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

# Ultrasound Multiple Point Target Detection and Localization using Deep Learning

Jihwan Youn, Martin Lind Ommen, Matthias Bo Stuart, Erik Vilain Thomsen,
Niels Bent Larsen, Jørgen Arendt Jensen

Department of Health Technology, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark

***Abstract*—Super-resolution imaging (SRI) can achieve sub-wavelength resolution by detecting and tracking intravenously injected microbubbles (MBs) over time. However, current SRI is limited by long data acquisition times since the MB detection still relies on diffraction-limited conventional ultrasound images. This limits the number of detectable MBs in a fixed time duration. In this work, we propose a deep learning-based method for detecting and localizing high-density multiple point targets from radio frequency (RF) channel data. A Convolutional Neural Network (CNN) was trained to return confidence maps given RF channel data, and the positions of point targets were estimated from the confidence maps. RF channel data for training and evaluation were simulated in Field II by placing point targets randomly in the region of interest and transmitting three steered plane waves. The trained CNN achieved a precision and recall of 0.999 and 0.960 on a simulated test dataset. The localization errors after excluding outliers were within $\pm\,46\,\mu$m and $\pm\,27\,\mu$m in the lateral and axial directions. A scatterer phantom was 3-D printed and imaged by the Synthetic Aperture Real-time Ultrasound System (SARUS). On measured data, a precision and recall of 0.976 and 0.998 were achieved, and the localization errors after excluding outliers were within $\pm\,101\,\mu$m and $\pm\,75\,\mu$m in the lateral and axial directions. We expect that this method can be extended to highly concentrated microbubble (MB) detection in order to accelerate SRI.**

## I. Introduction

Super-resolution imaging (SRI), often referred to as ultrasound localization microscopy (ULM), has demonstrated that it is possible to surpass the diffraction limit of conventional ultrasound imaging. Microvessels laying closer than a half-wavelength apart have been resolved by deploying microbubbles (MBs) as a contrast agent and using SRI [1]–[5]. The centroids of individual MBs can be easily found as MB echoes are much stronger than surrounding tissues when insonified, and their sizes are much smaller than a wavelength. Sub-wavelength imaging is achieved by accumulating the detected MB positions over time, revealing the fine structure of the microvasculature.

The MB detection in SRI, however, is still diffraction-limited because it is performed in conventional ultrasound images which are commonly formed by delay-and-sum (DAS) beamforming [6]. For accurate and reliable detection and localization, the MBs need to be more than a wavelength apart to avoid the overlaps of MB point spread functions (PSFs). Diluted concentrations of MBs are commonly used to satisfy this criteria as the behavior of MBs is hard to control. The number of detectable MBs, therefore, is constrained and this leads to very long data acquisition times in order to map the entire microvasculature.

In this work, we propose a deep learning-based method for detecting and localizing multiple ultrasound point targets. The method especially aims to identify high-density point targets whose PSFs are overlapping, by feeding radio frequency (RF) channel data directly as input. A fully convolutional neural network (CNN) was designed to return 2-D confidence maps given RF channel data. The pixel values of the confidence maps correspond to the confidence of point targets existing in the pixels. The point target positions were extracted from the confidence maps by identifying local maxima. The CNN was trained and evaluated using simulated RF channel data. To further investigate the method on measured data, a phantom experiment was performed using a 3-D printed PEGDA $700\,\mathrm{g/mol}$ hydrogel phantom [7].

## II. Method

### A. Simulated Dataset

*1) RF channel data:* The Field II ultrasound simulation program [8], [9] was used to simulate RF channel data for generating a training and a test datasets. The datasets were composed of a certain number of frames. One frame was created by transmitting three steered plane waves after placing 100 point targets randomly within a region of $6.4 \times 6.4\,\mathrm{mm}^2$ (an average target density of $2.44\,\mathrm{mm}^{-2}$) where the center was $18\,\mathrm{mm}$ away from a transducer. The transducer was modeled after a commercial 192-element linear array, and the measured impulse response [10], [11] was applied to make the RF data as close to real measured data as possible. The parameters used in simulation are listed in Table I.

