2019 IEEE International Ultrasonics Symposium (IUS)
Glasgow, Scotland, October 6-9, 2019

Deep-Learning High-Dynamic-Range Ultrasound

Eduardo Prado
Computer Aided Medical Procedures, Technische Universität München
Boltzmannstraße 3,
85748 Garching bei München, Germany
eduardo.prado@tum.de

Ivan Pavlov
Computer Aided Medical Procedures, Technische Universität München
Boltzmannstraße 3,
85748 Garching bei München, Germany
ivan.pavlov@tum.de

Nassir Navab
Computer Aided Medical Procedures, Technische Universität München
Boltzmannstraße 3,
85748 Garching bei München, Germany
Computer Aided Medical Procedures, Johns Hopkins University,
3400 North Charles Street,
Baltimore, MD 21218, USA
nassir.navab@tum.de

Guillaume Zahnd
Computer Aided Medical Procedures, Technische Universität München
Boltzmannstraße 3,
85748 Garching bei München, Germany
g.zahnd@tum.de

***Abstract*—In recent years ultrasound imaging has achieved an increasing acceptance across medical specialties. For this reason new techniques keep being tested in the field. Among these techniques we found High Dynamic Range (HDR) imaging where the range of luminosity levels is augmented by combining multiple expositions of a scene. Current ultrasound techniques present limitations that are not compatible with traditional implementations of HDR imaging. In this paper, we asses the use of a deep learning (DL) neural network (U-net architecture) on predicting HDR values from low dynamic range (LDR) input images. In addition, an image acquisition pipeline to create the data set from which the network was trained is described. We demonstrated that this type of networks can be trained to predict HDR out from a minimal number of input expositions, while the obtained results showed to be comparable with more traditional approaches.**

***Index Terms*—Deep learning, Ultrasound, High Dynamic Range**

I. Introduction

High Dynamic Range (HDR) imaging is a technique inspired from computational photography, where multiple images of the same scene are acquired with a different exposure and combined to generate a final image where each luminance level can be observed [1]. Raw ultrasound (US) signal natively contains a high dynamic range. To be able to display the image on commercial monitors, a logarithmic compression is applied to the raw input. This conversion thus reduces the range of tones that can be used to represent and differentiate between objects or tissue containing a wide echogenicity range [2].

Besides pre- and post-processing algorithms, in recent years, multiple DL approaches have been tested on US imaging for a variety of applications [3]. Nevertheless, current US techniques present limitations that are not compatible with traditional implementations of HDR. For instance, HDR reconstruction will require a series of images where both US probe and target remain stable for relative long periods of time. At the same time, the necessity for exposition variation hinder HDR computation to be performed in real time. Finally the absence of research applying HDR to US imaging translates into a lack of images that could be collected from different sources for their later integration into problem specific datasets for DL and ML algorithms.

The main contributions of this paper can be summarized as:

1) The construction of a *in vivo* US training dataset combining low dynamic range (LDR) and HDR images acquired at different exposures.
2) The augmentation of an LDR-HDR US dataset from previously acquired single-exposure images, artificially compressed to have enough expositions for their HDR computation.
3) The implementation and evaluation of a convolutional neural network (CNN), able to reconstruct HDR information from a minimum number of input expositions.

II. Related Work

US imaging works under the principle of sound wave scattering and propagation. Mechanical waves such as US waves propagate along a longitudinal path. As US waves penetrate the body, they exert pressure over the different tissues they encounter. Different tissues present different impedance values Z, resulting in a variance in US speed across a tissue.

On the absence of a wave it can be said that the acoustic pressure p is zero. For longitudinal waves, the underlying particle velocity v can be expressed as:

$$p = vZ \tag{1}$$

This work was supported by the NVIDIA Corporation GPU grant program

Where Z is called the characteristic impedance described by the pressure and the speed of sound c at different materials:

$$Z = \rho c \tag{2}$$

The pressure of a scattered wave by a target can be computed by:

$$p_s(r,t) = A_0 \frac{Re^{-\mu_a(d+r)}}{r} F\left(t - \frac{d+r}{c}\right) \tag{3}$$

Only the amplitude A_0 will be a parameter intrinsic to an US transducer. As illustrated by Equation 3, the rest of the terms are specific to the media being imaged and are independent of the US machine tunable parameters. With c representing the speed of sound, d the distance between the target and the transducer, F the acoustic envelope, r the radius from the target, R the reflection coefficient, and μ_a is the amplitude attenuation factor. All this specific of the tissue in the wave trajectory.

