Deep Learning Beamforming for Sub-Sampled Ultrasound Data

Walter Simson
Computer Aided Medical Procedures
Technische Universität München
Boltzmannstr. 3, 85748 Garching, Germany
walter.simson@tum.de

Magdalini Paschali
Computer Aided Medical Procedures
Technische Universität München
Boltzmannstr. 3, 85748 Garching, Germany

Nassir Navab
Computer Aided Medical Procedures
Technische Universität München
Boltzmannstr. 3, 85748 Garching, Germany
Johns Hopkins University
3400 North Charles Street
Baltimore, MD 21218, USA

Guillaume Zahnd
Computer Aided Medical Procedures
Technische Universität München
Boltzmannstr. 3, 85748 Garching, Germany

***Abstract*—In medical imaging tasks, such as cardiac imaging, ultrasound acquisition time is crucial, however traditional high-quality beamforming techniques are computationally expensive and their performance is hindered by sub-sampled data. To this end, we propose DeepFormer, a method to reconstruct high quality ultrasound images in real-time on sub-sampled raw data by performing an end-to-end deep learning-based reconstruction. Results on an in vivo dataset of 19 participants show that DeepFormer offers promising advantages over traditional processing of sub-sampled raw-ultrasound data and produces reconstructions that are both qualitatively and visually equivalent to fully-sampled DeepFormed images.**

***Index Terms*—Ultrasound, Deep Learning, Beamforming, Compressed Sensing**

I. Introduction

Ultrasound (US) imaging is a well established imaging modality that is used for a variety of applications due to its low cost, lack of ionizing radiation and ease of use. Beamforming is the process of generating and receiving sonic tissue responses and transforming them into images of the medium scanned which can be interpreted by the human eye. The response, also called raw scan line (2D), is transformed into a single radio frequency signal (RF, 1D), which makes up a column of the ultrasound image. To create a complete image of the target anatomy, an ensemble of neighboring scan lines is collected and arranged to make up a 2D-image.

Currently, beamforming algorithms can be classified into two groups, real-time, such as Delay and Sum [1] and offline methods such as Delay Multiply and Sum and Minimum Variance [2], [3]. The latter of the two classes offers higher image fidelity, a narrower main lobe and reduced side lobes at the expense of higher computational cost. This added computational cost prohibits its wide acceptance by the medical community, which has until now preferred the lower quality and higher imaging rates of real-time methods.

It is well established that the frame rate of ultrasound acquisitions depends on the number of scan lines and the imaging depth of the image. The acquisition time t_a can be generally defined by the equation:

$$t_a = (2 \times c_t \times d_i + t_d) \times n_{sl} \tag{1}$$

where c_t describes the speed of sound of the medium, d_i the imaging depth of the scanner, t_d the idle time, and n_{sl} the number of scan lines.

In applications, such as cardiac ultrasound, high temporal resolution is crucial to examine the anatomy during the entire cardiac cycle [4]. Furthermore, it has been shown that tracking the motion of the wall of the carotid artery can serve as strong indicator of coronary artery disease [5]. A higher temporal resolution allows for a more precise tracking of this motion. In order to increase the frame rate of ultrasound imaging and reduce the total acquisition time per frame, one could reduce the number of acquired scan lines n_{sl}. However, using traditional beamforming techniques on sub-sampled data would lead to significant decrease in the lateral resolution of the captured frame [6].

Deep Convolutional Neural Networks (DCNN) have been utilized recently for a plethora of applications in medical imaging, providing state-of-the-art end-to-end solutions to tasks such as segmentation [7], classification [8], retrieval [9] and image reconstruction [10]. Their strong learning capabilities enable them to capture the underlying structures of the data successfully, even in cases where their amount is limited [11], while they can operate in real-time due to their efficient inference functionality.

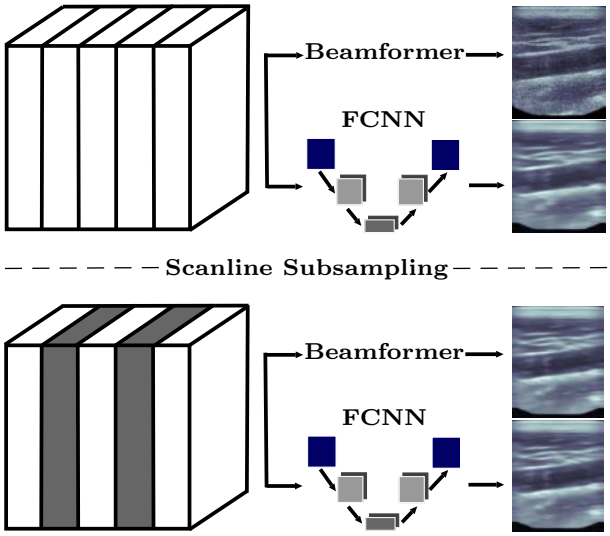

Fig. 1. Overview of the proposed framework, DeerFormer, and comparison with traditional beamforming.

