High-Quality Plane Wave Compounding using Convolutional Neural Networks

Maxime Gasse, Fabien Millioz, Emmanuel Roux, Damien Garcia, Hervé Liebgott, Denis Friboulet

High-Quality Plane Wave Compounding using Convolutional Neural Networks

Maxime Gasse, Fabien Millioz, Emmanuel Roux,
Damien Garcia, Hervé Liebgott and Denis Friboulet

Abstract—Single plane wave (PW) imaging produces ultrasound (US) images of poor quality at high frame rates (ultrafast). High-quality PW imaging usually relies on the coherent compounding of several successive steered emissions (typically more than ten), which in turn results in a decreased frame rate. We propose a new strategy to reduce the number of emitted PWs by learning a compounding operation from data, i.e. by training a convolutional neural network (CNN) to reconstruct high quality images using a small number of transmissions. We present experimental evidence that this approach is promising, as we were able to produce high-quality images from only 3 PWs, competing in terms of contrast ratio and lateral resolution with the standard compounding of 31 PWs (10x speed-up factor).

I. INTRODUCTION

Ultrasound images obtained from single PW transmissions exhibit lower quality (in terms of resolution and contrast) than those obtained from classical focused transmissions. To overcome this limitation, a standard approach consists in performing spatial coherent compounding of multiple RF images resulting from successive transmissions [1]. The cost of such an approach is a substantially decreased frame rate, since many PW transmissions are involved in this process. In this work, we formulate PW compounding as a supervised learning problem. In particular, we demonstrate experimentally the ability for a convolutional neural network (CNN) to reconstruct high-quality US images from a small number of PW acquisitions.

II. METHODOLOGY

We consider m PW acquisitions from a linear US probe, each yielding q RF signals of length p. The compounding operation then takes m beamformed RF images as an input, that is, $\mathbf{x} \in \mathbb{R}^{m \times p \times q}$, and produces one compounded RF image as an output, that is, $\mathbf{h}(\mathbf{x}) \in \mathbb{R}^{p \times q}$. We adopt a supervised learning setting, and seek for the optimal compounding operation $\mathbf{h}^\star$ with respect to a target image $\mathbf{y} \in \mathbb{R}^{p \times q}$. Finding $\mathbf{h}^\star$ then amounts to solving the following risk-minimization problem:

$$\mathbf{h}^\star = \arg\min_{\mathbf{h}} \, \mathbb{E}_{\mathbf{x},\mathbf{y}} \left[L(\mathbf{h}(\mathbf{x}), \mathbf{y}) \right], \qquad (1)$$

with $\mathbf{x} \in \mathcal{X}$ a point in the input space, $\mathbf{y} \in \mathcal{Y}$ a point in the output space, $\mathbf{h}$ a $\mathcal{X} \mapsto \mathcal{Y}$ mapping, and L a loss function. A common approach to solve this problem is to consider a limited set of mappings $\mathbf{h} \in \mathcal{H}$ (i.e., a model), and a limited set of input / output pairs $(\mathbf{x}, \mathbf{y}) \in \mathcal{D}$ (i.e., a dataset) sampled from the joint distribution $p(\mathbf{x}, \mathbf{y})$, and then perform empirical loss minimization:

$$\hat{\mathbf{h}}^\star = \arg\min_{\mathbf{h} \in \mathcal{H}} \sum_{(\mathbf{x},\mathbf{y})}^{\mathcal{D}} L(\mathbf{h}(\mathbf{x}), \mathbf{y}). \qquad (2)$$

Solving (2) is equivalent to solving (1) under two assumptions: 1) the dataset $\mathcal{D}$ is representative of the joint distribution $p(\mathbf{x}, \mathbf{y})$; and 2) the model $\mathcal{H}$ contains the optimal mapping $\mathbf{h}^\star$. Since none of these assumptions can be measured in practice, a compromise must be made on the model complexity with respect to the number of available training samples in order to avoid the under/over-fitting problem. In this work, our mapping space $\mathcal{H}$ consisted in a fixed CNN structure with parameter space Θ, therefore the search for $\mathbf{h}^\star$ boils down to a search for the optimal parameters $\theta^\star$. In order to reconstruct high-quality images, we chose the reference $\mathbf{y}$ as the image obtained from the standard compounding of $n \gg m$ PW acquisitions. The main justification for this approach is that there may be useful information from $\mathbf{x}$ for reconstructing a high-quality image $\mathbf{y}$ which is not exploited by standard compounding, but may be learned from data with an adequate model.

