DEEP LEARNING FOR SUPER-RESOLUTION VASCULAR ULTRASOUND IMAGING

Ruud JG van Sloun[⋆]*, Oren Solomon*[†]*, Matthew Bruce*[‡]*, Zin Z Khaing*[§]*,*
Yonina C. Eldar[†] *and Massimo Mischi*[⋆]

[⋆]Department of Electrical Engineering, Eindhoven University of Technology, The Netherlands
[†]Department of Electrical Engineering, Technion, Haifa, Israel
[‡]Applied Physics Laboratory, The University of Washington, Seattle, WA, USA
[§]Department of Neurological surgery, The University of Washington, Seattle, WA, USA

ABSTRACT

Based on the intravascular infusion of gas microbubbles, which act as ultrasound contrast agents, ultrasound localization microscopy has enabled super resolution vascular imaging through precise detection of individual microbubbles across numerous imaging frames. However, analysis of high-density regions with significant overlaps among the microbubble point spread functions typically yields high localization errors, constraining the technique to low-concentration conditions. As such, long acquisition times are required for sufficient coverage of the vascular bed. Algorithms based on sparse recovery have been developed specifically to cope with the overlapping point-spread-functions of multiple microbubbles. While successful localization of densely-spaced emitters has been demonstrated, even highly optimized fast sparse recovery techniques involve a time-consuming iterative procedure. In this work, we used deep learning to improve upon standard ultrasound localization microscopy (Deep-ULM), and obtain super-resolution vascular images from high-density contrast-enhanced ultrasound data. Deep-ULM is suitable for real-time applications, resolving about 1250 high-resolution patches (128x128 pixels) per second using GPU acceleration.

Index Terms— Ultrasound, Deep learning, Super resolution, Contrast agents

1. INTRODUCTION

High-fidelity microvascular imaging is of great interest for a multitude diagnostic applications, having the potential to detect and phenotype diseases that are hallmarked by microvascular alterations. One such application is the detection of angiogenesis in cancer, characterized by a chaotic vascular network that excibits irregular geometries, and is a good predictor for the development of distal metastasis [1]. Likewise, assessment of microvascular deficits and subsequent hypoperfusion following traumatic spinal cord injury may provide new neurological monitoring opportunities [2].

Contrast-enhanced ultrasound imaging is a cost-effective imaging modality that enables microvascular imaging through the use of an intravascular contrast agent. This is made of microbubbles, which are sized similarly to red blood cells and therefore reach the full vascular network up to the capillary level. However, the attainable ultrasound resolution is typically not sufficient to resolve microvascular structures, being limited by wave diffraction and the transducer array geometry. Although the use of smaller wavelengths permits higher resolutions and therewith imaging at smaller scales, it comes at the cost of reduced penetration depth since higher frequencies suffer increasingly from absorption of the acoustic energy by tissue.

In ultrasound localization microscopy (ULM) [3, 4, 5, 6] this inherent physical trade-off between resolution and penetration depth is circumvented by exploiting super-resolution concepts from optics (e.g. Photoactivation Localization Microscopy [7]). Instead of constructing a diffraction-limited vascular image by taking the average or maximum intensity across an ensemble of ultrasound frames over time, ULM aims at localizing the centroids of individual microbubbles on a frame-by-frame basis to construct a super-resolved image. While ULM was shown to enable a 10-fold improvement in resolution as compared to its diffraction-limited counterpart [3], it typically relies on long acquisition times (minutes) to cover the full vascular bed. Attaining a motion-free acquisition across such a time-span is often infeasible in a clinical setting. The strong constraint that ULM poses on acquisition time could be alleviated significantly if one were able to drastically increase the microbubble concentration. Unfortunately, similar to single-molecule localization in optics, single-microbubble localization in ULM breaks down when the imaging point-spread-functions (PSFs) of the detected microbubbles display substantial overlaps. This scenario is increasingly likely to happen for higher concentrations. As such, standard ULM methods requiring isolated microbubbles with non-overlapping PSFs dictate relatively long acquisition times.

