



Deep Learning Models for Fast Ultrasound Localization Microscopy

Youn, Jihwan; Luijten, Ben ; Stuart, Matthias Bo; Eldar, Yonina C. ; van Sloun, Ruud J. G. ; Jensen, Jørgen Arendt

Published in:
2020 IEEE International Ultrasonics Symposium

Link to article, DOI:
[10.1109/IUS46767.2020.9251561](https://doi.org/10.1109/IUS46767.2020.9251561)

Publication date:
2020

Document Version
Peer reviewed version

[Link back to DTU Orbit](#)

Citation (APA):
Youn, J., Luijten, B., Stuart, M. B., Eldar, Y. C., van Sloun, R. J. G., & Jensen, J. A. (2020). Deep Learning Models for Fast Ultrasound Localization Microscopy. In *2020 IEEE International Ultrasonics Symposium* IEEE. <https://doi.org/10.1109/IUS46767.2020.9251561>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Deep Learning Models for Fast Ultrasound Localization Microscopy

1st Jihwan Youn
Dept. of Health Technology
Technical University of Denmark
Lyngby, Denmark
jihyoun@dtu.dk

2nd Ben Luijten
Dept. of Electrical Engineering
Eindhoven University of Technology
& *Philips Research*
Eindhoven, The Netherlands
w.m.b.luijten@tue.nl

3rd Matthias Bo Stuart
Dept. of Health Technology
Technical University of Denmark
Lyngby, Denmark
mbst@dtu.dk

4th Yonina C. Eldar
Dept. of Computer Science
and Applied Mathematics
Weizmann Institute of Science
Rehovot, Israel
yonina.eldar@weizmann.ac.il

5th Ruud J. G. van Sloun
Dept. of Electrical Engineering
Eindhoven University of Technology
& *Philips Research*
Eindhoven, The Netherlands
r.j.g.v.sloun@tue.nl

6th Jørgen Arendt Jensen
Dept. of Health Technology
Technical University of Denmark
Lyngby, Denmark
jaje@dtu.dk

Abstract—Ultrasound localization microscopy (ULM) can surpass the resolution limit of conventional ultrasound imaging. However, a trade-off between resolution and data acquisition time is introduced. For microbubble (MB) localization, centroid detection is commonly used. Therefore, low-concentrations of MBs are required to avoid overlapping point spread functions (PSFs), leading to a long data acquisition time due to the limited number of detectable MBs in an image frame. Recently, deep learning-based MB localization methods across high-concentration regimes have been proposed to shorten the data acquisition time. In this work, a data-driven encoder-decoder convolutional neural network (deep-ULM) and a model-based deep unfolded network embedding a sparsity prior (deep unfolded ULM) are analyzed in terms of localization accuracy and computational complexity. The results of simulated test data showed that both deep learning methods could handle overlapping PSFs better than centroid detection. Additionally, thanks to its model-based approach, deep unfolded ULM needed much fewer learning parameters and was computationally more efficient, and consequently achieved better generalizability than deep-ULM. It is expected that deep unfolded ULM will be more robust *in-vivo*.

Index Terms—deep unfolded network, high-concentration microbubble localization, model-based neural network, super-resolution ultrasound imaging, ultrasound localization microscopy

I. INTRODUCTION

Ultrasound localization microscopy (ULM) has shown great potential as a breakthrough in super-resolution ultrasound imaging (SRUS) by imaging microvasculature whose vessels are spaced closer than the resolution limit of conventional ultrasound imaging [1]–[6]. ULM is achieved by localizing gas-filled microbubbles (MBs) that are injected into the bloodstream and accumulating their centroids from multiple frames in an image. The resulting super-resolution images can be used

to diagnose early-stage cancer [7], ischemic kidney disease [8], and diabetes [9].