III. EXPERIMENTAL SETUP

A. Dataset acquisition

A linear-array probe of 128 elements with size 7×0.283mm and kerf 0.025mm (ATL L7-4 38mm, bandwidth 4-7 MHz, transmitted frequency 5.2 MHz) was interfaced with a Verasonics system research scanner (Vantage 256™) to perform the acquisitions. For each PW transmission, the acquired raw RF signals were sampled at 20 MHz and beamformed using the method from Lu [2], resulting in RF beamformed images with resolution $p \times q = 1332 \times 128$ and a corresponding spatial span of size of about 5×4 cm. Each acquisition was performed using 31 steered PW spanning $\pm 15°$ in one degree steps. Reference images $\mathbf{y}$ were obtained from the standard compounding of all $n = 31$ PWs while the input $\mathbf{x}$ was made up of only a small subset of $m = 3$ PWs, corresponding to steering angles $(-15°, 0°, +15°)$. The dataset used in the experiments corresponds to a total of 7000 $(\mathbf{x}, \mathbf{y})$ samples representative of US images. Specifically, 5000 acquisitions were performed of in-vivo tissues from three healthy subjects (carotid, thyroid and liver regions) and 2000 images of a Gammex phantom (410 SCG). The samples were obtained in batches of 250 images by continuously moving the probe around the target area while acquiring images at a rate of 50 fps. From the 7000 samples, 6000 were used for training the CNN, i.e. for solving (2). The remaining 1000 samples were used for testing, i.e. to evaluate the quality of the CNN-based compounding.