Recently, algorithms based on sparse recovery were pro-

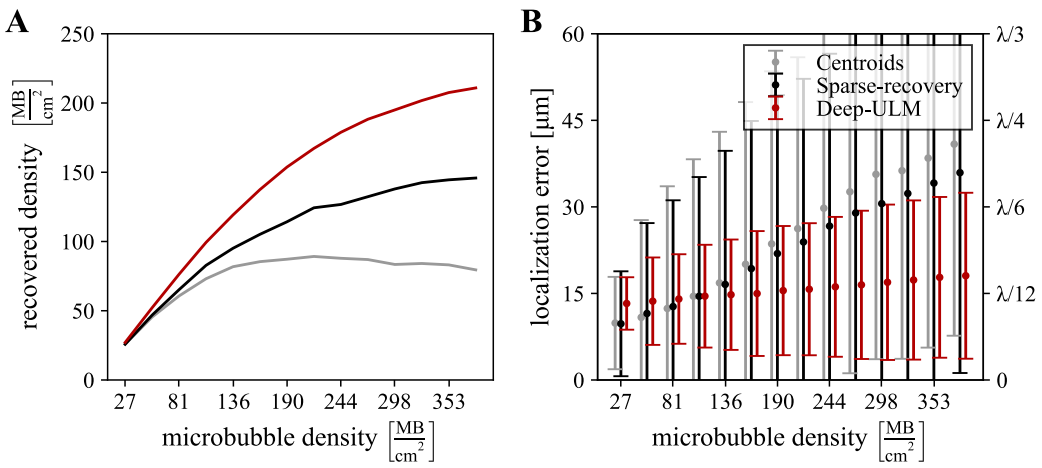

Fig. 1. Detection rate and localization precision of Deep-ULM (red) compared to ULM based on centroids (gray) and sparse recovery (black). (A) Recovered density as a function of simulated microbubble (MB) density, and (B) corresponding median localization errors with bars representing the standard deviation.

posed that permit the use of higher microbubble concentrations [8, 9, 10]. However, even highly optimized sparse recovery methods have an iterative nature, with a computation time that grows significantly with the field of view. As such, real-time implementation remains a major challenge.

In this paper, we propose Deep-ULM, an ultrasound localization microscopy strategy based on deep learning, designed and trained to cope with high-concentration contrast-enhanced ultrasound acquisitions. This method exploits an end-to-end convolutional neural network that maps individual low-resolution input frames to high resolution outputs. Deep-ULM is fast, offering the advantages of sparse-recovery at a much lower computation cost, reaching inference rates that permit real-time implementation.

2. DEEP-ULM

2.1. Network architecture

We adopt a fully convolutional U-net style architecture [11] that consists of an encoder network which captures essential image information into a latent feature layer, and an expanding decoder network which maps this latent representation to precise localizations on a high-resolution grid. The encoder follows a contracting path made of 3 layer-blocks, each block comprising two 3x3 convolution layers with leaky rectified linear unit (ReLU) activations, and one 2x2 Max-pooling operation. We use leaky ReLUs rather than regular ReLUs across all convolution layers in the network to avoid inactive neurons/nodes that effectively decrease the model capacity [12]. In addition, batch normalization is used before all activations to boost the network's trainability by enabling higher learning rates and requiring less-strict hyper-parameter optimization [13]. The subsequent latent layer includes two 3x3 convolutional layers, followed by a dropout layer (probability 0.5) which randomly disables about 50% of the latent features during training [14]. This latent space is then transformed to a high-resolution localization image by the decoder. The decoder again consists of 3 blocks; the first two blocks encompassing two 5x5 deconvolution layers (transposed convolution) of which the second has an output stride of 2 rather than 1, followed by a 2x2 up-sampling layer which simply repeats the image rows and columns. The last block consists of two deconvolution layers, of which the second again has an output stride of 2, preceding another 5x5 convolution which maps the feature space to a single-channel image through a linear activation function. The full network effectively scales the input image dimensions up by a factor 8.

2.2. Training

We train the network using simulated pairs of high-resolution targets an corresponding ultrasound acquisitions. To this end, random microbubble concentrations between 0 and 260 microbubbles/cm^2 are generated, with each microbubble having a random location and backscatter intensity. The latter reflects the backscatter intensity variations of a polydisperse microbubble population imaged at various distances from the elevational beam axis, and ranged between 0.4 and 1 (a.u.).

The microbubble locations $(x_0, y_0) \in \mathcal{X}$ are then converted to a high-resolution target, I_{tar}, by assigning its backscatter value to an image pixel on the desired high-resolution grid. The corresponding radiofrequency (RF) signals are then obtained through the modulated PSF:

$$r(t, y) = \sum_{(x_i, y_i) \in \mathcal{X}} P(x_i, y_i | \phi) \sin\left[2\pi f_0 (t - x_i / c_0)\right], \quad (1)$$