The fidelity of ULM depends on the number of detected MBs and their localization precision and sensitivity. Standard ULM methods ordinarily locate the centroids of isolated MBs, therefore, overlapping point spread functions (PSFs) need to be avoided. Diluted low-concentrations of MBs are commonly employed to minimize the overlapping PSFs for accurate localization. Even so, some overlapping PSFs still appear since MBs cannot easily be controlled after injection. The high-resolution of ULM is related to precise MB localization, so the overlapping MB PSFs are often rejected. However, low-concentrations of MBs and overlapping PSF rejection limit the number of detectable MBs in an image frame, and eventually require a long data acquisition time. To cope with this limitation, there have been efforts to achieve SRUS at high-concentrations of MBs [10]–[12].

Recently, several deep learning-based methods have been proposed to localize MBs across high-concentration regimes with overlapping PSFs [13]–[16]. Here we analyze two models and assess their capability in terms of localization accuracy and computational complexity. One approach is a data-driven encoder-decoder convolutional neural network (deep-ULM) [13], and the other is a model-based deep unfolded network that embeds a sparsity prior (deep unfolded ULM) [14]. These algorithms were compared along with the centroid detection method as baseline under challenging simulation scenarios.

II. METHOD

A. Data Generation

Ultrasound data were simulated in Field II pro [17]–[19] for training and evaluating deep learning models. The simulated data were chosen over measured data for training because it

is difficult to obtain ground-truth (i.e., MB positions) from the measured data. Radiofrequency (RF) channel data were simulated using a transducer modeled following the Verasonics L11-4v and a single cycle 6.9 MHz sinusoidal pulse. For one image frame, eleven plane waves with different angles were transmitted after placing ultrasound scatterers randomly in the region of interest. The RF channel data were then delay-and-sum beamformed with a dynamic apodization on a $\lambda/4$ grid, and the beamformed images were subsequently coherently compounded. The simulation parameters are presented in Table I. For the training set, 256 image frames were generated.

TABLE I
FIELD II SIMULATION PARAMETERS

TABLE I FIELD II SIMULATION PARAMETERS		
	Parameter	Value
Transducer	Transmit frequency	6.9 MHz
	Pitch	30 mm
	Element height	5 mm
	Element width	27 mm
	Number of elements	128
Imaging	Wave type	Plane
	Steering angles	$2 \cdot i^\circ, i \in \{-5, \dots, 5\}$
	F#	0.5
	# of elements in TX	128
	Apodization in TX	Hann window
	Apodization in RX	Hann window
Environment	Speed of sound	1480 m/s
	Field II sampling frequency	180 MHz

B. Deep learning-based Localization

Deep learning methods were designed to estimate MB positions from beamformed RF data. The MB positions (i.e., output) that were used to train networks were quantized and represented in a $\lambda/16$ image grid. The values of pixels containing MBs were set to one, and the others were zero. The higher-resolution grid was used than the beamformed images (i.e., input) to increase localization precision of estimated MBs.

The deep neural networks were trained by minimizing the difference between true MB positions and estimated MB positions using the ADAM [20] optimizer. The difference was captured by a loss function,

$$\mathcal{L}(\mathbf{x}, \mathbf{y}; \theta, \sigma) = \frac{1}{N} \sum_{i=1}^N \|G(\mathbf{y}_i; \sigma) - f(\mathbf{x}_i; \theta)\|_F^2, \quad (1)$$

where \mathbf{x}_i and \mathbf{y}_i are the i -th ultrasound image and MB positions, N is the number of samples, G is the 2-D Gaussian filtering with a standard deviation of σ , $f(\cdot; \theta)$ is the neural network function with learning parameters θ , and $\|\cdot\|_F$ is the Frobenius norm. Smoothing was applied to the true MB positions to provide larger gradients to ensure training stability.