B. CNN architecture and training

The model we employed was a 2D fully-convolutional neural network (no spatial pooling) with 4 hidden layers and 4-pieces maxout activation units [3], whose architecture is described in Table I. Maxout units are piecewise-linear convex functions which can approximate many popular activation functions and most often outperform these, yet at a higher computational cost [4]. The absence of spatial pooling ensures that spatial information is preserved at the same resolution throughout the network, which may be desirable for preserving phase in RF signals. Also, with such a model we assumed the compounding operation to be spatially invariant, with a resulting receptive field (input patch size for each output
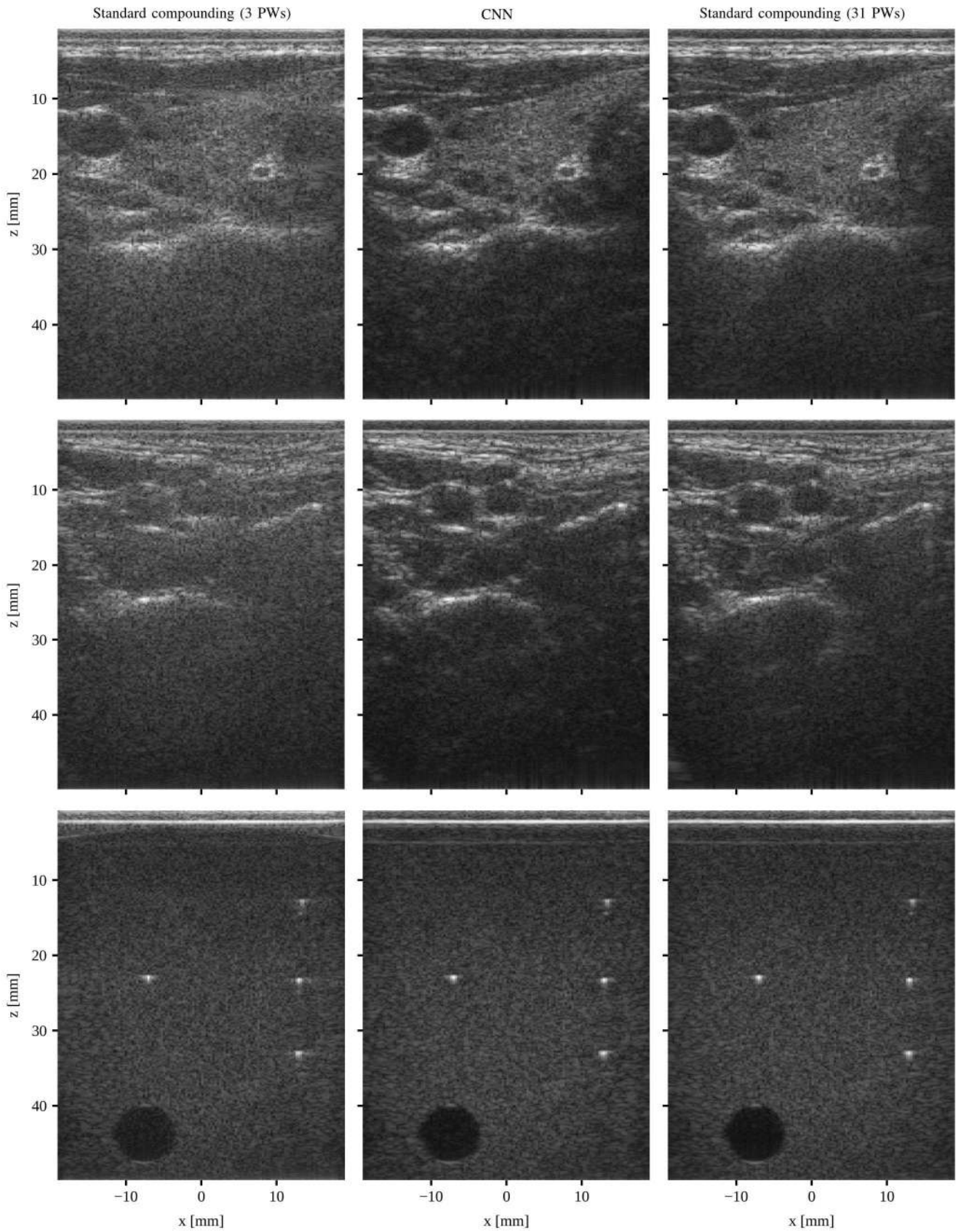
Fig. 1. B-mode images obtained using our CNN model and standard compounding. Imaged media are respectively, from top to bottom: in-vivo tissues from the thyroid region; in-vivo tissues from the carotid region; and in-vitro tissues from the Gammex phantom. All images were normalized within $[0, 1]$ and gamma-compressed using $\gamma = 0.3$, with no dynamic range clipping.

TABLE I
CNN ARCHITECTURE.

layer size (depth × height × width)	kernel size (height × width)	nb. of kernels	activation
$m \times p \times q$			
$64 \times p \times q$	9×3	256	maxout 4
$32 \times p \times q$	17×5	128	maxout 4
$16 \times p \times q$	33×9	64	maxout 4
$8 \times p \times q$	65×17	32	maxout 4
$1 \times p \times q$	1×1	4	maxout 4

pixel) of 121×31 (about 0.5×1 cm). The model was implemented using the Theano library [5] and trained via stochastic gradient descent with the Adam optimizer [6] on a Nvidia GPU (Geforce GTX 1080 Ti) to minimize the expected L_2 loss, resulting in training times of about two days.

IV. RESULTS

On representative test samples, the images produced by the CNN (Fig. 1, middle column) are visually very close to the target images (Fig. 1, right column). In particular, it can be observed how the CNN improved the contrast and enhanced anatomical structures as compared to the images obtained from the standard compounding of the same 3 PWs (Fig. 1, left column). We report two quantitative evaluation measures to assess this improvement: the contrast ratio (CR) as defined in [7], and the lateral resolution (LR), i.e. the full width at -6dB of the point spread function, taken from test samples acquired on the Gammex phantom. As a reference, the CNN is compared in terms of CR and LR to the standard coherent compounding of an increasing number of PWs, i.e. 1 to 31 PWs spread uniformly within a $\pm 15°$ sector (Fig. 2). The evolution of CR and LR is consistent with the behavior observed for uniform sector spanning by Zhang et al. [8]: LR quickly improves with 2 and 3 PWs while CR degrades, and then both indicators tend to stabilize to an optimal value as the number of PWs increases. Interestingly, while exploiting only 3 PWs the CNN resulted in a CR equivalent to that of the standard compounding of about 20 PWs, and a LR corresponding to the standard compounding of 31 PWs.