In traditional photography, the intensity of a pixel is determined by a non linear relation between the radiance on a scene and a camera intrinsic parameters [4]. When a camera captures a scene, the amount of light captured by the camera is known as dynamic range. The nonlinear relation between intensity values and exposure is generally known as the camera response function and explains a non-linear mapping of image irradiance. This response function can be use as a mapping function to compress irradiance values in a determined dynamic range to relative smaller range.

In recent years multiple studies have extended the analysis on LDR-to-HDR conversion. For instance, HDR in ultrasound imaging is possible as the specific irradiance response from various tissues are often source of either over- or under-exposition [5]. In the same way similar domain transformation problems have been approached by means of CNNs [6]. Although approaches for DL-based HDR have been attempted [7] [8] [9], literature on approaches to use this technique for US imaging are so far sparse.

III. Methods

A. Ultrasound HDR images computation

By taking certain restrictions into account, such as probe position, depth, and frequency, a comparable response curve can be obtained for US probes. By only varying signal amplitude and capturing an image per amplitude value a probe-specific response curve can be computed from US images.

Preparing US HDR images requires a minimum of exposition using the approach proposed by Degirmenci et al [5]: Fifteen images were taken by varying the acoustic power knob on an Ultrasonix (Ultrasonix Medical Corporation, Richmond BC, Canada) scanner with a linear array (L14-5/38) as well as a convex array (C5-2/60). For both probes a frequency of 6.6 MHZ was fixed, at a 5 cm depth. Other parameters such as dynamic range and gain were fixed only during a single study, but later modified between studies to obtain different response curves for better dataset generalization. The *in vivo* targets were placed in a container full of warm water and then fixed in a position that assured minimum movement for the length of the experiment. Fifteen images were taken corresponding to 15 levels of acoustic power each power level was converted to relative amplitude A_r by computing:

$$A_r = 10^{\frac{Power}{20dB}} \tag{4}$$

A sample of the recovered camera curve for one of the test probes is shown in Figure 1. This process yielded two tone-mapped LDR images and a HDR file containing the whole range of the compressed images used. A tone-mapping operation was needed to display an LDR version of the HDR reconstruction that contained the highest (as possible) dynamic range. Local versions of Reinhard and Durand operators were computed for this purpose. By measuring the peak signal-to-noise ratio as suggested in [5], these operators showed the widest range of power levels on our experiments.

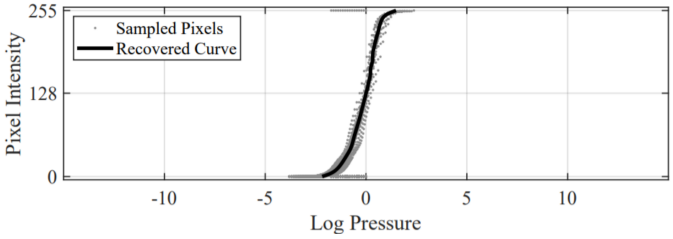

Fig. 1. Probe recovered response curve.

The approach from Degirmenci et al. [5] using the MATLAB (MathWorks, Natick, MA, USA) [10] `HDRUS.m` script was modified to read batches of US studies of the Ultrasonix scanner export format *b8* and *b32*. A total of 32 *in vivo* experiments where done following this method resulting in 480 LDR images of with multiple sizes no smaller than 400 $\times$ 400 pixels and 32 HDR files.

B. Synthetic intensity compression

At the present time HDR computation for US requires a minimum of exposures, being the minimal number for computation matter of experimentation and consideration for the final application. For this reason, the following strategy was adopted to synthesize a collection of images necessary for neural network training.

Given a single *in vivo* US image acquired with optimal settings (referred to as the 0 dB standard), two series of seven images were generated with controlled power levels by applying a known compression (respectively, decompression) on the pixel values. This set of 15 input images was used to construct the HDR ground truth label for network training, namely by performing the analytical HDR computation method previously mentioned and validated. The results of the described process are shown in Figure 2.

