A Deep Learning Approach to Ultrasound Image Recovery

Dimitris Perdios*, Adrien Besson*, Marcel Arditi*, and Jean-Philippe Thiran*†
*Signal Processing Laboratory (LTS5), Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland
†Department of Radiology, University Hospital Center (CHUV) and University of Lausanne (UNIL), Lausanne, Switzerland

***Abstract*—Based on the success of deep neural networks for image recovery, we propose a new paradigm for the compression and decompression of ultrasound (US) signals which relies on stacked denoising autoencoders. The first layer of the network is used to compress the signals and the remaining layers perform the reconstruction. We train the network on simulated US signals and evaluate its quality on images of the publicly available PICMUS dataset. We demonstrate that such a simple architecture outperforms state-of-the art methods, based on the compressed sensing framework, both in terms of image quality and computational complexity.**

***Index Terms*—Compressed sensing, deep learning, ultrafast ultrasound imaging, stacked denoising autoencoders.**

I. Introduction

In recent years, the problem of recovering ultrasound (US) signals from undersampled measurements have raised a growing interest due to the emergence of the compressed sensing (CS) framework [1]. Formally, let us consider a US signal as $\boldsymbol{m} \in \mathbb{R}^N$ where N denotes the number of time samples. US signal recovery amounts to retrieving $\boldsymbol{m}$ from $\boldsymbol{y} \in \mathbb{R}^M$ where $M < N$ and $\boldsymbol{y} = \mathcal{C}(\boldsymbol{m})$ with $\mathcal{C} : \mathbb{R}^N \to \mathbb{R}^M$ a compression operator.

In this context, the CS framework demonstrates that a perfect reconstruction of $\boldsymbol{m}$ from $\boldsymbol{y}$ is possible, providing that $\boldsymbol{y}$ is k-sparse, i.e. that $\|\boldsymbol{y}\|_0 = k$ where the ℓ_0-norm accounts for the number of non-zero coefficients, and that $\mathcal{C}$ is a linear operator $\mathsf{C} \in \mathbb{R}^{M \times N}$ which satisfies the restricted isometry property (RIP) of order $2k$ [2]. CS also provides a way to recover $\boldsymbol{m}$ by solving the following optimization problem:

$$\min_{\hat{\boldsymbol{m}} \in \mathbb{R}^N} \|\hat{\boldsymbol{m}}\|_1 \;\text{ subject to }\; \|\boldsymbol{y} - \mathsf{C}\hat{\boldsymbol{m}}\|_2 < \epsilon, \tag{1}$$

where $\epsilon \in \mathbb{R}_+$.

In the context of US imaging, Liebgott *et al.* [3] have shown that the wave-atom frame is a sparsity model particularly suited to US signal recovery. More recently, Besson *et al.* [4] have explored sparsity of US signals in a convolutional dictionary composed of shifted US pulses. Many researchers have also used the CS framework in the image reconstruction process exploiting structural properties of radio-frequency (RF) images or tissue reflectivity functions while undersampling the US signals directly [5], [6]. In a similar effort, researchers have proposed methods where the RF images are undersampled. Lorintiu *et al.* [7] have used a line-wise undersampling and a learned dictionary for the sparsity prior. Quinsac *et al.* [8] have exploited a similar undersampling approach but with a sparsity prior in the Fourier domain. Chen *et al.* [9] have exploited CS for deconvolution purpose.

The CS framework suffers from several major drawbacks that severely limit its applicability in US imaging. First, checking that the matrix C satisfies the RIP is a NP-hard problem. It has been demonstrated that random Gaussian or Bernoulli matrices satisfy the RIP with high probability, providing that M is sufficiently high. But constraints in the US signal acquisition process make the design of such matrices rather impossible in practice. Moreover, sparsity of US signals is very hard to obtain due to statistical dependencies inside specific regions (speckle) and wide variability between different regions inside an image. Finally, the resolution of Problem (1) involves the use of convex optimization algorithms that require hundreds of iterations to converge and a very precise fine-tuning of hyper-parameters, which prevent their use in real-time scenarios.

