A Portable Ultrasound Imaging System Utilizing Deep Generative Learning-Based Compressive Sensing On Pre-Beamformed RF Signals

Saurabh Kumar Gupta[1*], Kundan Kumar[1*], Chandra Sekhar Seelamantula[2] and Chetan Singh Thakur[1]
{saurabhgupta,kundankumar,chandrasekhar,csthakur}@iisc.ac.in

***Abstract*— Recent advances in the unsupervised and generative models of deep learning have shown promise for application in biomedical signal processing. In this work, we present a portable resource-constrained ultrasound (US) system trained using Variational Autoencoder (VAE) network which performs compressive-sensing on pre-beamformed RF signals. The encoder network compresses the RF data, which is further transmitted to the cloud. At the cloud, the decoder reconstructs back the ultrasound image, which can be used for inferencing. The compression is done with an undersampling ratio of 1/2, 1/3, 1/5 and 1/10 without significant loss of the resolution. We also compared the model by state-of-the-art compressive-sensing reconstruction algorithm and it shows significant improvement in terms of PSNR and MSE. The innovation in this approach resides in training with binary weights at the encoder, shows its feasibility for the hardware implementation at the edge. In the future, we plan to include our field-programmable gate array (FPGA) based design directly interfaced with sensors for real-time analysis of Ultrasound images during medical procedures.**

I. INTRODUCTION

High-resolution ultrasound imaging systems generate a large amount of data with a high frame rate to provide high-resolution imaging for bio-medical applications. It requires computational intensive resources and high-speed data transfer links, further making the system bulky and power consuming. This results in a lack of portability and deployment of the system on the remote-location where power budget is limited, further restricting the health-care accessibility. There has been a trade-off between the quality of image and cost of equipment, which limits the use of ultrasound imaging at an affordable cost with good image resolution. In recent year, the problem of recovering under-sampled measurements has shown a growing interest along with the emergence of compressed sensing (CS) framework [1]. In ultrasound imaging, compressive sensing framework has been used for compressed data acquisition [2] and beamforming [3], opening a path to the reconstruction of high-resolution images with under-sampled data.

A signal $x \in R^N$ can be represented as a linear combination of the elements of some basis function ψ, $x = \Psi\alpha$, where $\psi \in R^{N \times N}$ and $\alpha \in R^N$. With a correct choice of basis ψ, a large number of entries in α are zero or close to zero, with only $K \ll N$ coefficients having useful information. The compressive sensing framework utilizes this sparsity of signal in ψ basis and reconstructs the original signal from a few measurements sampled at a low rate. A simple approach of compression could be computing all the α from x and then storing only the values and positions of K significant α. In the compressive sensing approach, we need not compute all N coefficients. Instead, we collect $M \ll N$ measurements of x in basis $\phi \in R^{M \times N}$ (original representation of signal), which is independent and incoherent to basis ψ.

M measurements of signal x can be obtained by multiplying a $M \times N$ measurement matrix to $N \times 1$ column vector x.

$$y = \phi x \tag{1}$$

The signal x can be accurately recovered from it's compressed measurements y under certain conditions on matrices ϕ and ψ. First, the matrix $\theta = \phi \times \psi$ should obey the Restricted Isometry Property (RIP) of order K (sparsity of x in basis ψ) and the matrices ϕ and ψ should be incoherent. The incoherence between ϕ and ψ is assured by generating matrix ϕ randomly with the Bernoulli (± 1) or i.i.d Gaussian entries and choosing ψ an N $\times$ N DCT or Haar matrix (when ϕ and ψ are sufficiently large matrices). The reconstruction of original signal x from compressed measurement y relies on solving the following convex optimization problem:

$$min\|\widehat{\alpha}\|_1 \left(= \sum |\widehat{\alpha_n}|\right) \ such \ that \ y = \phi\psi\widehat{\alpha} \tag{2}$$

However, the compressive sensing framework suffers some major bottlenecks when it comes to ultrasound imaging. The matrix ϕ satisfies RIP properties only if M is sufficiently large. Real-time, high frame-rate US systems impose difficulty in generating such large matrices. Also, the CS reconstruction involves the use of convex optimization algorithms that require hundreds of iterations to converge, which limits its use in real-time implementation in ultrasound imaging systems at the health-care node.