The simulated raw RF data were not beamformed but delayed, based on the time-of-flight calculated by

$$\tau_i(x,z) = \left(\sqrt{(x-x_i)^2 + z^2} + z\right)/c \qquad (1)$$

where $\tau_i$ is the time-of-flight of the $i$-th transmission, $(x, z)$ is the point, $x_i$ is the center of the $i$-th transmission aperture, and $c$ is the speed of sound. The delayed RF data were then sampled to have the same number of samples as that of confidence maps along the axial direction. The size of resulting RF data for one frame was $256 \times 64 \times 3$.

978-1-7281-4595-2/19/$31.00 ©2019 IEEE

TABLE I
RF CHANNEL DATA SIMULATION PARAMETERS

| Category | Parameter | Value |
|---|---|---|
| Transducer | Center frequency | $5.2\,\mathrm{MHz}$ |
| | Pitch | $0.20\,\mathrm{mm}$ |
| | Element width | $0.18\,\mathrm{mm}$ |
| | Element height | $6\,\mathrm{mm}$ |
| | Number of elements | 192 |
| Imaging | Number of TX elements | 32 |
| | Number of RX elements | 64 |
| | Steering angles | $-15°$, $0°$, $15°$ |
| Environment | Speed of sound | $1480\,\mathrm{m/s}$ |
| | Field II sampling frequency | $120\,\mathrm{MHz}$ |
| | RF data sampling frequency | $29.6\,\mathrm{MHz}$ |
| Scatterer | Number of scatterers | 100 |
| | Lateral position range | $(-3.2, 3.2)\,\mathrm{mm}$ |
| | Axial position range | $(14.8, 21.2)\,\mathrm{mm}$ |

*2) Confidence Map:* Non-overlapping Gaussian confidence maps were used as labels for training CNNs. Initially, binary confidence maps were created, where pixel values of one indicated a point target and the remaining pixel values were zero. A $21 \times 21$ Gaussian filter with a standard deviation of six was then applied at each point target position in the binary confidence maps. The filter values from the targets will be overlapped when some targets are closer than a half of the filter size in the confidence maps. In that case, the maximum value at each pixel location was taken. This maintained local maxima at target positions as opposed to the overlapping PSFs of DAS beamforming, and enabled the CNN to resolve targets closer than the diffraction limit.

The pixel size of the confidence maps was set to $25\,\mu\mathrm{m}$, and the image size of them became $256 \times 256$, given the pixel size and the region of interest.

### B. Convolutional Neural Network

*1) Network Architecture:* The proposed CNN is adapted from U-Net [12] which has an encoder-decoder structure. The feature maps are downsampled while the number of feature maps increases in the encoding path. Then, the feature maps are upsampled to their original size while the number of feature maps decreases in the decoding path. U-Net has a large receptive field, an effective input size that is covered by a convolution operation in an unit, for the sake of this structure. This is beneficial because a partial view of RF data is not enough to determine point target existence.

A detailed CNN architecture is illustrated in Fig. 1. Convolution and rectified linear unit (ReLU) layers in U-Net were replaced with pre-activation residual units (Fig. 1a) [13]. The pre-activation residual units ease optimization problem by introducing shortcuts, thereby improving performance. The proposed CNN (Fig. 1e) mainly consisted of four *down-blocks* (Fig. 1b), one *conv-block* (Fig. 1c), and four *up-blocks* (Fig. 1d). The skip-connections in U-Net was removed since it hindered the training. Instead, CoordConv [14] was added to transfer spatial information over convolution layers. Dropout [15] was attached after the shortcut in residual blocks for regularization. For pooling and unpooling, strided convolution and pixel shuffle [16] were chosen, respectively. Leaky rectified linear units (Leaky ReLU) [17] were applied as non-linear activation to avoid dying ReLU problem causing nonactivated units.

*2) Training Details:* The CNN was trained by minimizing the mean squared error (MSE) between true confidence maps and CNN outputs. The training dataset consisted of a total of $10,240$ frames. The kernel weights were initialized with orthogonal initialization [18] and optimized with ADAM [19] by setting $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\epsilon = 10^{-7}$. The initial learning rate was $10^{-4}$ and it was halved at every 100 epoch while limiting the minimum learning rate to $10^{-6}$. The number of epochs was 600 and the mini-batch size was 32.