Additionally, a variety of ultrasound reconstruction approaches utilizing deep learning have been proposed. Nair *et al.* [12] reconstruct ultrasound images from simulated raw ultrasound data using a fully convolutional neural network, Ortha *et al.* [4] reconstruct pre-delayed raw data for cardiac imaging applications, while approaches such as those presented in [13], [14] improve conventional beamforming by raw data pre-processing.

In this work, we propose an end-to-end framework, which learns a mapping from a sparse collection of raw data scan lines to a high quality Minimum Variance ultrasound image leveraging a state-of-the-art fully convolutional network. The proposed method, called *DeepFormer*, works towards enabling the reconstruction of high-quality ultrasound images from sub-sampled data, with significantly lower computational expense, while maintaining a high lateral resolution. The overview of our approach can be seen in Fig. 1. DeepFormer has been thoroughly trained and evaluated on an in vivo dataset, as will be further discussed in Section III-A.

II. Methods

A. Sub-Sampled DeepFormer

Traditionally, scan line beamforming is performed by activating a sub-aperture of a_s elements of an array of a_t total piezo elements, and recording the tissue response. Afterwards, the sub-aperture is shifted a set number of elements s to acquire the next neighboring scan line. Often, $s < 0.5 \times a_s$ leads to a spacial overlap of tissue through which a sonic signal is propagated.

The resulting mutual information between scan lines is not taken into account during the beamforming process, even though there is strong semantic correlation between the raw data of neighboring scan lines. However, a fully convolutional neural network can effectively capture and leverage this correlation and be trained on even sub-sampled raw data to interpolate skipped scan lines via inference of correlations in neighboring ones [10].

B. Network Architecture and Optimization

The efficient fully convolutional network architecture QuickNat [15] is employed for the reconstruction, due to its improved gradient flow provided by the long- and short-term skip connections between the encoder and decoder blocks. Batch normalization and dropout are additionally utilized to decrease the risk of overfitting. In order to improve the image quality, a hybrid loss was employed using a combination of the ℓ_1 loss and a differentiable formulation of the Multi-Scale Structural Similarity Imaging Metric (MS-SSIM) proposed in [16]. For an image P consisting of N pixels p, the ℓ_1 loss function is calculated by:

$$\mathcal{L}_1(P) = \frac{1}{N} \sum_{p \in P} |x(p) - y(p)|,$$

with $x(p)$ and $y(p)$ being the pixel intensity values in the reconstructed image and the ground truth respectively. The SSIM loss is calculated in windows of the images, where the measure between two windows x and y is given by:

$$\mathcal{L}_{\text{SSIM}} = 1 - \frac{2\mu_x\mu_y + c_1}{\mu_x^2 + \mu_y^2 + c_1} \cdot \frac{2\sigma_{xy} + c_2}{\sigma_x^2 + \sigma_y^2 + c_2} = 1 - l(p) \cdot cs(p).$$

The variables μ_x and μ_y are the means of windows x and y, σ_x^2 and σ_y^2 are their respective variances, σ_{xy} is their covariance. The constants c_1, c_2 ensure the numerical stability of the division operations. The MS-SSIM loss is a multi-scale version of the SSIM loss, calculated in a pyramid of M levels, formulated as:

$$\mathcal{L}_{\text{MS-SSIM}} = 1 - l_M(p) \cdot \prod_{j=1}^{M} cs_j(p).$$

The composite loss used for training DeepFormer is:

$$\mathcal{L}_{\text{DF}} = \alpha \mathcal{L}_{\text{MS-SSIM}} + (1 - \alpha)\mathcal{L}_1,$$

where α is set to 0.84, as suggested by Zhao *et al.* in [17].