In this paper, we propose to exploit stacked denoising autoencoders (SDA), successfully applied to recovery of structured signals [10], for the recovery of US images. To do so, the compression is considered to be the first layer of the proposed architecture. The hidden and output layers are used for the reconstruction. We explore both a linear measurement case where the compression matrix is not learned (SDA–CNL) and a non-linear measurement case where the compression is learned (SDA–CL). We show that a 4-layer SDA–CL outperforms a state-of-the-art CS algorithm in terms of both quality and reconstruction time, without the need to tune any hyper-parameter.

The remainder of the paper is organized as follows. Section II details the proposed networks, their trainings as well as the synthetic training set generation. Section III describes the experiments and performance of the networks. Concluding remarks are given in Section IV.

II. Stacked Denoising Autoencoders for Ultrasound Image Recovery

A. Proposed Architectures

Both proposed architectures, i.e. SDA–CNL and SDA–CL, are composed of 4 fully-connected layers as described on Fig. 1. The compressed measurements $\boldsymbol{y} = \mathcal{C}(\boldsymbol{m})$ are the output of the first layer. In the case of SDA–CNL, $\mathcal{C}(\boldsymbol{m}) = \Phi\boldsymbol{m}$, where $\Phi \in \mathbb{R}^{M \times N}$ is a random Gaussian matrix, not learned during training. In the case of SDA–CL, $\mathcal{C}(\boldsymbol{m}) = \mathcal{T}(\mathsf{W}_{in}\boldsymbol{m} + \boldsymbol{b}_{in})$, where $\mathsf{W}_{in} \in \mathbb{R}^{M \times N}$ is a weight

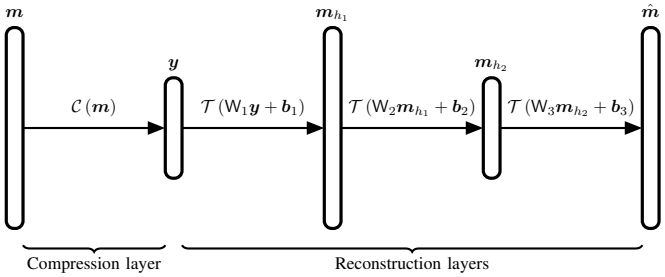

Fig. 1. Proposed 4-layer architecture.

matrix, $\boldsymbol{b}_{in} \in \mathbb{R}^M$ is a bias and $\mathcal{T}(\,.\,)$ is a non-linear function. The reconstruction layers are composed of two hidden layers $\boldsymbol{m}_{h_1} = \mathcal{T}(\mathsf{W}_1\boldsymbol{y} + \boldsymbol{b}_1)$ and $\boldsymbol{m}_{h_2} = \mathcal{T}(\mathsf{W}_2\boldsymbol{m}_{h_2} + \boldsymbol{b}_2)$, and an output layer $\hat{\boldsymbol{m}} = \mathcal{T}(\mathsf{W}_3\boldsymbol{m}_{h_2} + \boldsymbol{b}_3)$, where $\mathsf{W}_1 \in \mathbb{R}^{N\times M}$, $\mathsf{W}_2 \in \mathbb{R}^{M\times N}$, $\mathsf{W}_3 \in \mathbb{R}^{N\times M}$, $\boldsymbol{b}_1 \in \mathbb{R}^N$, $\boldsymbol{b}_2 \in \mathbb{R}^M$ and $\boldsymbol{b}_3 \in \mathbb{R}^N$. Due to the zero-mean centering of US signals, we choose the non-linearity function $\mathcal{T}(\,.\,)$ to be the hyperbolic tangent function.

It is clear that $\hat{\boldsymbol{m}}$ has a relatively high number of degrees of freedom corresponding to the weight matrices and bias vectors, which will be learned during the training phase.

Once trained, the first layer is used to compress each of the element-raw-data signals independently and the remaining layers are used for the recovery of these signals. Both the compression and the recovery operations can therefore be performed in parallel for a set of element-raw-data signals acquired by a US probe. Following the decompression step, the RF image is retrieved using any state-of-the-art US image reconstruction algorithm.

B. Training of the Proposed Networks

We consider the configuration of the plane-wave imaging challenge in medical ultrasound (PICMUS) [11], which is summarized in Table I. It can be seen that the sampling frequency of the element raw-data is around 4 times the center frequency, hence extremely close to the Nyquist frequency of the received US signals, considering the bandwidth of the transducer elements. The number of measurement samples N is fixed to 1024 in order to fit to typical sizes of deep neural networks.

