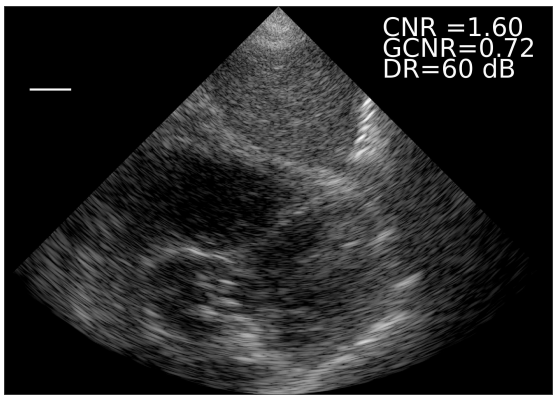


Fig. 4: VCZ curves of diffuse scatterers and outputs mapped by two different generators. This figure summarizes the visual differences observed in Fig. 3, showing the average coherence computed for a rectangular window centered around the transmit focus.

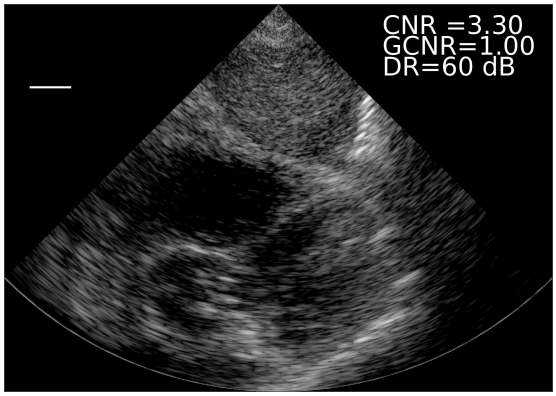
References

[1] M. Sharifzadeh, H. Benali, and H. Rivaz, "Phase aberration correction: A convolutional neural network approach," *IEEE Access*, vol. 8, pp. 162252–162260, 2020.

[2] W. A. Simson, M. Paschali, V. Sideri-Lampretsa, N. Navab, and J. J. Dahl, "Investigating pulse-echo sound speed estimation in breast ultrasound with deep learning," *arXiv:2302.03064*, 2023.



(a) DAS



(b) Proposed

Fig. 5: Delay-and-sum (DAS) and proposed beamformer applied to a test image. The scale bar is two cm.

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

[4] J. Jensen, "Field: A program for simulating ultrasound systems," *Medical & Biological Engineering & Computing*, vol. 34, pp. 351–353, 1996.

[5] J. Jensen and N. Svendsen, "Calculation of pressure fields from arbitrarily shaped, apodized, and excited ultrasound transducers," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 39, no. 2, pp. 262–267, 1992.

[6] B. E. Treeby and B. T. Cox, "k-wave: Matlab toolbox for the simulation and reconstruction of photoacoustic wave-fields," *Journal of Biomedical Optics*, vol. 15, no. 2, p. 021314, 2010.

[7] J. Tierney, A. Luchies, C. Khan, J. Baker, D. Brown, B. Byram, and M. Berger, "Training deep network ultrasound beamformers with unlabeled in vivo data," *IEEE Transactions on Medical Imaging*, vol. 41, no. 1, pp. 158–171, 2022.

[8] R. Mallart and M. Fink, "The van Cittert–Zernike theorem in pulse echo measurements," *The Journal of the Acoustical Society of America*, vol. 90, pp. 2718–2727, 11 1991.

[9] B. Byram and J. Shu, "Pseudononlinear ultrasound simulation approach for reverberation clutter," *Journal of Medical Imaging*, vol. 3, no. 4, p. 046005, 2016.

[10] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, Oct 2017.

[11] H. Daumé III, "Frustratingly easy domain adaptation," in *Conference of the Association for Computational Linguistics (ACL)*, (Prague, Czech Republic), 2007.

[12] G. F. Pinton, G. E. Trahey, and J. J. Dahl, "Spatial coherence in human tissue: implications for imaging and measurement," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 61, no. 12, pp. 1976–1987, 2014.

[13] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein generative adversarial networks," in *Proceedings of the 34th International Conference on Machine Learning* (D. Precup and Y. W. Teh, eds.), vol. 70 of *Proceedings of Machine Learning Research*, pp. 214–223, PMLR, 06–11 Aug 2017.