1) *Deep-ULM*: Deep-ULM uses an encoder-decoder convolutional neural network (CNN), which is widely used for computer vision and image processing problems such as segmentation [21], [22] and image generation [23]. It mainly

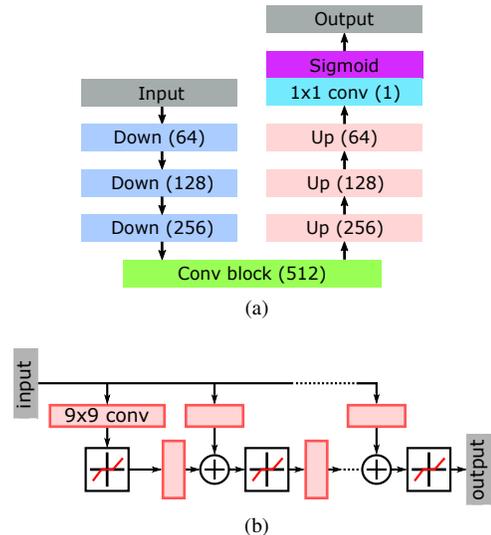


Fig. 1. Deep neural networks for MB localization. (a) Deep-ULM: encoder-decoder convolutional neural network and (b) Deep unfolded ULM: model-based neural network.

consists of *down*, *conv*, and *up* blocks, as shown in Fig. 1a. In the encoding path, the *down* blocks extract features using a series of convolution layers while downsampling the features from the previous layer by a factor of 2. In the decoding path, the MB positions are reconstructed based on the extracted features in the encoding path. To obtain the MB positions in the higher-resolution grid, the first *up* block upsamples the features by a factor of 2 and the other *up* blocks perform upsampling by a factor of 4. A detailed description of *down*, *conv*, and *up* blocks can be found in [15], [16].

The encoder-decoder CNN is a fully data-driven method and requires millions of learning parameters, which has a high chance of overfitting to the training data distributions. Therefore, considering the training data were simulated, deep-ULM may work well on the data simulated in the same way but not on data simulated differently or measured data.

2) *Deep unfolded ULM*: Deep unfolded ULM has been proposed to overcome the limitations of generalizability of deep-ULM [14], [24]. It solves ULM as a sparse coding problem, which can be formalized as

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n}, \quad (2)$$

where \mathbf{y} is the low-resolution MB ultrasound image, \mathbf{A} represents the PSF, \mathbf{x} is the MB positions on the high-resolution grid, and \mathbf{n} is noise.

It can be assumed that \mathbf{x} is sparse because the MB positions are represented in a higher-resolution grid. The optimal \mathbf{x} can then be estimated by solving an optimization problem with a sparsity prior, i.e., the ℓ_1 -penalty:

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x}} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \lambda \|\mathbf{x}\|_1, \quad (3)$$

where λ is the regularization coefficient. The problem (3) can be solved using the proximal gradient method. However, such

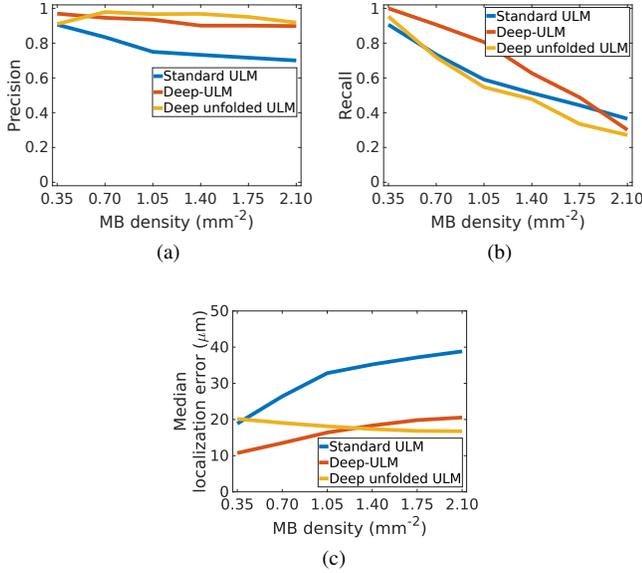


Fig. 2. Comparison of the methods on test sets simulated by placing scatterers randomly at different MB densities where (a) is precision, (b) is recall, and (c) is the median of localization error.

iterative methods may take a long time to converge and their performance highly depends on the hyper-parameters such as the regularization coefficient, the PSF model, and the step size at each iteration, so that empirical tuning is necessary.