In the recent literature Bora *et al.*[4] have shown compressed-sensing using generative models, they have shown VAE based compressive-sensing on non-biomedical images. Perdios *et al.*[5] have shown a Denoising Autoencoder based architecture for compressive-sensing on ultrasound signals. The goal of our work is to present a non-iterative algorithm for recovering under-sampled ultrasound signals using generative models, Variational Autoencoders [6][7], and we also propose a cloud and edge-based distributed Ultrasound imaging System architecture, where RF

*These authors contributed equally to the manuscript
[1] Department of Electronic Systems Engineering, Indian Institute of Science, Bangalore-560012, India,
[2]Department of Electrical Engineering, Indian Institute of Science, Bangalore560012, India.

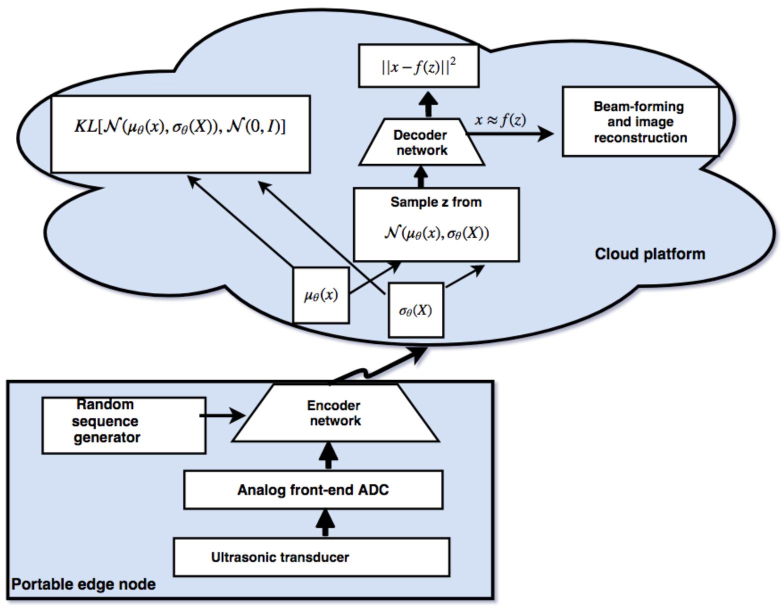

Fig. 1: Proposed System Architecture. The encoder network is implemented at a portable edge node and the decoder and image reconstruction algorithm is implemented on the cloud platform.

data reconstruction and image formation is done on cloud and edge device is responsible for compressed data acquisition. Variational Autoencoders provide an efficient, method to recover a latent representation z ("encoding") of our data-points x (the "decoded" observations). By training VAE on a large dataset, we develop an effective encoding mechanism for our observations, which can be used to either generate realistic new data or reconstruct a data item from a compressed measurement.

The proposed framework can also be used for other medical imaging areas like computed tomography, rapid MRI and neuronal spike train recovery. We will further explore the system level implementation of hardware with less computational complexity and low power consumption.

II. Architecture

A. Proposed Architecture

In the proposed cloud-based system, we implement the first layer of the encoder part of the VAE at edge devices, which we call it data acquisition layer. In our design, we fix the weights of this layer with randomly generated binary entries (+1,-1), which reduces the computations at the edge nodes, making the design of the edge devices compact, low-power and portable. The edge device acquires an analog RF data from US transducer probes, performs pre-processing and converts an analog signal into digital bit streams by Analog to Digital converters (ADC). The digital bit stream is further under-sampled at the first layer of the encoder. The M/N ($M \ll N$) undersampling is achieved by multiplying input vector $(N \times 1)$ with the encoder matrix (M×N). The encoder matrix is formed by randomly generated entries of '+1' and '-1'. The random entries can be generated using a conventional Linear feedback shift register (LFSR) based random sequence generators in real time and we do not need to store encoder weights in system memory, which reduces the system memory requirement at edge nodes significantly.

The decoder layers of VAE and image reconstruction algorithms using delay-and-sum beamforming can be implemented on the cloud platform. Figure 1. presents a block diagram of our proposed architecture.

