Ultrasound Domain Adaptation Using Frequency Domain Analysis

Mostafa Sharifzadeh, Ali K. Z. Tehrani, Habib Benali, and Hassan Rivaz

Department of Electrical and Computer Engineering

Concordia University

Montreal, QC, Canada

mostafa.sharifzadeh@concordia.ca

Abstract—**A common issue in exploiting simulated ultrasound data for training neural networks is the domain shift problem, where the trained models on synthetic data are not generalizable to clinical data. Recently, Fourier Domain Adaptation (FDA) has been proposed in the field of computer vision to tackle the domain shift problem by replacing the magnitude of the low-frequency spectrum of a synthetic sample (source) with a real sample (target). This method is attractive in ultrasound imaging given that two important differences between synthetic and real ultrasound data are caused by unknown values of attenuation and speed of sound (SOS) in real tissues. Attenuation leads to slow variations in the amplitude of the B-mode image, and SOS mismatch creates aberration and subsequent blurring. As such, both domain shifts cause differences in the low-frequency components of the envelope data, which are replaced in the proposed method. We demonstrate that applying the FDA method to the synthetic data, simulated by Field II, obtains an 3.5% higher Dice similarity coefficient for a breast lesion segmentation task.**

I. INTRODUCTION

The performance of deep neural networks (DNNs) is highly correlated with the amount of available data for training, where a larger training set leads to better results. Therefore, insufficient data have always been a concern particularly in the medical domain, where clinical data are not easily accessible due to multiple concerns, and the performance of these methods is hindered by this issue. Using synthetic data for training networks is a popular approach to address this challenge. However, in most cases, the trained models on synthetic data are not well generalizable in real-world applications, where we need to deal with real imaging data acquired by various scanners and different protocols [1].

This issue originally comes from the fact that DNNs typically assume that both training and test sets have been drawn from the same distribution [2], which is not necessarily true, especially regarding the recent trend of using synthetic data for training. This problem is usually referred to as the domain shift problem, which induces a dramatic performance drop [3].

Domain adaptation methods are a well-known solution to address the domain shift problem and have been investigated in the medical ultrasound domain. Tierney *et al.* proposed a scheme that incorporates both simulated and unlabeled *in vivo* data to train a beamformer. They employed cycle-consistent generative adversarial networks to map between simulated and *in vivo* data in both the input and ground truth target domains

[4], [5]. Ying *et al.* introduced a multi-scale self-attention unsupervised network for domain adaptation between labeled thyroid ultrasound images and unlabeled ones in a different domain [6]. Meng *et al.* introduced mutual information-based disentangled neural networks for classifying unseen categories of fetal ultrasound images in different domains. They extracted generalizable categorical features by explicitly disentangling categorical and domain features via mutual information minimization to transfer knowledge to unseen categories in a target domain [7]. Zhang *et al.* proposed a deep-stacked transformation approach for generalizing medical image segmentation models to unseen domains and evaluated it on segmentation tasks involving MRI and ultrasound modalities [8].

Recently, Yang *et al.* [9] introduced the Fourier Domain Adaptation (FDA) method in the field of computer vision. They proposed that moving sample A from the source distribution to the distribution of sample B in the target dataset can be achieved by computing the Fast Fourier Transform (FFT) of both samples and substituting the magnitude of the low-frequency spectrum of the source sample with the target sample and finally reconstructing the modified source sample using inverse FFT (IFFT). This method is much faster than DNN-based methods, and they demonstrated promising results to adapt synthetic dataset GTA5 [10] to the real domain dataset CityScapes [11], which both contain urban street scenes.

We believe this method can perform even better on ultrasound images than urban street scenes because two common differences between synthetic and real ultrasound data are caused by unknown values of attenuation and speed of sound (SOS) in real tissues. Attenuation leads to slow variations in the amplitude of the B-mode image, and a mismatch between the nominal and true values of the SOS creates aberration and subsequent blurring. As such, both of these domain shifts are low-frequency in nature and can be compensated by swapping the low-frequency spectrum of the synthetic and real image.

In this work, for the first time, we exploit the FDA method to mitigate the domain shift problem of synthetic ultrasound images and evaluate its performance in a breast lesion segmentation task.