To change the dynamic range of an input image a mapping function that worked strictly as an increasing function (when decompressing) or as a decreasing function (when compressing) was required. These type of mapping functions

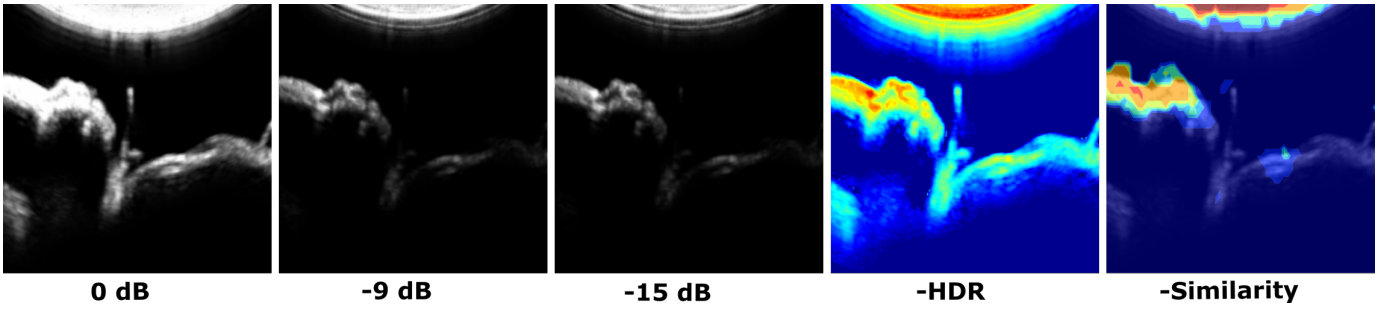

Fig. 2. Artificial multi-exposure LDR set with tone-mapped versions of HDR on a baby doll (toy).

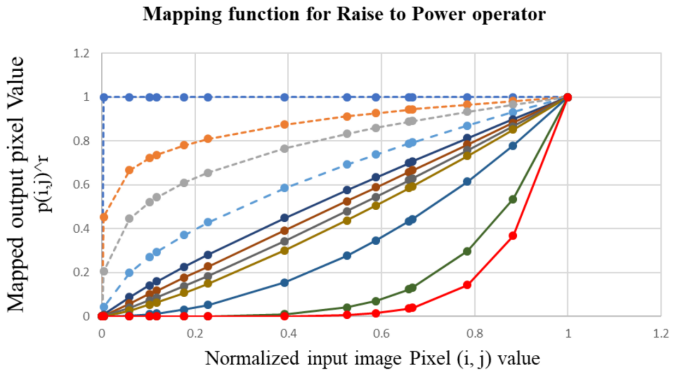

Fig. 3. Probe recovered response curve.

are also known as anamorphic operators [11]. A "Raise-to-Power" operator described by Equation 5 was chosen due to its capacity of changing an image dynamic range while its mapping operation can be described by an exponential curve.

$$Q(i,j) = cP(i,j)^r \tag{5}$$

Here, P refers to the input image and Q to the output image. The term r represents the fix power operator which will determine the operation as a compression for large values and a decompression or gain operator for small r values, c represents a scaling constant that takes into account the maximum and minimum pixel value of the input image.

All collected images were obtained across two external [12] [13] and one internal [14] US datasets. After manually discarding images that were smaller than 224×224 pixels and/or that contained annotations over the area of interest, the overall size of the dataset was to 1901 "0 dB" standards images. Subsequently, processing these images with a known compression, a total of 28,515 LDR images where obtained (15 compressions per original image). Figure 3 shows the mapping function for eleven values of r for each curve. As r approaches smaller values, this type of operator performs less efficiently. From this behavior we can safely assume that a "Raise-to-Power" operator will increase contrast in areas with high pixel values at cost of the contrast in low pixel values.