B. VAE Architecture

The Variational Autoencoder is a deep Bayesian generative model framework, which tries to learn the distribution from which data has been generated. It models the similarity between two random variable x and latent variable z. The Mathematical representation of VAE is the marginal distribution on observation x defined by a prior distribution $p(z)$ and a conditional distribution $P_\theta(x|z)$.

$$P_\theta(x) = \int P_\theta(x|z) P_\theta(z)\ dx \tag{3}$$

A prior $p(z)$ distribution is usually assumed to be a multi-variate unit Gaussian with zero mean $\mathcal{N}(0,\ I)$. Also, in VAE a conditional distribution $P_\theta(x|z)$ is a generator network parameterized by θ and is usually a Gaussian $\mathcal{N}(\mu_\theta(x),\ \sigma^2_\theta(x))$ where $\mu_\theta(x)$ and $\sigma^2_\theta(x)$ is modeled using the network.
In VAE standard training procedure involves maximizing the evidence of the lower bound $(ELBO)$ on the true posterior $P_\theta(z|x)$ with the help of an auxiliary known posterior distribution $q_\phi(z|x)$, where $q_\phi(z|x)$ is parameterized by ϕ and is assumed to be a Gaussian $\mathcal{N}(\mu_\phi(x),\ \sigma^2_\phi(x))$, also $\mu_\phi(x)$ and $\sigma^2_\phi(x)$ are derived by fitting a network.
The $ELBO$ $(\mathcal{L})$ is defined in equation (4-6).

Under-sampling Ratio	PSNR in dB		MSE	
	Model (a)	Model (b)	Model (a)	Model (b)
1/2	23.19	23.35	312.10	300.86
1/3	23.31	23.19	303.74	311.92
1/4	22.96	23.22	328.68	309.17
1/5	23.30	23.39	304.07	297.71
1/10	22.85	22.63	337.48	354.76
1/30	21.92	21.92	417.68	417.68
1/100	21.40	21.62	470.69	447.89

TABLE I:
Peak-Signal-To-Noise Ratio and Mean-Square-Error computed on the reconstructed images for different under-sampling ratios. Model (a) corresponds to encoder layer with trainable weights and model (b) corresponds to encoder layer with fixed binary weights as described in section II

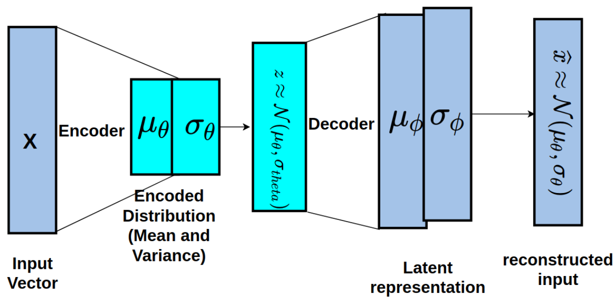

Fig. 2: Layer-wise representation of VAE network

$$\mathcal{L}(x:\theta,\phi) \;\geq\; log_{P_\theta}(x) - KL[q_\phi(z|x) \;||\; P_\theta(z|x)] \quad (4)$$

$$= E_{q_\phi(z|x)}\;[\;log_{P_\theta}(x) \;+\; log_{P_\theta}(z|x) \;-\; log_{q_\phi}(z|x)] \quad (5)$$

$$= \;E_{q_\phi(z|x)}\;[\;log_{P_\theta}(z|x) \;-\; log_{q_\phi}(z|x)] \quad (6)$$

The model defines the KullbackLeibler divergence (KL) term in the loss equation (4), which is also intractable due to the term involving calculation of true posterior $P_\theta(x|z)$. Further, with the use of reparameterization trick on z we can redefine optimization with θ and ϕ as equation (7)

$$\mathcal{L}(\theta,\phi:x) \;\approx \underset{\theta,\;\phi}{\arg\max} \sum_{m=1}^{M} log_{P_\theta}(x|z^m) + KL[q_\phi(z^m|x)\;||\;P_\theta(z)] \quad (7)$$

Where M denotes a total number of samples used for sampling. Single layer encoder and decoder network is implemented via the $\tanh$ perceptron with θ and ϕ as parameters respectively. Figure 2. shows the layer-wise representation of implemented VAE network.