III. Experimental Setup

A. Data acquisition

Ultrasound raw data acquisition was performed with the publicly available GPU-based beamforming software SUPRA [18] with a cQuest Cicada scanner (Cephasonics, CA, USA) equipped with a 128 element linear transducer (CPLA12876, 7 MHz). The imaging depth was 40 [mm], the pulse frequency 7 MHz with single focus at 20 mm and a dynamic range of 50 dB. Ground truth images were generated offline with a Minimum Variance beamforming method [3] also implemented in SUPRA.

TABLE I
COMPARISON OF BEAMFORMED AND DEEPFORMED IMAGES. FOR THIS EVALUATION DEEPFORMER SUB-SAMPLED WAS COMPARED TO DEEPFORMER FULLY-SAMPLED AND BEAMFORMED SUB-SAMPLED WAS COMPARED TO BEAMFORMED FULLY-SAMPLED.

	SSIM ± SD	PSNR ± SD
Beamforming	0.7580 ± 0.0138	30.7719 ± 0.5098
DeepFormer	0.9452 ± 0.0058	34.6326 ± 0.9744

B. In Vivo Data

With the above mentioned parameters, a corpus of in vivo US data was generated, comprised of data from 19 healthy participants (age: 32 ± 13 y.o.). All scans were collected on both sides (left, right) of the body and in both the longitudinal and transverse orientation of the target anatomies, namely common carotid artery, thyroid gland, bicep muscle fibers and forearm. The scan duration was at least 10 seconds leading to a minimum number of 19 frames per scan. Every scan consists of an ultrasound sweep over the anatomy, in order to introduce variability in the imaging plane. Acquisitions for the corpus were performed free-hand. A patient-level split was introduced to separate the data into a training set consisting of 14 subjects and an independent test set of 5 subjects.

C. Network Training

DeepFormer has been implemented in PyTorch [19] and trained until convergence on an NVIDIA Titan Xp GPU using Stochastic Gradient Descent with momentum 0.9 and learning rate set to 0.01, as suggested in [15].

IV. RESULTS AND DISCUSSION

A. Beamforming and DeepForming of sub-sampled data

Table I showcases the difference of image quality between sub-sampled images and their fully-sampled counterparts using conventional beamforming and DeepFormer respectivly. The sub-sampled DeepFormer achieved an SSIM score that was 0.2 higher and displayed a 5 dB improvement over conventionally sub-sampled Minimum Variance beamforming. The consistency in DeepForming image quality not only on full but also on sub-sampled data displays the strength of the approach in ultrasound reconstruction and highlights its suitability to be applied when a decreased number of scan lines is acquired.

B. Effect of different loss functions

Choosing the proper loss function for a given reconstruction can have a significant impact on the resulting reconstructed images and is strongly dependent on the modality at hand.

Traditionally the ℓ_2 loss function is employed for reconstruction tasks, however Zhao *et al.* [17] reported extensive results showcasing its poor convergence properties and stronger sensitivity to outliers in the training data. On the contrary, their experiments highlighted the superiority of the ℓ_1 loss function, which we preferred to utilize in this paper. Exhaustive ablative testing was performed to compare the efficacy of composite reconstruction losses as described by Zhao *et al.* [17], the results of which can be seen in Table II. Our evaluation metrics, namely SSIM and PSNR, showed that the composite $\mathcal{L}_1 + \mathcal{L}_{\text{MS-SSIM}}$ loss function outperformed other available options.

TABLE II
ABLATIVE TESTING OF THE DIFFERENT DEEPFORMER LOSS FUNCTIONS.

	SSIM ± SD	PSNR ± SD
$\mathcal{L}_1$ Sub-sampling	0.5315 ± 0.0201	23.9067 ± 0.6419
$\mathcal{L}_{\text{MS-SSIM}}$ Sub-sampling	0.5174 ± 0.0199	21.7697 ± 0.8644
$\mathcal{L}_1 + \mathcal{L}_{\text{MS-SSIM}}$ Full	0.5554 ± 0.0227	26.5537 ± 0.8060
$\mathcal{L}_1 + \mathcal{L}_{\text{MS-SSIM}}$ Sub-sampling	0.5550 ± 0.0228	26.3224 ± 0.8383

Individually $\mathcal{L}_1$ and $\mathcal{L}_{\text{MS-SSIM}}$ display lower performance, which can be attributed to the lack of contrast in images trained on $\mathcal{L}_1$ and the sensitivity of $\mathcal{L}_{\text{MS-SSIM}}$ to changes in local structures, which can be readily found in the ground truth [17].

Furthermore, Table II highlights the fact that the reconstruction difference in image quality of both fully- and sub-sampled DeepFormed images is infinitesimal small (0.0004 SSIM, 0.2313 PSNR) and that the effect of excluding one half of the input data before reconstruction is negligible.