Loss	Epochs	Batch size	Learning rate	Train size	Validation size
MSE	40	15	0.0001	22815	5700
$L1_\theta$ Loss	64	15	0.001	22815	5700

TABLE I
NETWORK PARAMETERS SUMMARY

C. Deep learning HDR

Our approach is based on the use of a U-net architecture [15] due to its capacity to efficiently perform domain transformations. At the same time connections between the down sampling and up sampling sections of the network facilitate transfer and preservation of low- and high-level features. Compared to approaches that produce visually convincing results [6], U-net provides more predictable results between iterations [16]. At the same time it does not require restrictions on output intensity for training stabilization, which could result on non-optimal HDR predictions.

In order to determine the ability of the network to predict HDR values based on DLR inputs two different loss functions where tested a mean square error (MSE) loss based on the approach by Eilertsen et al. [8] and a combination of L1 loss combined with a cosine similarity term ($\mathbf{L1}_\theta$) as suggested by Marnerides et al in [7].

IV. RESULTS

The network was trained with a different loss function on each setup. All experiments were done using 80% of samples taken from the synthetic dataset ($22,815$ training pairs). Evaluation of the model was performed using the remaining 20% of data not seen during training ($5,700$ samples). Every train pair consisted of a HDR label and one of the 15 synthetic expositions used to create the label. Concisely each label was fed 15 times to the network with intend of training the network to predict HDR from one input exposition. A second data loader where each LDR set is combined into a 15 channel input vector is possible. This can be considered as teaching the network to predict HDR values from camera functions since every set r was different when compression was applied, althougt more experimentation is needed in order to provide consistent results. Details on training parameters are summarized on Table I.

To assess the quality of the predicted HDR images, a modification of peak signal-to-noise ratio (PSNR) that takes human visual system (HVS) contrast perception into account was used [17]. The PSN-HVS was computed for one sample per epoch consisting of `Input`, `Ground truth`, and `Prediction`. Two values were obtained, the first from comparing the input image to the ground truth HDR. The second by comparing the input to the predicted HDR. The PSNR-HVS were averaged by the total number of measured samples (same value as number of epochs) and are summarized in Table II.

Experiment	Input - Ground Truth HDR PSNR-HVS	Input - Prediction HDR PSNR-HVS
MSE loss	9.4715 dB	9.5486 dB
$L1_\theta$ Loss	8.8698 dB	7.9117 dB

TABLE II
PEAK SIGNAL-TO-NOISE RATIO SCORE SUMMARY

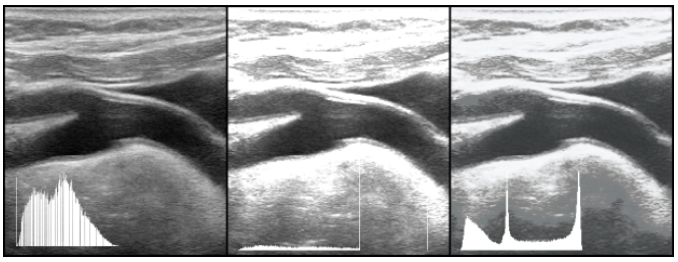

Fig. 4. Intensity distribution histogram for randomly picked sample trained with $L1_\theta$ Loss. Left: Input, Center: Label HDR, Right: Reconstruction HDR.

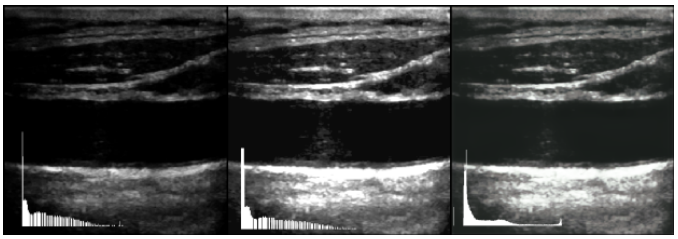

Fig. 5. Intensity distribution histogram for randomly selected sample trained with the MSE loss. Left: Input, Center: Label HDR, Right: Reconstruction HDR.

V. Discussion and Conclusion

From Table II is possible to infer that using an MSE loss the model was able to preserve and even increase the overall contrast perception across predictions. In contrast the model trained with a $L1_\theta$ loss was not capable of replicating the contrast distribution. A quick analysis of the predictions histogram presented in Figures 4 showed how the $L1_\theta$ loss shifted the distribution of pixel intensities. On the other hand Figure 5 shows MSE loss results where a better intensity distribution is seen with a tonality similar to the reference label. So far referenced approaches determine HDR reconstruction quality base on qualitative methods, however objective approaches have been presented [18] in recent years and are worth considering for HDR quality measurement.