II. METHODOLOGY

Let $I \in \mathbb{R}^{W \times H}$ and $\hat{S} \in \{0,1\}^{W \times H}$ denote a sample input image and the corresponding output segmentation mask,

respectively, where W and H are width and height of the image. The segmentation problem can be formulated as

$$\hat{S} = f_{seg}(I, \boldsymbol{\theta}) \tag{1}$$

where $f_{seg} : \mathbb{R}^{W \times H} \to \{0, 1\}^{W \times H}$ is the segmentation convolutional neural network (CNN), and $\boldsymbol{\theta}$ are the network's parameters. By training the CNN, an optimizer is utilized to find optimal parameters $\boldsymbol{\theta}^*$ that minimize the error, measured by a loss function L, between predicted mask $\hat{S}$ and ground truth S

$$\boldsymbol{\theta}^* = \underset{\theta}{argmin} \, L(S, \hat{S}) \tag{2}$$

A. Datasets

1) Synthetic Dataset: We simulated 1000 ultrasound images using the publicly available Field II simulation package [12], [13] containing 100,000 scatterers uniformly distributed inside a phantom of size 50 mm × 10 mm × 50 mm in x, y, and z directions, respectively. All phantoms were centered at the focal point, positioned at an axial depth of 20 mm from the face of the transducer, and each contained an anechoic region with a different shape. To generate those anechoic regions, we took 1000 samples with only one salient object from a publicly available dataset, denoted as XPIE [14], which contained segmented natural images. Then we discarded natural images and resampled only their ground truth masks with the same size as the phantom. Finally, we assigned a zero weight to the amplitude of scatterers which were located inside the mask. The advantages of this method were twofold: First, the mask could be considered as the ground truth of the simulated images. Second, we provided the network with an extended range of features as opposed to regions with limited shapes. Finally, we resampled all images to a size of 256×256 and split the dataset into two training and validation sets, each containing 800 and 200 images, respectively. Note that this data was only used for training and validation and did not contain a test set. The simulation parameters are summarized in Table I.

TABLE I
FIELD II PARAMETERS FOR DATA SIMULATION.

Parameter	Value
Center Frequency	3.5 MHz
Total/Active Number of Elements	192/64
Element Height	5 mm
Element Width	Equals to wavelength
Kerf	0.05 mm

2) In vivo Dataset: We exploited an ultrasound breast images dataset, known as Dataset B [15]. The dataset was publicly available and collected in 2012 from the UDIAT Diagnostic Centre with a Siemens ACUSON Sequoia C512 system and a 17L5 HD linear array transducer. It included 163 breast B-mode ultrasound images containing lesions of different sizes at different locations, with a mean image size

of 760×570 pixels. Lesions were categorized into benign and cancerous classes, with 110 and 53 samples in each class, respectively. The dataset also contained respective ground truth masks of the breast lesions, manually obtained by experienced radiologists. We resampled all images to a size of 256×256, and split the dataset into three training, validation, and test sets, each containing 20, 20, and 123 images, respectively.

B. Fourier Domain Adaptation

To mitigate the domain shift problem, the FDA method [9] suggests replacing the magnitude of the low-frequency spectrum of source samples with target samples. Let $I_s \in \mathbb{R}^{W \times H}$, and $I_t \in \mathbb{R}^{W \times H}$ represent a simulated image using Field II (source) and an *in vivo* image (target), respectively. Besides, let $\mathcal{F}_M(I) : \mathbb{R}^{W \times H} \to \mathbb{R}^{W \times H}$ and $\mathcal{F}_P(I) : \mathbb{R}^{W \times H} \to \mathbb{R}^{W \times H}$ be the magnitude and phase of the Fourier transform $\mathcal{F}$ of the image I:

$$\mathcal{F}(I)(m, n) = \sum_{w=0}^{W-1} \sum_{h=0}^{H-1} I(w, h) e^{-j2\pi(\frac{h}{H}n + \frac{w}{W}m)} \tag{3}$$

Accordingly, given $\mathcal{F}_M(I)$ and $\mathcal{F}_P(I)$, $\mathcal{F}^{-1}$ is the inverse Fourier transform that converts back the signal from the frequency domain to the image domain.

$$I = \mathcal{F}^{-1}(\mathcal{F}_M(I), \mathcal{F}_P(I)) \tag{4}$$

where $j^2 = -1$, and (3) and (4) can be implemented using FFT [16] and IFFT algorithms, respectively.
Further, let denote with M_α a mask matrix of size $W \times H$:

$$M_\alpha(w, h) = \begin{cases} 1, & -\alpha < \frac{2w}{W} - 1 < \alpha, -\alpha < \frac{2h}{H} - 1 < \alpha \\ 0, & otherwise \end{cases} \tag{5}$$

Finally, given a pair of simulated and *in vivo* images, the FDA method can be formalized as:

$$I_{s \to t} = \mathcal{F}^{-1}(M_\alpha \cdot \mathcal{F}_M(I_t) + (1 - M_\alpha) \cdot \mathcal{F}_M(I_s), \mathcal{F}_P(I_s)) \tag{6}$$

where $\alpha \in (0, 1)$. In this work, we set $\alpha = 0.014$. Fig. 1 illustrates the FDA method. It shows (a) a simulated image I_s, and (b) a real ultrasound image I_t from the *in vivo* dataset. After taking the FFT of both images, the magnitude of the low-frequency spectrum of the simulated image has been replaced with the real one. Finally, by taking the IFFT, the output $I_{s \to t}$ has been obtained.