C. Quality of the reconstructed images

Just as beamformed images display characteristic properties that result from the reconstruction process, DeepFormed images offer distinct features that stem from the reconstruction pipeline. As can be seen in Fig. 2, all major structures in the ultrasound frame were successfully reconstructed. Whereas traditional scan lines are beamformed independently from one another, leading to an anisotropic image, DeepFormed images are generated based on all captured raw data, leading to higher coherence between neighboring scan lines and an overall smoother lateral image texture.

Speckle, which results from the constructive and destructive interference of propagating waves, contributes to the characteristic texture of ultrasound images. However, the convolutional filters and pooling layers of DeepFormer discard this high frequency information in exchange for globally stable features, leading to an even smoother resulting image.

Moreover, the utilized composite loss has been proposed for applications involving smoothing and demosaicing of natural images, where the ground truth does not suffer from significant noise. In the case of DeepFormer, the ground truth itself contains noise, which contributes to the smoothing of the resulting reconstruction and values of SSIM, without however, decreasing the usability of the information contained in the image.

Ultimately, visual consistency between the fully- and sub-sampled DeepFormed images displays DeepFormer's capability to successfully perform reconstruction with limited data.

V. CONCLUSION

In this work, DeepFormer was proposed as a highly effective method to reconstruct ultrasound images from sub-sampled

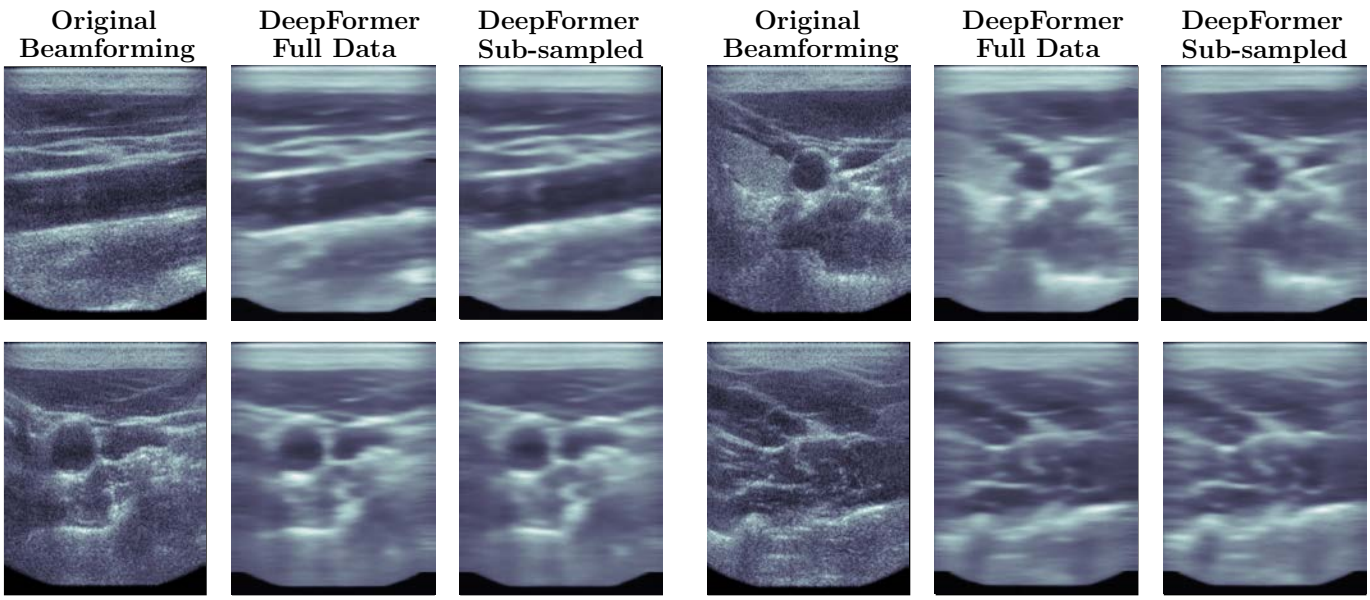

Fig. 2. Qualitative comparison between the original beamformed images using the Minimum Variance Method and the reconstructed images from DeepFormer, utilizing all the scan lines, or a subset.

raw data. A composite loss function, consisting of $\mathcal{L}_1$ and $\mathcal{L}_{\text{MS-SSSIM}}$ was able to achieve the highest SSIM and PSNR scores on an in vivo dataset of 19 participants on 4 different anatomies. Our results showcase that sub-sampled ultrasound acquisition using DeepFormer can allow for accelerated acquisitions with comparable image quality. DeepForming is applicable in a plethora of medical applications, such as cardiac imaging, where acquisition speed and lower processing time are critical. Future work includes exploring the performance of DeepFormer on different levels of sub-sampling and evaluating it in a more extensive clinical study.