III. Experiments, Results and Performance Evaluation

For the training and validation of proposed architecture we use the PICMUS challenge [9][10][11][12], simulation resolution/distortion test and PICMUS challenge: in vivo carotid cross-section dataset. Both the dataset consists of 75 plane-wave sequences. One plane wave sequence consists of 128 RF samples corresponding to 128 channels in the

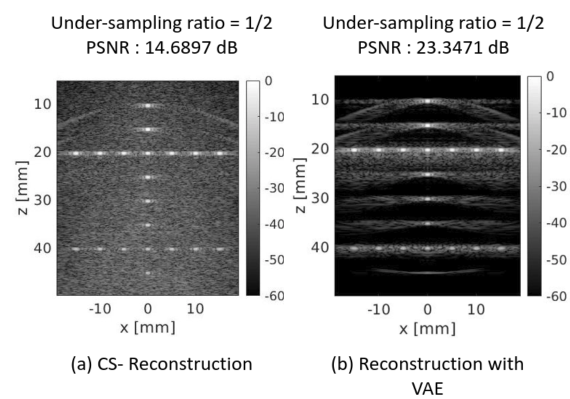

Fig. 3: (a) represents traditional CS-reconstruction by solving equation (2) as described in section I and (b) represents reconstruction with VAE, Under-sampling ratio = 1/2

transducer probe. For generalized training of VAE, we used all the samples from both the datasets, we extracted a total $2 * 75 * 128 \;=\; 19200$ RF samples, which we split into training and test dataset (15360 samples used for training and 3840 for testing). The VAE model was trained for undersampling ratios of 1/2, 1/3, 1/4, 1/5, 1/10, 1/30, and 1/100. We validated the image reconstruction on simulation resolution/distortion test dataset. After reconstruction, image formation was done using standard delay and sum beamforming. Further, the envelope detection is done using Hilbert transform, which is normalized and log compressed with 60 dB range to generate the final B-mode image.

We examined the performance of VAE architecture with two models : (a) Trainable weights at Data acquisition layers, (b) fixed binary weights (+1,-1) from the Bernoulli distribution, at data acquisition layers. It is noticed that model (b) outperforms over (a) in performance and is also suitable for low power portable hardware at edge devices. Figure 5. presents the reconstructed images for each compression ratio (using binary weights). Table I summarizes the performance evaluation for both models. Figure 4. shows the plot of PSNR vs undersampling ratio for models (a) and (b). We also compared our reconstruction algorithm with reconstruction using CS-framework (compressing with random i.i.d Gaussian matrix ϕ and reconstruction by solving equation (2)). Figure 3. compares the reconstruction using the CS

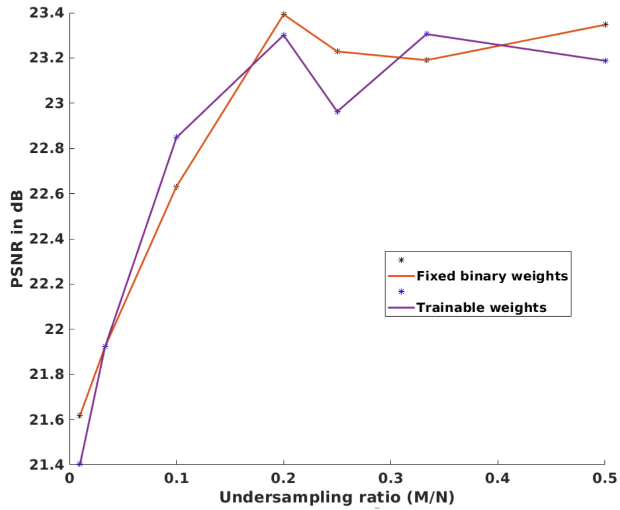

Fig. 4: PSNR vs undersampling ratio for model (a) and (b)

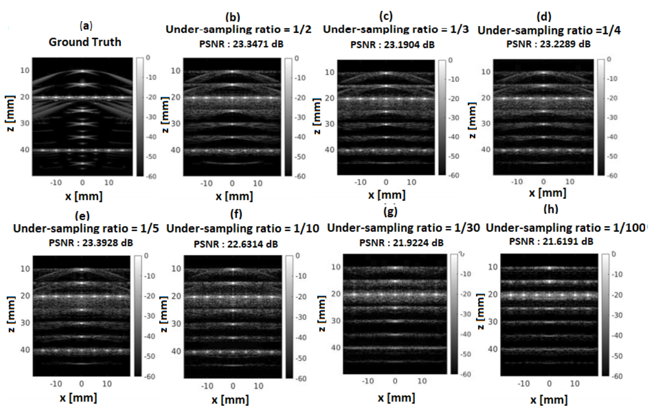

Fig. 5:
(a) is the ground truth image without any compression: (b)-(g) shows reconstructed images with VAE from undersampled signals with undersampling ratios of (b) 1/2, (c) 1/3, (d) 1/4, (e) 1/5, (f) 1/10, (g) 1/30 and (h) 1/100 respectively.