TABLE I
PLANE WAVE IMAGING CONFIGURATION PARAMETERS

Parameter	**L11-4v**
Element number	128
Pitch	300 µm
Center frequency	5.133 MHz
Bandwidth	67 %
Element width	0.27 mm
Transmit frequency	5.208 MHz
Excitation	2.5 cycles
Sampling frequency	20.832 MHz
Sample number	1024

The training set is simulated using the open-source k-Wave toolbox [12] on a configuration mimicking the acquisition system described in Table I. The attenuation coefficient is set to 0.5 dB MHz^{-1} cm^{-1} and the simulation accounts for the element directivity. The insonified medium is simulated from a randomly generated phantom containing:

- a fully diffusive background which defines the echogenicity reference;
- one to three circular inclusions of variable radius and echogenicity;
- zero to five point reflectors.

The number of inclusions and point reflectors as well as their positions within the field of view are randomly determined. Each inclusion has a radius drawn between 5 and 50 wavelengths and is either anechoic (80 %), between −6 dB and 6 dB (15 %) or between 10 dB and 20 dB (5 %). The transmit scheme used to insonify the phantom is plane wave with normal incidence. Each synthetic acquisition is composed of 128 element-raw-data, mimicking the US signals received at each transducer element of the probe described in Table I, for the considered transmit scheme.

20 000 synthetic acquisitions are generated using the above procedure. This corresponds to 2.5 M element-raw-data signals and enables us to handle 2.1 M parameters, i.e. the number of weights of the SDA in the case of an undersampling ratio M/N of 0.5.

The networks are implemented[1] using the Python API of TensorFlow. Both networks are trained for each considered undersampling ratio M/N, namely from 0.05 to 0.5. Time gain compensation is applied to the element-raw-data to compensate for the attenuation. Every acquisition is normalized between −1 and 1 to fit the range of the non-linearity used in both SDAs. The training is performed on a NVIDIA GeForce GTX 1080 Ti with a learning rate of 0.001 over 20 epochs using mini-batch learning and a batch size of 4096. The trainable weights are initialized with the Xavier initialization [13] and the biases to zero. We use Adam optimizer and the ℓ_2-loss as the loss function.

III. RESULTS

A. Experimental Settings

The proposed architectures are tested on one numerical image, three *in vitro* images and two *in vivo* images provided by the PICMUS dataset[2] [11] acquired using the configuration defined in Table I.

Three different approaches are compared, namely SDA–CL, SDA–CNL, both described in Section II, and a standard CS reconstruction based on a sparsity prior in a convolutional dictionary made of shifted pulses [4]. The element raw-data corresponding to 1 plane-wave insonification are compressed with undersampling ratios (M/N) ranging between 0.05 and 0.5. For SDA–CNL and the CS reconstruction, an i.i.d. Gaussian random matrix with zero mean and a variance equal to $1/M$ is used. For SDA–CL, the compression is performed

[1] https://github.com/dperdios/us-rawdata-sda

[2] https://www.creatis.insa-lyon.fr/EvaluationPlatform/picmus/index.html

by the compression layer of the network, learned during the training phase.

In the case of SDA–CNL and SDA–CL, the signal recovery is achieved by the reconstruction layers. In the case of CS reconstruction, the signal is retrieved by performing 1000 iterations of the primal-dual forward backward algorithm [14] where the hyper-parameters are empirically tuned to obtain the highest image quality. A standard delay-and-sum algorithm, with spline interpolation for delay calculations and taking into account element-directivity, is performed on the recovered signals to obtain the RF image. The envelope is extracted using the Hilbert transform, normalized and log-compressed to obtain the final B-mode image. The considered dB ranges are 60 dB for the numerical and the *in vitro* images and 40 dB for the *in vivo* images.

The image quality is evaluated using the peak-signal-to-noise ratio (PSNR), computed on the final B-mode image against a reference B-mode image obtained from the element-raw data without compression.