Deep unfolded ULM solves the optimization problem using Learned ISTA (LISTA) [25]. LISTA is constructed by unfolding the iteration part as a K -layer neural network, as shown in Fig. 1b. In this work, a 10-layer network was used. LISTA is fast and tuning-free since the iteration is not required and the hyper-parameters, which need to be tuned in the proximal gradient scheme, are embedded in the model, as learning parameters. That allows more robust MB localization by learning more diverse PSF models, unlike the proximal gradient methods which require a specific PSF model [26]. Deep unfolded ULM does not include upsampling in the model, so the input data were upsampled by a factor of 4 before being applied to the network.

III. RESULTS

The trained deep learning models were compared with standard ULM (centroid detection) on two different simulated test sets. One test set comprised independent frames simulated in the same way as the training data at different MB densities. The other test set was composed of consecutive frames simulated using a pair of closely spaced parallel tubes in which scatterers flowed in the opposite directions to each other.

A. Randomly Placed Scatterers

The capability of the models at various MB densities was investigated using a randomly placed scatterer test set. Three evaluation metrics were used: precision, recall, and the median of localization error, defined as

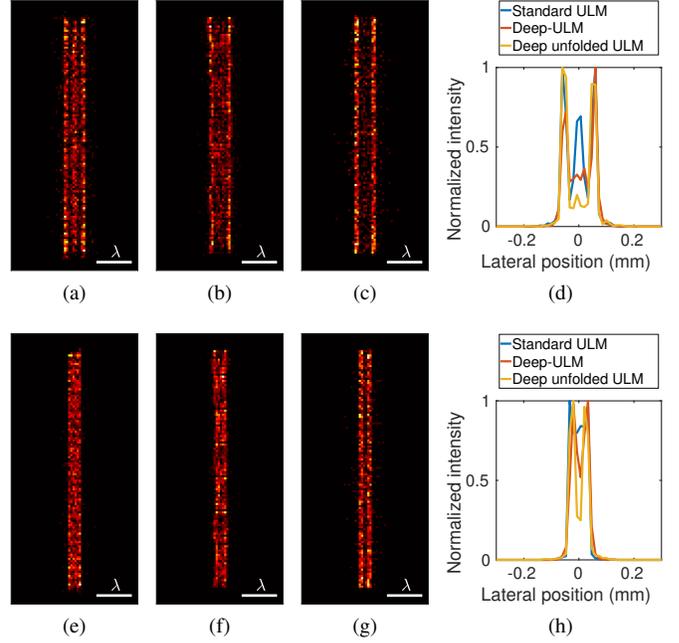


Fig. 3. Comparison of the methods on the simulation of a pair of parallel tubes. (a) - (d) are the results of tubes separated by $\lambda/2$ and (e) - (h) are the results of tubes separated by $\lambda/4$, where (a), (e) are stand ULM, (b), (f) are deep-ULM, (c), (g) are deep unfolded ULM, and (d), (h) are the intensity profile of each method along the lateral direction.

$$\text{precision} = \frac{TP}{TP + FP}, \quad \text{recall} = \frac{TP}{TP + FN}, \quad (4)$$

where TP is the number of true positive (true detection), FP is the number of false positive (false detection), and FN is the number of false negative (missed target).

The results are shown in Fig. 2. For standard ULM, all three metrics got worse as the density increases. At higher densities, a larger number of overlapping PSFs appeared, so that MB localization became more challenging. On the other hand, deep-ULM was not degraded as much as standard ULM at the high densities because deep learning models can deal with a certain degree of overlapping PSFs in MB localization. Deep unfolded ULM achieved comparable precision and localization uncertainty to deep-ULM, but the recall was not as good as deep-ULM. This shows that deep-ULM can achieve better performance on the data set that have the same distribution as the training set, i.e., randomly placed scatterer data, by exploiting a larger number of learning parameters.

B. Parallel Tubes

For more realistic experiments, 1024 consecutive frames were simulated using a pair of parallel tubes separated by $\lambda/2$ and $\lambda/4$. The resulting super-resolution images of each method and their MB intensity profile along the lateral direction are shown in Fig. 3.