where c_0 is the speed of sound, f_0 is the transmit fre-

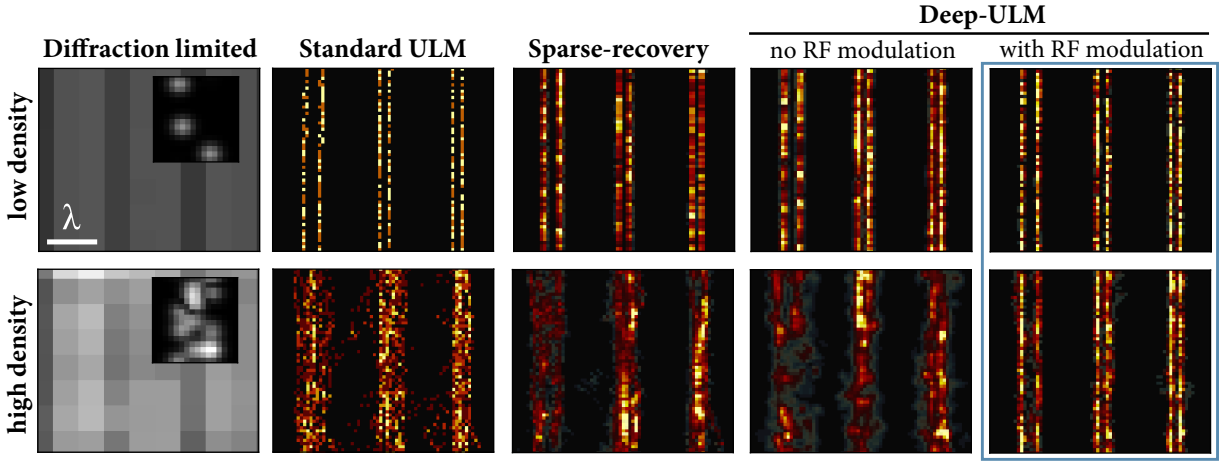

Fig. 2. Comparison of Deep-ULM against other ULM methods for contrast-enhanced ultrasound simulations of 6 closely spaced vessels, for low ($\approx$ 2 microbubbles/frame) and high ($\approx$ 15 microbubbles/frame) densities. Note that all methods are deployed on an image-domain sequence of frames.

quency and ϕ contains the parameters of the PSF P, estimated by manually pinpointing several isolated microbubbles and fitting a 2D anisotropic rotated Gaussian to the data. Uncertainty in this estimate was incorporated in the training procedure by introducing variance in the PSF parameters through a multiplicative random component, i.e. $\phi = \phi_m \left[1 + \mathcal{N}(\mu = 0, \sigma = 0.1)\right]$. The RF signals are then envelope detected and subsequently down-sampled to an 8 times courser grid than the high-resolution targets to yield the input patches:

$$I\left[n_x, n_y\right] = \sqrt{\left\{r(t,y)\right\}^2 + \left\{\mathcal{H}_t\left[r(t,y)\right]\right\}^2}\Big|_{t=\frac{n_x\Delta_x}{c_0}, y=n_y\Delta_x}, \tag{2}$$

where $\mathcal{H}_t\left[\cdot\right]$ denotes the Hilbert transform across t, and Δ_x is the pixel size. We then added white and colored background noise with relative standard deviations of 2% and 5%, respectively. Colored noise was produced by spatially filtering white noise with a 2D Gaussian having a standard deviation of 1.2 pixels.

Given this training data, we used the Adam optimizer [15] with learning rate 0.001, stochastically optimizing across batches of 128 patches to minimizing the following loss function:

$$\mathcal{L}(I, I_{tar}|\theta_{nn}) = \left\|f(I|\theta_{nn}) - G * I_{tar}\right\|_2^2 + \lambda \left\|f(I|\theta_{nn})\right\|_1, \tag{3}$$

where $f(I|\theta_{nn})$ is the nonlinear neural network function with parameters θ_{nn}, and λ is a regularization parameter that promotes network predictions that yield sparse images, and was (conservatively) set to 0.01. The operator G denotes a 2D Gaussian filter of which the standard deviation was set to one pixel. In practice, we observed that applying such a mild 2D filtering operation on the sparse target data improved training, ensuring proper back propagation of loss gradients in which large localization errors are penalized more than small deviations. The mean-squared-error-based regression strategy enables joint estimation of microbubble locations and their backscatter intensities. The latter is particularly useful to emphasize localizations near the elevational beam axis during image reconstruction.

Training (and inference) were implemented in Python using the Tensorflow backend (Google, Mountain View, CA), and run on a computation server, equipped with an NVIDIA Titan X Pascal GPU that has 12 GB of video memory.