B. Performance Evaluation

The PSNR values, summarized in Table II, show that SDA–CL outperforms both the CS reconstruction and SDA–CNL in many cases considered in this study. Regarding the evolution of the PSNR against the undersampling ratio, it is interesting to note the difference in terms of behaviour between SDA–CL and the CS reconstruction. For undersampling ratios below 0.30, the CS reconstruction tends to stagnate at a relatively low PSNR while the PSNR of SDA–CL is increasing considerably. This leads to a difference of 4.5 dB to 5.5 dB at an undersampling ratio of 0.30. For higher undersampling ratios, the CS reconstruction enters in its phase transition regime with significant increase of the PSNR while the quality of SDA–CL tends to stagnate. This leads to similar PSNR between the two approaches at high undersampling ratios.

Fig. 2 displays the B-mode images of the cross-section of the carotid and of an *in vitro* phantom for an undersampling ratio of 0.30, reconstructed with SDA–CL on Fig. 2(b) and Fig. 2(e) and with CS on Fig. 2(c) and Fig. 2(f). The difference in terms of PSNR between SDA–CL and the CS reconstruction is confirmed by a visual assessment of the corresponding B-mode images. Indeed, SDA–CL recovers a speckle patterns close to the reference, whereas the CS reconstruction does not. This comes from the fact that fully developed speckle patterns are not sparse in the model considered for the CS reconstruction method. However, SDA–CL seems to suffer from oscillating artifacts around sharp hyperechoic regions.

In terms of computational complexity, the reconstruction layers of SDA–CL involve 3 matrix-vector products, each of which has a complexity of $\mathcal{O}(MN)$ which corresponds to the minimal computational complexity of one iteration of an optimization algorithm [10]. Thus the complexity of the proposed SDA–CL architecture is two to three orders of magnitude lower than iterative algorithms, which makes it suitable for real-time imaging without the need to finely tune hyper-parameters.

IV. Conclusion

In this work, we propose to use stacked denoising autoencoders, composed of four layers, for ultrasound image compression and recovery. The first layer of the network is used to compress the signal and the three remaining layers are exploited during the reconstruction process. We suggest two architectures, namely SDA–CNL where a linear compression layer is not learned and SDA–CL where a non-linear compression is learned during the training process. We describe an imaging pipeline into which the proposed architectures may be integrated. The proposed methods are evaluated on the PICMUS dataset and we demonstrate that SDA–CL outperforms a state-of-the-art compressed-sensing-based reconstruction both in terms of quality and reconstruction time.

Acknowledgements

This work was supported in part by the UltrasoundToGo RTD project (no. 20NA21_145911), evaluated by the Swiss NSF and funded by Nano-Tera.ch with Swiss Confederation financing.

References

[1] E. J. Candès and M. Wakin, "An introduction to compressive sampling," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 21–30, 2008.

[2] M. A. Davenport, M. F. Duarte, Y. C. Eldar, and G. Kutyniok, *Introduction to compressed sensing*. Cambridge University Press, 2012, pp. 1–64.

[3] H. Liebgott, R. Prost, and D. Friboulet, "Pre-beamformed RF signal reconstruction in medical ultrasound using compressive sensing," *Ultrasonics*, vol. 53, no. 2, pp. 525–533, 2013.

[4] A. Besson, R. E. Carrillo, D. Perdios, M. Arditi, Y. Wiaux, and J.-P. Thiran, "A compressed-sensing approach for ultrasound imaging," in *Signal Processing with Adaptive Sparse Structured Representations (SPARS) workshop*, 2017.

[5] G. David, J.-L. Robert, B. Zhang, and A. F. Laine, "Time domain compressive beam forming of ultrasound signals," *J. Acoust. Soc. Am.*, vol. 137, no. 5, pp. 2773–2784, 2015.

[6] A. Besson, R. E. Carrillo, O. Bernard, Y. Wiaux, and J.-P. Thiran, "Compressed delay-and-sum beamforming for ultrafast ultrasound imaging," in *IEEE Int. Conf. Image Process.*, 2016, pp. 2509–2513.

[7] O. Lorintiu, H. Liebgott, M. Alessandrini, O. Bernard, and D. Friboulet, "Compressed sensing reconstruction of 3D ultrasound data using dictionary learning and line-wise subsampling," *IEEE Trans. Med. Imaging*, vol. 34, no. 12, pp. 2467–2477, 2015.