C. Network Architecture and Training Strategy

We used a vanilla U-Net [17] to evaluate the performance of the FDA method on ultrasound images for a segmentation task. The network was proposed specifically for biomedical image segmentation, where the number of available annotated samples for training is limited. It is composed of an encoder followed by a decoder, where skip connections are also adapted to concatenate low-level features of the encoder with high-level ones in the decoder.
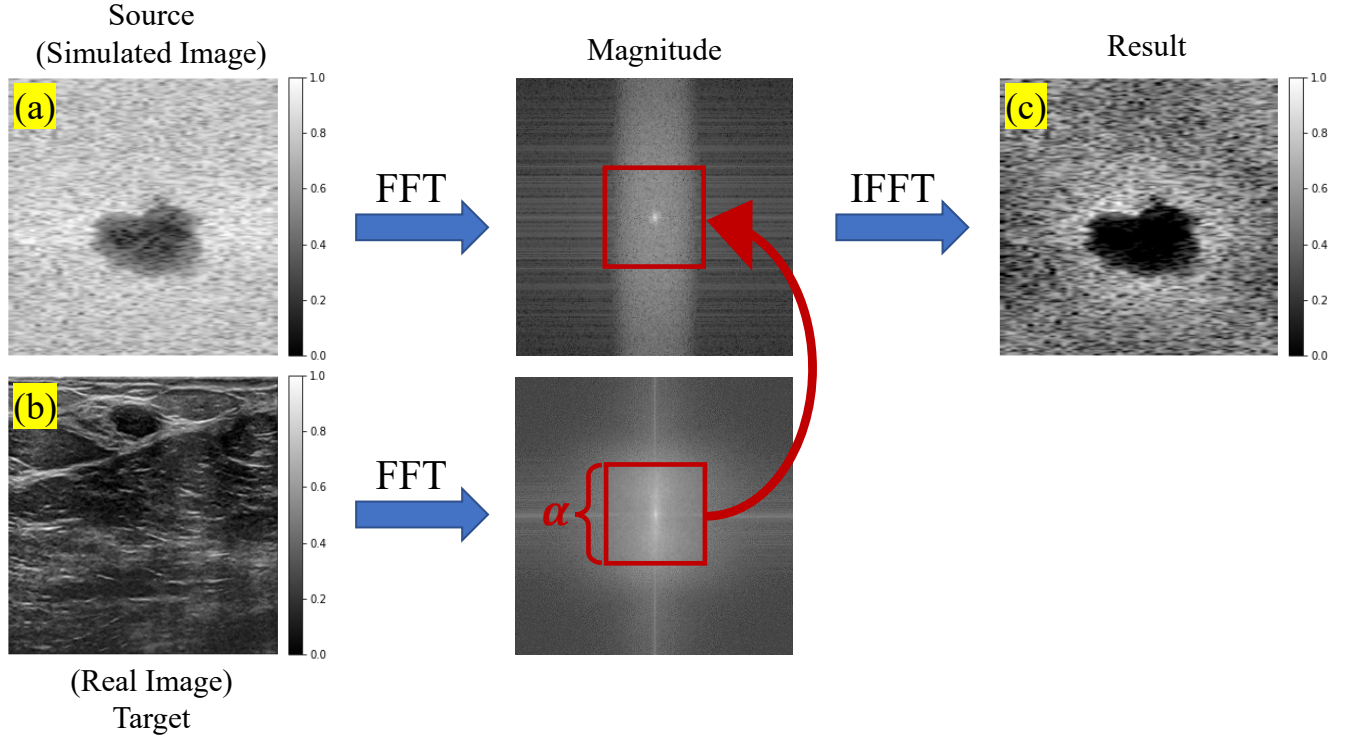
Fig. 1. The FDA method takes the FFT of simulated and real images, which belong to source and target distributions, respectively. Then it replaces the magnitude of the low-frequency spectrum of the simulated image with the real one. Finally, it obtains the output by taking the IFFT from the modified simulated image. (a) A synthetic ultrasound image, which is simulated using Field II and belongs to the source distribution. (b) A real ultrasound image, which belongs to the target distribution. (c) The output, which seems closer to the target distribution.