The proposed pipeline for an HDR US dataset demonstrated to be effective in training a DL model. In addition, the creation of an HDR-LDR data set from manually taken studies could enable the benchmarking of new approaches and ideas for HDR-US experimentation. Finally, the trained model results suggested that further experimentation combined with better training techniques can result in new implementations of HDR imaging in the field of US imaging.

References

[1] F. Banterle, A. Artusi, K. Debattista, and A. Chalmers, *Advanced High Dynamic Range Imaging: Theory and Practice (2nd Edition)*. Natick, MA, USA: AK Peters (CRC Press), July 2017.

[2] Y. Xiao, M. Boily, H. S. Hashemi, and H. Rivaz, "High-Dynamic-Range Ultrasound : Application for imaging tendon pathology," *Ultrasound in Medicine & Biology*, vol. 44, no. 7, pp. 1525–1532, 2018.

[3] L. J. Brattain, B. A. Telfer, M. Dhyani, J. R. Grajo, and A. E. Samir, "Machine learning for medical ultrasound: status, methods, and future opportunities," *Abdom. Radiol.*, vol. 43, no. 4, pp. 786–799, 2018.

[4] M. D. Grossberg and S. K. Nayar, "What is the space of camera response functions?," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2, pp. II–602–9, 2003.

[5] A. Degirmenci, D. P. Perrin, and R. D. Howe, "High dynamic range ultrasound imaging," *International Journal of Computer Assisted Radiology and Surgery*, vol. 13, pp. 721–729, may 2018.

[6] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative Adversarial Networks," jun 2014.

[7] D. Marnerides, T. Bashford-Rogers, J. Hatchett, and K. Debattista, "ExpandNet: A deep convolutional neural network for high dynamic range expansion from low dynamic range content," *Computer Graphics Forum*, vol. 37, no. 2, pp. 37–49, 2018.

[8] G. Eilertsen, J. Kronander, G. Denes, R. Mantiuk, and J. Unger, "HDR image reconstruction from a single exposure using deep CNNs," vol. 36, no. 6, p. 15, 2017.

[9] S. Shin, K. Kong, and W. J. Song, "CNN-based LDR-to-HDR conversion system," *2018 IEEE International Conference on Consumer Electronics, ICCE 2018*, vol. 2018-Janua, pp. 1–2, 2018.

[10] The Mathworks, Inc., Natick, Massachusetts, *MATLAB version (R2019a)*, 2019.

[11] R. Fisher, S. Perkins, A. Walker, and E. Wolfart, "71 - hypermedia image processing reference (hipr)," 1997.

[12] L. Pedraza, C. Vargas, F. Narvez, O. Durn, E. Muoz, and E. Romero, "An open access thyroid ultrasound-image database," vol. 9287, 01 2015.

[13] M. Zukal, R. Benes, P. Cika, and K. Riha, "Common carotid artery (cca) ultrasound image database," 2011.

[14] W. Simson, M. Paschali, N. Navab, and G. Zahnd, "Deep learning beamforming for sub-sampled ultrasound data," in *2018 IEEE International Ultrasonics Symposium, Kobe, Japan*, 2018.

[15] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 9351, pp. 234–241, 2015.

[16] C. Yang, X. Lu, Z. Lin, E. Shechtman, O. Wang, and H. Li, "High-resolution image inpainting using multi-scale neural patch synthesis," *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, vol. 2017-Janua, pp. 4076–4084, 2017.

[17] N. Ponomarenko, F. Silvestri, K. Egiazarian, M. Carli, J. Aastola, and V. Lukin, "On between-coefficient contrast masking of DCT basis functions," *Multimedia Tools and Applications*, vol. 72, no. 1, pp. 1–4, 2013.

[18] M. Narwaria, M. Perreira Da Silva, and P. Le Callet, "HDR-VQM: An objective quality measure for high dynamic range video," *Signal Processing: Image Communication*, vol. 35, pp. 46–60, 2015.