The limitation of standard ULM at a high MB density is clearly shown. In the middle of the tubes where no MBs were supposed to be detected, a larger number of false detection

TABLE II
SUMMARY OF DEEP-ULM AND DEEP UNFOLDED ULM

Scheme	Deep-ULM	Deep unfolded ULM
	Fully data-driven	Model-based data-driven
# of learning parameters	5 998 785	1735
Floating point operations (FLOPs)	788 259 839	3462
Generalizability to out of data distributions	Not good	Good

appeared and high MB intensity along the lateral direction was achieved. Both deep learning models worked better than standard ULM and deep unfolded ULM resulted in better-resolved images with much fewer parameters. This shows that deep unfolded ULM achieves better generalization to various data distributions that are different from the training data, consistent with [24], [26].

IV. DISCUSSION

A summary of deep-ULM (a fully data-driven method) and deep unfolded ULM (a model-based data-driven method) are shown in Table II. Deep unfolded ULM, required much fewer parameters and operations while achieving comparable results to deep-ULM. The model-based approach allowed not only to reduce the number of learning parameters and operations, but also to achieve better generalizability to out of training data distributions. Deep unfolded ULM showed better performance on the test set of parallel tubes, which had scatterers located inside the tubes contrary to the training data which had randomly placed scatterers. Under the better generalizability, deep unfolded ULM will possibly be able to achieve more robust MB localization than deep-ULM on measured data.

ACKNOWLEDGMENT

We gratefully acknowledge the support of NVIDIA Corporation with the donation of the Titan V Volta GPU used for this research.

REFERENCES

- [1] O. Couture, B. Besson, G. Montaldo, M. Fink, and M. Tanter, "Microbubble ultrasound super-localization imaging (MUSLI)," in *Proc. IEEE Ultrason. Symp.*, 2011, pp. 1285–1287.
- [2] O. M. Viessmann, R. J. Eckersley, K. Christensen-Jeffries, M. X. Tang, and C. Dunsby, "Acoustic super-resolution with ultrasound and microbubbles," *Phys. Med. Biol.*, vol. 58, pp. 6447–6458, 2013.
- [3] M. A. O'Reilly and K. Hynynen, "A super-resolution ultrasound method for brain vascular mapping," *Med. Phys.*, vol. 40, no. 11, pp. 110701–7, 2013.
- [4] 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, pp. 499–502, November 2015.
- [5] K. Christensen-Jeffries, R. J. Browning, M. Tang, C. Dunsby, and R. J. Eckersley, "In vivo acoustic super-resolution and super-resolved velocity mapping using microbubbles," *IEEE Trans. Med. Imag.*, vol. 34, no. 2, pp. 433–440, February 2015.
- [6] K. Christensen-Jeffries, O. Couture, P. A. Dayton, Y. C. Eldar, K. Hynynen, F. Kiessling, M. O'Reilly, G. F. Pinton, G. Schmitz, M. Tang *et al.*, "Super-resolution ultrasound imaging," *Ultrasound Med. Biol.*, vol. 46, no. 4, pp. 865–891, 2020.