The sigmoid function was chosen as the activation function of the output layer, and the learning rate and batch size were 1×10^{-4} and 16, respectively. The AdamW [18], a variant of Adam [19], with a weight decay of 10^{-2} was exploited as the optimizer.

We used Dice similarity coefficient (DSC) for evaluating the segmentation performance, and the loss function was also defined based on this metric, which quantifies the area overlap between the ground truth and predicted masks:

$$DSC(S, \hat{S}) = \frac{2 \left| S \cap \hat{S} \right| + \varepsilon}{|S| + \left| \hat{S} \right| + \varepsilon} \tag{7}$$

where ε is a small number that prevents numerical instability for small masks. For each epoch of training or fine-tuning, the model weights were stored only when the validation loss had been improved, and finally, the best weights were used for testing. Experiments were implemented using the PyTorch package [20], and training was performed on an NVIDIA TITAN Xp GPU with 12 GB of memory.

III. RESULTS

To assess the effect of the FDA method for mitigating the domain shift problem, we conducted three different experiments. In the first experiment, labeled as the baseline, a network was trained merely using 20 training samples of the *in vivo* dataset for 200 epochs. In the next experiment,

first, the network was trained using 800 training samples of the simulated dataset without applying the FDA method for 150 epochs. Then it was fined-tined using 20 training samples of the *in vivo* dataset for 50 epochs. Finally, as the third experiment, we applied FDA to the simulated images and repeated the previous experiment. To apply the FDA method, for each training simulated image and at each iteration, the target image was randomly chosen from 40 images in the training and validation sets of the *in vivo* dataset. However, we used a fixed set of target images for applying the FDA method on the validation samples. Since the main purpose of using the validation set was to find and save the best model across different epochs, injecting randomness caused by choosing random target images was not desired.

Fig. 2 illustrates the DSC results for 123 test set images of the *in vivo* dataset for the aforementioned three experiments. As we expected, the baseline method led to the lowest DSC due to training on only 20 real ultrasound images without taking advantage of pre-training on simulated images. The second experiment achieved a better performance by pre-training on simulated data and using *in vivo* images for fine-tuning. However, it suffered from the domain shift problem, where there was a high discrepancy between the distribution of the simulated data and the *in vivo* data. The third experiment showed a 3.5% improvement in mean DSC, obtained by applying the FDA method.
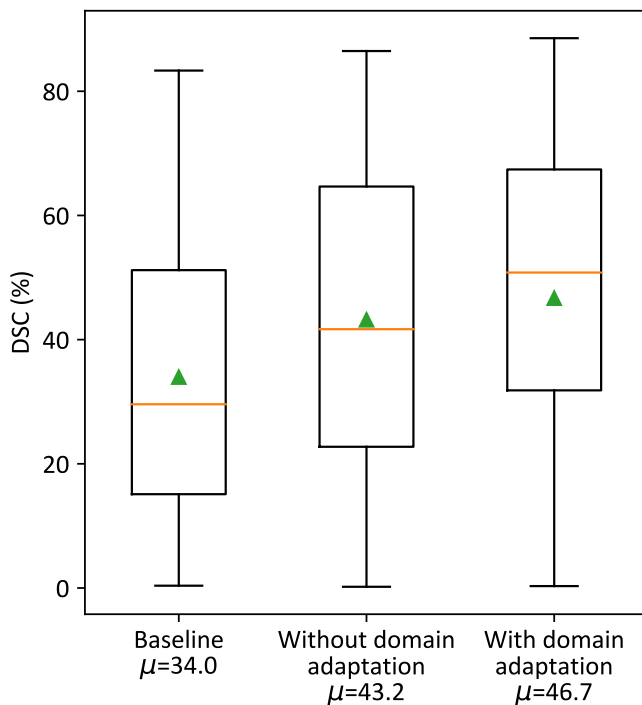
Fig. 2. Quantitative comparison of DSC over the *in vivo* test set. (Left) Training the network from scratch merely using *in vivo* images. (Middle) Use pre-trained weights obtained from training on simulated images without applying the FDA method. (Right) Use pre-trained weights obtained from training on simulated images with applying the FDA method. The triangle and horizontal line represent the mean and median, respectively.

IV. Conclusion

In conclusion, we claimed that important differences between simulated and real ultrasound data are low-frequency in nature. We employed the FDA method, which replaces the magnitude of the low-frequency spectrum of a synthetic image with a real one to tackle the issue of domain shift. For the first time, we exploited the FDA method in segmentation of ultrasound images, and more generally in ultrasound imaging. We demonstrated that applying this fast and simple method on simulated ultrasound data can improve the mean DSC as high as 3.5% compared to using simulated data without applying any domain adaptation method.