[8] C. Quinsac, A. Basarab, and D. Kouamé, "Frequency domain compressive sampling for ultrasound imaging," *Adv. Acoust. Vib.*, vol. 2012, pp. 1–16, 2012.

[9] Z. Chen, A. Basarab, and D. Kouamé, "Compressive deconvolution in medical ultrasound imaging," *IEEE Trans. Med. Imaging*, vol. 35, no. 3, pp. 728–737, 2016.

[10] A. Mousavi, A. B. Patel, and R. G. Baraniuk, "A deep learning approach to structured signal recovery," in *2015 53rd Annu. Allert. Conf. Commun. Control. Comput.*, 2015, pp. 1336–1343.

[11] H. Liebgott, A. Rodriguez-Molares, F. Cervenansky, J. Jensen, and O. Bernard, "Plane-wave imaging challenge in medical ultrasound," in *2016 IEEE Int. Ultrason. Symp.*, 2016, pp. 1–4.

[12] B. E. Treeby and B. T. Cox, "k-Wave: MATLAB toolbox for the simulation and reconstruction of photoacoustic wave fields," *J. Biomed. Opt.*, vol. 15, no. 2, p. 021314, 2010.

[13] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," *Proc. 13th Int. Conf. Artif. Intell. Stat.*, vol. 9, pp. 249–256, 2010.

[14] P. L. Combettes, L. Condat, J.-C. Pesquet, and B. C. Vu, "A forward-backward view of some primal-dual optimization methods in image recovery," in *IEEE Int. Conf. Image Process.*, 2014, pp. 4141–4145.

TABLE II
PEAK-SIGNAL-TO-NOISE RATIO COMPUTED ON THE IMAGES OF THE PICMUS DATASET FOR DIFFERENT UNDERSAMPLING RATIOS

Test case	**Algorithm**	**Undersampling ratio**									
		0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
Carotid cross	SDA–CNL	16.08	16.03	16.06	16.49	17.40	19.60	22.76	26.10	28.59	30.43
	SDA–CL	**17.57**	**19.20**	**20.93**	**22.73**	**24.91**	**28.33**	**32.74**	**34.11**	34.84	35.57
	CS	15.70	16.27	16.64	17.74	19.64	22.98	27.13	31.73	**36.08**	**39.24**
Carotid long	SDA–CNL	13.87	14.07	14.27	14.60	15.41	17.05	19.88	22.77	24.92	26.53
	SDA–CL	**15.43**	**17.03**	**18.47**	**20.09**	**22.18**	**25.33**	**29.03**	**30.54**	**31.38**	**32.07**
	CS	14.17	14.55	15.07	16.14	18.14	20.61	23.79	27.87	30.01	31.56
In vitro type 1	SDA–CNL	16.10	17.85	18.80	19.69	20.60	22.03	23.82	25.41	26.78	27.73
	SDA–CL	14.48	**18.21**	**20.17**	**22.38**	**24.92**	**28.15**	**30.79**	**31.83**	**32.44**	**33.25**
	CS	**16.84**	17.73	18.42	19.25	20.54	22.42	24.88	27.51	30.50	33.12
In vitro type 2	SDA–CNL	16.91	17.66	17.80	18.11	18.62	19.87	21.38	23.24	24.74	25.93
	SDA–CL	15.35	**18.18**	**20.02**	**21.98**	**24.07**	**26.58**	**29.11**	**30.21**	**30.91**	31.72
	CS	**17.37**	17.65	18.03	18.67	19.76	21.71	24.00	26.94	30.29	**33.30**
In vitro type 3	SDA–CNL	16.90	17.86	18.39	19.02	19.85	21.14	22.55	24.25	25.46	26.59
	SDA–CL	15.33	**18.14**	**20.01**	**22.20**	**24.54**	**27.42**	**29.91**	**31.02**	**31.79**	32.47
	CS	**17.33**	17.90	18.48	19.22	20.46	22.21	24.44	27.27	30.31	**33.38**
Numerical	SDA–CNL	13.96	13.61	13.98	14.82	16.30	19.14	22.58	26.06	27.82	28.78
	SDA–CL	**16.76**	**17.52**	**19.54**	**21.74**	**24.30**	**26.90**	**28.39**	**28.69**	28.61	28.58
	CS	13.83	14.83	16.39	17.93	20.07	22.51	25.26	28.21	**30.94**	**33.21**