framework(using Orthogonal Matching Pursuit Algorithm (OMP)) [8] and VAE network for the under-sampling ratio of 1/2. It is interesting to note that the proposed algorithm shows a significant improvement of 8.6 dB in PSNR over CS-reconstruction for the undersampling ratio of 1/2. As the US RF signals are less sparse, the CS-reconstruction method could not reconstruct the signals sampled at lower undersampling ratios. In terms of computational efficiency, for the under-sampling ratio of M/N, The VAE decoder network involves two matrix-vector multiplications of computational complexity $\mathcal{O}(M^2)$ and $\mathcal{O}(MN)$ which is in order of the computational complexity of single iteration in CS-reconstruction algorithms.

IV. Conclusions

In this work, we propose a cloud-based Ultrasound Imaging System with portable edge devices. The proposed system uses a compressed data acquisition model using Variational Autoencoders. The first layer of the VAE implemented using non-trainable fixed binary weights is used for compressed data acquisition. These fixed weights are generated in real time with a LFSR based random sequence generator. Reconstruction of the compressed samples and image formation algorithms are implemented at the cloud platform. The proposed VAE architecture for compression is evaluated using open-source PICMUS dataset, and we demonstrate that VAE architecture outperforms CS-reconstruction both in terms of PSNR of the reconstructed image and computational complexity. Using VAE architecture, we achieved an improvement of 8.6 dB in PSNR for the undersampling ratio of 1/2, over CS-reconstruction. The proposed architecture also promises an efficient hardware implementation using binary weights at the encoder.

V. Acknowledgement

This work was supported by the INSPIRE faculty fellowship (DST/INSPIRE/04/2016/000216) from the Department of Science & Technology, India.

References

[1] E. J. Candes and M. B. Wakin, ”An Introduction To Compressive Sampling,” in IEEE Signal Processing Magazine, vol. 25, no. 2, pp. 21-30, March 2008.
[2] 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.
[3] C. Wang et al., ”An easily-achieved time-domain beamformer for ultrafast ultrasound imaging based on compressive sensing,” 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Milan, 2015, pp. 7490-7493.
[4] Bora, A., Jalal, A., Price, E., Dimakis, A.G. ”Compressed sensing using generative models.” In: ICML, pp. 537-546 (2017)
[5] D. Perdios, A. Besson, M. Arditi and J. Thiran, ”A deep learning approach to ultrasound image recovery,” 2017 IEEE International Ultrasonics Symposium (IUS), Washington, DC, 2017, pp. 1-1.
[6] D. P. Kingma, M. Welling, ”Auto-encoding variational bayes”, International Conference on Learning Representations (ICLR), pp. 3, 2014.
[7] D.P. Kingma, D.J. Rezende, S. Mohamed, M. Welling, ”Semi-supervised Learning with Deep Generative Models”, Neural Information Processing Systems, 2014.
[8] Y. C. Pati, R. Rezaiifar, P. S. Krishnaprasad, ”Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition”, Proc. 27th Asilomar Conf. Signals Syst. Comput., vol. 1, pp. 40-44, 1993.
[9] H. Liebgott, A. Rodriguez-Molares, F. Cervenansky, J. A. Jensen, O. Bernard, ”Plane-wave imaging challenge in medical ultrasound”, Proc. IEEE Int. Ultrason. Symp., pp. 1-4, Sep. 2016.
[10] J. Camacho and C. Fritsch, ”Phase coherence imaging of grained materials,” in IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control, vol. 58, no. 5, pp. 1006-1015, May 2011.
[11] R. Mallart, M. Fink, ”Adaptive focusing in scattering media through sound-speed in homogeneitiesThe vancittert-zernike approach and focusing criterion”, J Acoust Soc Am, vol. 96, no. 6, pp. 3721-3732, Dec. 1994.
[12] A. Rodriguez-Molares et al., ”The UltraSound ToolBox,” 2017 IEEE International Ultrasonics Symposium (IUS), Washington, DC, 2017, pp. 1-4.