- [7] F. Lin, S. E. Shelton, D. Espindola, J. D. Rojas, G. Pinton, and P. A. Dayton, "3-D ultrasound localization microscopy for identifying microvascular morphology features of tumor angiogenesis at a resolution beyond the diffraction limit of conventional ultrasound," *Theranostics*, vol. 7, no. 1, pp. 196–204, 2017.
- [8] S. B. Andersen, C. A. V. Hoyos, I. Taghavi, F. Gran, K. L. Hansen, C. M. Sørensen, and J. A. J. M. B. Nielsen, "Super-resolution ultrasound imaging of rat kidneys before and after ischemia-reperfusion," in *Proc. IEEE Ultrason. Symp.*, 2019, pp. 1–4.
- [9] D. Ghosh, J. Peng, K. Brown, S. Sirsi, C. Mineo, P. W. Shaul, and K. Hoyt, "Super-resolution ultrasound imaging of skeletal muscle microvascular dysfunction in an animal model of type 2 diabetes," *J. Ultrasound Med.*, vol. 38, no. 10, pp. 2589–2599, 2019.
- [10] A. Bar-Zion, C. Tremblay-Darveau, O. Solomon, D. Adam, and Y. C. Eldar, "Fast vascular ultrasound imaging with enhanced spatial resolution and background rejection," *IEEE Trans. Med. Imag.*, vol. 36, no. 1, pp. 169–180, 2016.
- [11] A. Bar-Zion, O. Solomon, C. Tremblay-Darveau, D. Adam, and Y. C. Eldar, "Sushi: Sparsity-based ultrasound super-resolution hemodynamic imaging," *IEEE Trans. Ultrason., Ferroelec., Freq. Contr.*, vol. 65, no. 12, pp. 2365–2380, 2018.
- [12] O. Solomon, R. J. G. van Sloun, H. Wijkstra, M. Mischi, and Y. C. Eldar, "Exploiting flow dynamics for super-resolution in contrast-enhanced ultrasound," *IEEE Trans. Ultrason., Ferroelec., Freq. Contr.*, vol. 60, no. 10, pp. 1573–1586, 2019.
- [13] R. J. G. van Sloun, O. Solomon, M. Bruce, Z. Z. Khaing, H. Wijkstra, Y. C. Eldar, and M. Mischi, "Super-resolution ultrasound localization microscopy through deep learning," *arXiv:1804.07661v2 [eess.SP]*, 2018.
- [14] R. J. G. van Sloun, R. Cohen, and Y. C. Eldar, "Deep learning in ultrasound imaging," *IEEE Proc.*, vol. 108, no. 1, pp. 11–29, 2020.
- [15] J. Youn, M. L. Ommen, M. B. Stuart, E. V. Thomsen, N. B. Larsen, and J. A. Jensen, "Ultrasound multiple point target detection and localization using deep learning," in *Proc. IEEE Ultrason. Symp.*, 2019.
- [16] J. Youn, M. L. Ommen, M. B. Stuart, E. V. Thomsen, N. B. Larsen, and J. A. Jensen, "Detection and localization of ultrasound scatterers using convolutional neural networks," *IEEE Trans. Med. Imag.*, 2020.
- [17] J. A. Jensen and N. B. Svendsen, "Calculation of pressure fields from arbitrarily shaped, apodized, and excited ultrasound transducers," *IEEE Trans. Ultrason., Ferroelec., Freq. Contr.*, vol. 39, no. 2, pp. 262–267, 1992.
- [18] J. A. Jensen, "Field: A program for simulating ultrasound systems," *Med. Biol. Eng. Comp.*, vol. 10th Nordic-Baltic Conference on Biomedical Imaging, Vol. 4, Supplement 1, Part 1, pp. 351–353, 1996.
- [19] —, "A multi-threaded version of Field II," in *Proc. IEEE Ultrason. Symp.* IEEE, 2014, pp. 2229–2232.
- [20] D. Kingma and L. Ba, "Adam: A method for stochastic optimization," *arXiv:1412.6980 [cs.LG]*, 2015.
- [21] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention*, 2015, pp. 234–241.
- [22] V. Badrinarayanan, A. Kendall, and R. Cipolla, "Segnet: A deep convolutional encoder-decoder architecture for image segmentation," *V. Badrinarayanan and A. Kendall and R. Cipolla*, vol. 39, no. 12, pp. 2481–2495, 2017.
- [23] P. Isola, J. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," *arXiv:1611.07004v3 [cs.CV]*, 2016.
- [24] V. Monga, Y. Li, and Y. C. Eldar, "Algorithm unrolling: Interpretable, efficient deep learning for signal and image processing," *arXiv:1912.10557v3 [eess.IV]*, 2019.
- [25] K. Gregor and Y. LeCun, "Learning fast approximations of sparse coding," in *Int. Conf. Machine Learning*, 2010, pp. 399–406.
- [26] G. Dardikman-Yoffe and Y. C. Eldar, "Learned sparcom: Unfolded deep super-resolution microscopy," *arXiv:2004.09270v2 [eess.IV]*, 2020.