3. REFERENCE ULM IMPLEMENTATIONS

3.0.1. Standard ULM

Standard ULM was implemented using a centroid localization approach, largely following the methodology described by Errico *et al.* [3]. The input images are upsampled by a factor 8 using the Lanczos kernel and subsequently deconvolved with a Gaussian low-pass filter that is based on the PSF. We threshold the deconvolved images at 50% of their 98^{th} percentiles, and perform a morphological opening operation to remove spurious peaks. From this we detect the local maxima, and select an area of 24x24 pixels (i.e. 3x3 pixels on the original data) around them to compute the image centroids.

3.0.2. Sparse-recovery ULM

Sparse-recovery ULM approaches the microbubble localization task as an inverse problem, by modelling each image frame as a superposition of translated and scaled PSFs according to microbubble locations and backscatter amplitudes on a high-resolution grid [8]. Assuming that the microbubbles are smaller than a pixel and sparsely distributed across the image,

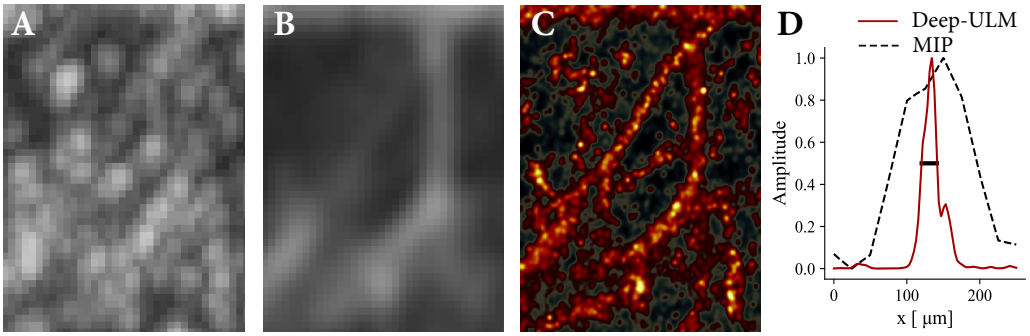

Fig. 3. *In-vivo* evaluation on a rat spinal cord. (A) Standard maximum intensity projection (MIP) image, (B) mean intensity image, and (C) Deep-ULM reconstruction. (D) Vessel intensity profile of Deep-ULM compared to MIP.

the following regularized inverse problem can be formulated by promoting a sparse solution through the addition of an ℓ_1 penalty:

$$\hat{\mathbf{i}}_{sr} = \arg\min_{\mathbf{i}_{sr}} \|\mathbf{P}\mathbf{i}_{sr} - \mathbf{i}\|_2^2 + \lambda_{sr} \|\mathbf{i}_{sr}\|_1 , \qquad (4)$$

where $\mathbf{i}_{sr}$ is the vectorized super-resolution frame on a high-resolution grid, $\mathbf{i}$ is the vectorized low-resolution image frame, and $\mathbf{P}$ is the measurement matrix in which each column is a shifted version of the PSF. To solve (4), we employed an optimized Fourier domain implementation of the Fast Iterative Shrinkage-Thresholding Algorithm (FISTA) [16],with a grid up-sampling factor of 8, and $\lambda = 0.01$.

4. RESULTS

Figure 1 shows the performance of Deep-ULM as compared to standard and sparse-recovery-based ULM for ultrasound simulations of randomly generated microbubble dispersions for various densities. For computing the recovered density, we only consider a microbubble detected if its localization was obtained within 30 μm (about 1/7th of the wavelength) of its true location. To determine the localization precision, each ULM-identified microbubble (i.e., also those beyond 30 μm) is associated to the closest ground-truth microbubble position, and their Euclidian distance is calculated. Deep-ULM outperforms both sparse recovery and standard ULM at both precision (localization error) and recall (recovered density), especially for higher densities.

We hypothesize that this performance gain originates from the neural network's ability to learn the image-domain implications of interference among backscattered RF signals of closely spaced microubbles. This situation is more likely to occur for the highest densities. We therefore now turn to investigating the impact of training the neural network with such knowledge of the interference patterns of overlapping RF-modulated point-spread-functions. Figure 2 displays super-resolution images (obtained by summing the sparse reconstructions for all frames) of a simulated ultrasound sequence (10 seconds, 100 Hz) of microbubbles flowing through closely spaced parallel vessels (3 pairs, separated by $\lambda/3$, $\lambda/4$, and $\lambda/5$), as obtained using Deep-ULM, trained with and without RF-modulation of the PSF. Interestingly, learning these interference patterns is indeed a crucial aspect of Deep-ULM's performance. We also compare these results against those obtained with sparse recovery and observe the same phenomena as for Deep-ULM when training without RF-modulation. Deep-ULM is moreover about 4 orders of magnitude faster than ULM by sparse recovery.