References

[1] M. Ghafoorian, A. Mehrtash, T. Kapur, N. Karssemeijer, E. Marchiori, M. Pesteie, C. R. Guttmann, F. E. de Leeuw, C. M. Tempany, B. van Ginneken, A. Fedorov, P. Abolmaesumi, B. Platel, and W. M. Wells, "Transfer learning for domain adaptation in MRI: Application in brain lesion segmentation," in *Lecture Notes in Computer Science*, vol. 10435 LNCS, 2017, pp. 516–524.

[2] Y. Chen, W. Li, C. Sakaridis, D. Dai, and L. Van Gool, "Domain Adaptive Faster R-CNN for Object Detection in the Wild," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2018, pp. 3339–3348.

[3] R. Gopalan, R. Li, and R. Chellappa, "Domain adaptation for object recognition: An unsupervised approach," in *Proceedings of the IEEE International Conference on Computer Vision*, 2011, pp. 999–1006.

[4] J. Tierney, A. Luchies, C. Khan, B. Byram, and M. Berger, "Accounting for domain shift in neural network ultrasound beamforming," in *IEEE International Ultrasonics Symposium, IUS*, vol. September, 2020.

[5] 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*, 2021.

[6] X. Ying, Y. Zhang, X. Wei, M. Yu, J. Zhu, J. Gao, Z. Liu, X. Li, and R. Yu, "MSDAN: Multi-Scale Self-Attention Unsupervised Domain Adaptation Network for Thyroid Ultrasound Images," in *IEEE International Conference on Bioinformatics and Biomedicine, BIBM*. IEEE, dec 2020, pp. 871–876.

[7] Q. Meng, J. Matthew, V. A. Zimmer, A. Gomez, D. F. Lloyd, D. Rueckert, and B. Kainz, "Mutual Information-Based Disentangled Neural Networks for Classifying Unseen Categories in Different Domains: Application to Fetal Ultrasound Imaging," *IEEE Transactions on Medical Imaging*, vol. 40, no. 2, pp. 722–734, 2021.

[8] L. Zhang, X. Wang, D. Yang, T. Sanford, S. Harmon, B. Turkbey, B. J. Wood, H. Roth, A. Myronenko, D. Xu, and Z. Xu, "Generalizing Deep Learning for Medical Image Segmentation to Unseen Domains via Deep Stacked Transformation," *IEEE Transactions on Medical Imaging*, vol. 39, no. 7, pp. 2531–2540, 2020.

[9] Y. Yang and S. Soatto, "FDA: Fourier domain adaptation for semantic segmentation," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2020, pp. 4084–4094.

[10] S. R. Richter, V. Vineet, S. Roth, and V. Koltun, "Playing for data: Ground truth from computer games," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 9906 LNCS, 2016, pp. 102–118.

[11] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele, "The Cityscapes Dataset for Semantic Urban Scene Understanding," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2016-Decem, 2016, pp. 3213–3223.

[12] J. A. Jensen, "FIELD: A program for simulating ultrasound systems," *Medical and Biological Engineering and Computing*, vol. 34, no. SUPPL. 1, pp. 351–352, 1996.

[13] J. A. Jensen and N. B. 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.

[14] C. Xia, J. Li, X. Chen, A. Zheng, and Y. Zhang, "What is and What is Not a Salient Object? Learning Salient Object Detector by Ensembling Linear Exemplar Regressors," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. Janua. IEEE, jul 2017, pp. 4399–4407.

[15] M. H. Yap, G. Pons, J. Martí, S. Ganau, M. Sentís, R. Zwiggelaar, A. K. Davison, and R. Martí, "Automated Breast Ultrasound Lesions Detection Using Convolutional Neural Networks," *IEEE Journal of Biomedical and Health Informatics*, vol. 22, no. 4, pp. 1218–1226, 2018.

[16] M. Frigo and S. G. Johnson, "FFTW: An adaptive software architecture for the FFT," in *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, vol. 3, 1998, pp. 1381–1384.

[17] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *Lecture Notes in Computer Science*, may 2015, vol. 9351, pp. 234–241.

[18] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," in *7th International Conference on Learning Representations, ICLR*, 2019.

[19] D. P. Kingma and J. L. Ba, "Adam: A method for stochastic optimization," in *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, 2015.

[20] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Köpf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, "PyTorch: An imperative style, high-performance deep learning library," in *Advances in Neural Information Processing Systems*, 2019.