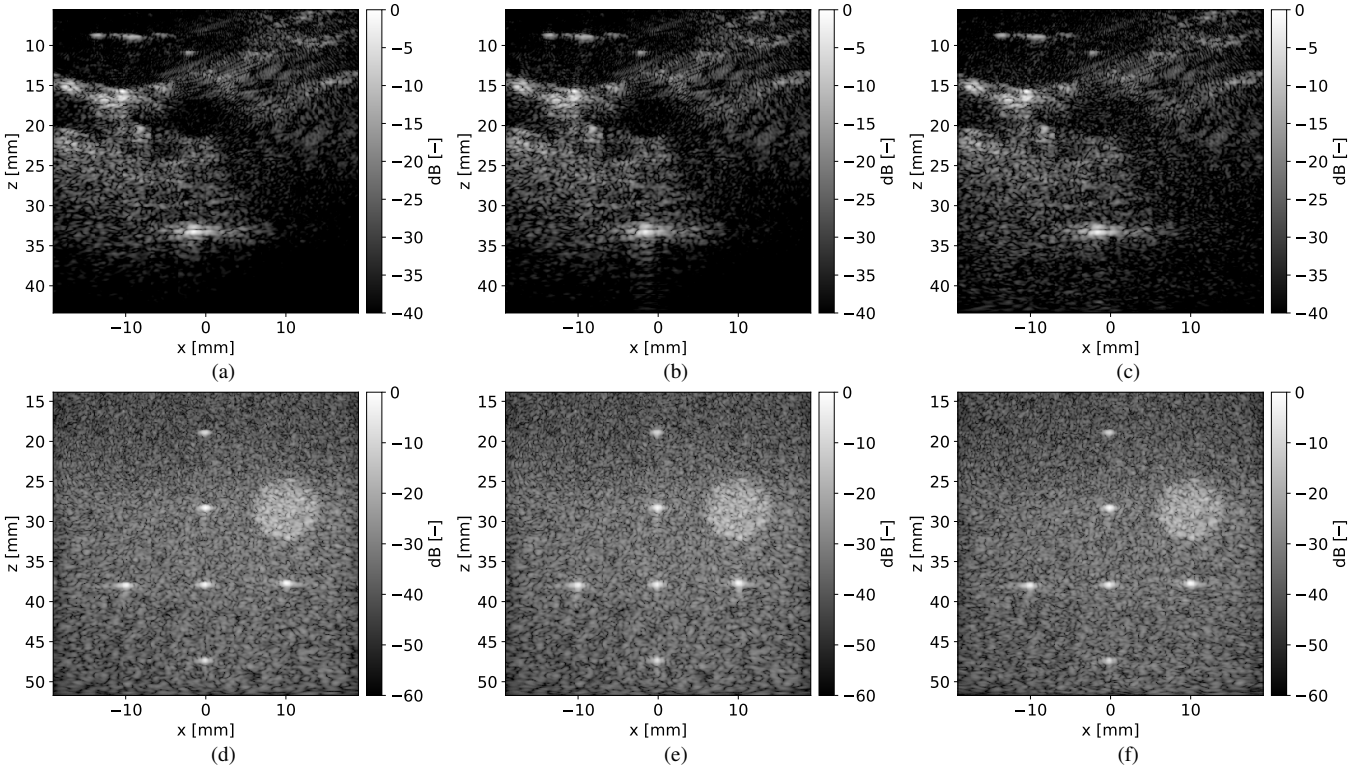

Fig. 2. B-mode images of the cross-section of the carotid ((a)-(c)) and of an *in vitro* phantom ((d)-(f)) obtained with 1 plane-wave insonification; (a)-(d) Reference images obtained without compression; (b)-(e) Images obtained with an undersampling ratio of 0.30 and reconstructed with SDA–CL; (d)-(f) Image obtained with an undersampling ratio of 0.30 and reconstructed with the CS algorithm.