V. CONCLUSION

A methodology for learning a PW compounding operation from data with supervised learning was presented, and experimental evidence was provided that a CNN model is able to exploit information from separate PW acquisitions more efficiently than standard compounding, resulting in a better trade-off between image quality and frame rate. Note that the proposed network architecture was not optimized in any principled way, and therefore should only be seen as a proof-of-concept rather than a definitive solution to the compounding problem. In our experiments the most critical part of the model seemed to be its resulting receptive field, the larger the better, which explains the increased kernel size in the last layers.

CNN-based US image reconstruction thus appears as a promising approach for improving ultrafast acquisition, and many future research directions can be developed to further
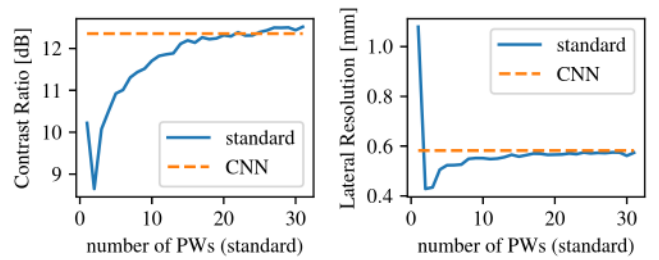

Fig. 2. Contrast ratio and lateral resolution reached by our CNN model with 3 PWs ($-15°, 0°, +15°$), compared to the standard compounding of m PWs spanning uniformly a $\pm 15°$ sector (uniform scheme). Contrast ratio is measured on a B-mode image (gamma-compression, $\gamma = 0.3$) using an anechoic region, while lateral resolution is measured on the envelope image of 0.1mm nylon fibers. Both measures were obtained from data acquired from the Gammex phantom.

investigate this approach, such as the compounding of diverging waves, the integration of the beamforming in the learning process or the exploitation of temporal information to further accelerate the image formation.

REFERENCES

[1] G. Montaldo, M. Tanter, J. Bercoff, N. Benech, and M. Fink, "Coherent plane-wave compounding for very high frame rate ultrasonography and transient elastography," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 56, no. 3, pp. 489–506, March 2009.

[2] J.-Y. Lu, "Experimental study of high frame rate imaging with limited diffraction beams," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 45, no. 1, pp. 84–97, Jan 1998.

[3] I. J. Goodfellow, D. Warde-Farley, M. Mirza, A. C. Courville, and Y. Bengio, "Maxout networks," in *Proceedings of the 30th International Conference on Machine Learning, ICML 2013, Atlanta, GA, USA, 16-21 June 2013*, 2013, pp. 1319–1327.

[4] P. Swietojanski, J. Li, and J.-T. Huang, "Investigation of maxout networks for speech recognition." in *ICASSP*. IEEE, 2014, pp. 7649–7653.

[5] Theano Development Team, "Theano: A Python framework for fast computation of mathematical expressions." *CoRR*, vol. abs/1605.02688, 2016.

[6] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *The International Conference on Learning Representations, ICLR 2015, San Diego*, 2015.

[7] M. van Wijk and J. Thijssen, "Performance testing of medical ultrasound equipment: fundamental vs. harmonic mode," *Ultrasonics*, vol. 40, no. 1, pp. 585–591, 2002.

[8] M. Zhang, F. Varray, A. Besson, R. E. Carrillo, M. Viallon, D. Garcia, J. P. Thiran, D. Friboulet, H. Liebgott, and O. Bernard, "Extension of fourier-based techniques for ultrafast imaging in ultrasound with diverging waves," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 63, no. 12, pp. 2125–2137, Dec 2016.