To show that Deep-ULM generalizes well beyond simulations, we display its application to an *in-vivo* contrast-enhanced ultrasound acquisition of a rodent spinal cord in Figure 3 [2]. Comparing panels A,B to C, we observe that Deep-ULM is able to resolve vessels that are not visible in the diffraction-limited images (A,B). This is further clarified in Figure 3D, where the line profiles of Deep-ULM and the maximum intensity projection are given for an isolated vessel.

5. CONCLUSION

In this work we show that the use of deep neural networks for ultrasound localization microscopy allows for improved performance as compared to a standard centroid-localization technique as well a sparse-recovery algorithm. The proposed method, Deep-ULM, yields higher precision and recall, in particular for areas with high contrast-agent densities, where the image PSFs of microbubbles overlap significantly. We show that adequate modeling of the radiofrequency interference patterns is crucial to achieve this performance gain with respect to the sparse-recovery method. Deep-ULM is moreover about 4 orders of magnitude faster, reconstructing about 1250 high-resolution patches per second. This makes real-time implementation feasible.

6. ACKNOWLEDGMENT

The authors would like to thank the NVIDIA Corporation and its academic GPU program for donating a Titan X Pascal.

7. REFERENCES

[1] J. Folkman, "Role of angiogenesis in tumor growth and metastasis," in *Seminars in oncology*, vol. 29, no. 6. Elsevier, 2002, pp. 15–18.

[2] Z. Z. Khaing, L. N. Cates, D. M. DeWees, A. Hannah, M. Bruce, and C. Hofstetter, "Intraoperative contrast-enhanced ultrasound to visualize hemodynamic changes after rodent spinal cord injury," *Journal of Neurosurgery: Spine*, 2018.

[3] C. Errico, J. Pierre, S. Pezet, Y. Desailly, Z. Lenkei, O. Couture, and M. Tanter, "Ultrafast ultrasound localization microscopy for deep super-resolution vascular imaging," *Nature*, vol. 527, no. 7579, pp. 499–502, 2015.

[4] O. M. Viessmann, R. J. Eckersley, K. Christensen-Jeffries, M. X. Tang, and C. Dunsby, "Acoustic super-resolution with ultrasound and microbubbles," *Physics in Medicine and Biology*, vol. 58, no. 18, pp. 6447–6458, sep 2013.

[5] K. Christensen-Jeffries, R. J. Browning, M.-X. Tang, C. Dunsby, and R. J. Eckersley, "In Vivo Acoustic Super-Resolution and Super-Resolved Velocity Mapping Using Microbubbles," *IEEE Transactions on Medical Imaging*, vol. 34, no. 2, pp. 433–440, feb 2015.

[6] D. Ackermann and G. Schmitz, "Detection and Tracking of Multiple Microbubbles in Ultrasound B-Mode Images," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 63, no. 1, pp. 72–82, jan 2016.

[7] E. Betzig, G. H. Patterson, R. Sougrat, O. W. Lindwasser, S. Olenych, J. S. Bonifacino, M. W. Davidson, J. Lippincott-Schwartz, and H. F. Hess, "Imaging intracellular fluorescent proteins at nanometer resolution," *Science*, vol. 313, no. 5793, pp. 1642–1645, 2006.

[8] R. Van Sloun, O. Solomon, Y. Eldar, H. Wijkstra, and M. Mischi, "Sparsity-driven super-resolution in clinical contrast-enhanced ultrasound," in *IEEE International Ultrasonics Symposium, IUS*, 2017.

[9] A. Bar-Zion, O. Solomon, C. Tremblay-Darveau, D. Adam, and Y. C. Eldar, "Sparsity-based Ultrasound Super-resolution Hemodynamic Imaging," *IEEE transactions on ultrasonics, ferroelectrics and frequency control*, in press, 2019.

[10] O. Solomon, R. J. G. Van Sloun, H. Wijkstra, and Y. C. Eldar, "Exploiting flow dynamics for super-resolution in contrast-enhanced ultrasound," *arXiv:1804.03134v1*.

[11] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.

[12] B. Xu, N. Wang, T. Chen, and M. Li, "Empirical Evaluation of Rectified Activations in Convolutional Network," may 2015.

[13] S. Ioffe and C. Szegedy, "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift," in *32nd International Conference on Machine Learning*, 2015, pp. 1–9.

[14] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A Simple Way to Prevent Neural Networks from Overfitting," *Journal of Machine Learning Research*, vol. 15, pp. 1929–1958, 2014.

[15] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[16] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM journal on imaging sciences*, vol. 2, no. 1, pp. 183–202, 2009.