REFERENCES

[1] K. Thomenius, "Evolution of ultrasound beamformers," in *1996 IEEE Ultrasonics Symposium. Proceedings*, IEEE, 1996.

[2] G. Matrone, A. S. Savoia, G. Caliano, and G. Magenes, "The delay multiply and sum beamforming algorithm in ultrasound B-mode medical imaging," *IEEE Transactions on Medical Imaging*, vol. 34, no. 4, pp. 940–949, 2015.

[3] J. Li and P. Stoica, *Robust adaptive beamforming*. John Wiley & Sons, Hoboken, NJ, 2006.

[4] O. Senouf, S. Vedula, G. Zurakhov, A. Bronstein, M. Zibulevsky, O. Michailovich, D. Adam, and D. Blondheim, "High frame-rate cardiac ultrasound imaging with deep learning," in *Medical Image Computing and Computer Assisted Intervention*, pp. 126–134, Springer International Publishing, 2018.

[5] G. Zahnd, K. Saito, K. Nagatsuka, Y. Otake, and Y. Sato, "Dynamic block matching to assess the longitudinal component of the dense motion field of the carotid artery wall in b-mode ultrasound sequences association with coronary artery disease," *Medical Physics*, 2018.

[6] T. Bjastad, S. Aase, and H. G. Torp, "The impact of aberration on high frame rate cardiac b-mode imaging," *IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control*, vol. 54, 2007.

[7] M. A. Degel, N. Navab, and S. Albarqouni, "Domain and geometry agnostic CNNs for left atrium segmentation in 3D ultrasound," in *Medical Image Computing and Computer Assisted Intervention*, pp. 630–637, Springer International Publishing, 2018.

[8] F. Navarro, S. Conjeti, F. Tombari, and N. Navab, "Webly supervised learning for skin lesion classification," in *Medical Image Computing and Computer Assisted Intervention*, pp. 398–406, Springer International Publishing, 2018.

[9] S. Conjeti, M. Paschali, A. Katouzian, and N. Navab, "Deep multiple instance hashing for scalable medical image retrieval," in *Medical Image Computing and Computer-Assisted Intervention*, pp. 550–558, Springer International Publishing, 2017.

[10] D. Ravì, A. B. Szczotka, D. I. Shakir, S. P. Pereira, and T. Vercauteren, "Effective deep learning training for single-image super-resolution in endomicroscopy exploiting video-registration-based reconstruction," *International Journal of Computer Assisted Radiology and Surgery*, vol. 13, no. 6, pp. 917–924, 2018.

[11] B. Kelly, T. P. Matthews, and M. A. Anastasio, "Deep learning-guided image reconstruction from incomplete data," 2017. arXiv: 1709.00584.

[12] A. A. Nair, T. D. Tran, A. Reiter, and M. A. L. Bell, "A deep learning based alternative to beamforming ultrasound images," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing, Calgary, AB, Canada, April 15-20, 2018*, pp. 3359–3363, 2018.

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

[14] Y. H. Yoon and J. C. Ye, "Deep learning for accelerated ultrasound imaging," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing, Calgary, AB, Canada, April 15-20, 2018*, pp. 6673–6676, 2018.

[15] A. G. Roy, S. Conjeti, N. Navab, and C. Wachinger, "Quicknat: Segmenting MRI neuroanatomy in 20 seconds," 2018. arXiv: 1801.04161.

[16] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

[17] H. Zhao, O. Gallo, I. Frosio, and J. Kautz, "Loss functions for neural networks for image processing," 2015. arXiv: 1511.08861.

[18] R. Göbl, N. Navab, and C. Hennersperger, "Supra: open-source software-defined ultrasound processing for real-time applications," *International Journal of Computer Assisted Radiology and Surgery*, vol. 13, no. 6, pp. 759–767, 2018.

[19] A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, and A. Lerer, "Automatic differentiation in pytorch," in *Neural Information Processing Systems 2017 – Autodiff Workshop*, 2017.