### C. 3-D Printed Scatterer Phantom

A PEGDA $700\,\mathrm{g/mol}$ hydrogel scatterer phantom [7] was 3-D printed to investigate the proposed method on measured data. The phantom contained water-filled cavities which acted as scatterers. A total of 100 scatterers were placed on a $10 \times 10$ grid with a spacing of $518\,\mu\mathrm{m}$ in the lateral direction and $342\,\mu\mathrm{m}$ in the axial direction, as illustrated in Fig. 2.

The 3-D printed phantom was scanned by the Synthetic Aperture Real-time Ultrasound System (SARUS) [20] to acquire RF channel data. The same imaging scheme and transducer described in Table I were used. The phantom was placed on a motion stage and scanned at different positions by moving the motion stage at a step of $50\,\mu\mathrm{m}$ in the lateral direction. A total of 33 frames were obtained.

## III. RESULTS

### A. Simulation Experiment

The trained CNN was initially evaluated on a simulated test dataset. It was simulated in the same way as the training dataset in Field II, and consisted of 3,840 frames. In Fig. 3, the result of applying the CNN method to a test frame is compared with simply using the conventional DAS beamforming on the same frame. The CNN method was able to identify highly concentrated point targets while the DAS beamforming failed due to the overlapping PSFs. Full width at half maximum (FWHM) of the DAS beamforming at a depth of 18 mm was $387\,\mu\mathrm{m}$ ($1.36\ \lambda$) in the lateral direction and $140\,\mu\mathrm{m}$ ($0.49\ \lambda$) in the axial direction.

The CNN's capability to detect and localize point targets were quantitatively evaluated. Detection was measured by precision and recall that are defined by

$$\text{Precision} = \frac{TP}{TP + FP} \tag{2}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{3}$$

where $TP$ is the number of true positives, $FP$ is the number of false positives, and $FN$ is the number of false negatives. The positive and negative detections were determined by comparing estimated target positions with true target positions based on their pair-wise distances. The CNN method achieved

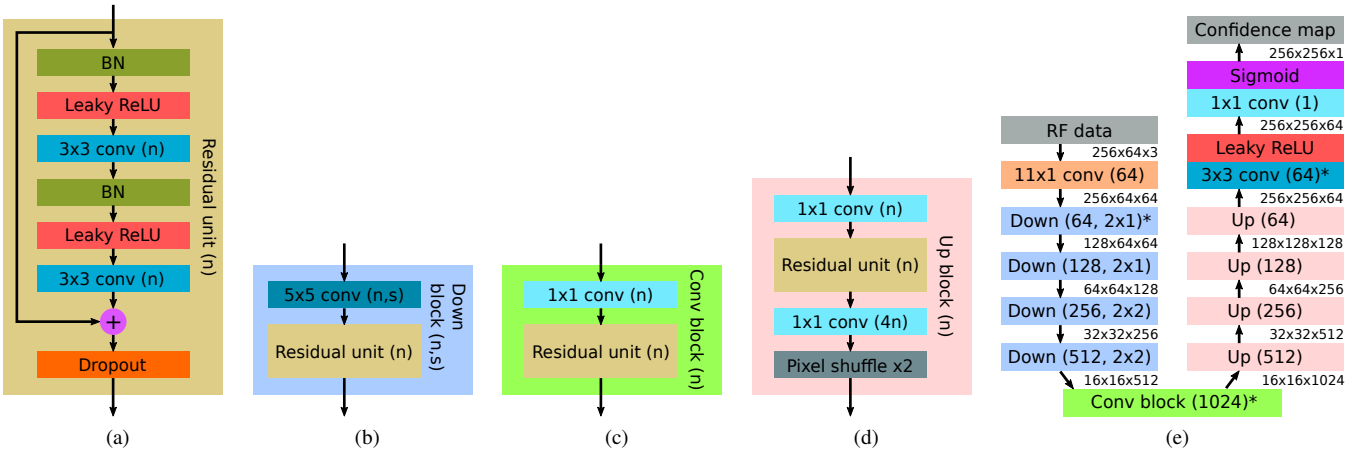


Fig. 1. The proposed CNN architecture and its components. (a) residual unit, (b) down-block, (c) conv-block, (d) up-block, and (e) the network overview. n and s in the parenthesis are the number of kernels and stride. The asterisk in (e) indicates that its first convolution in the block is CoordConv. The three numbers between blocks in (e) represent feature map size in the order of height, width, and the number of feature maps.

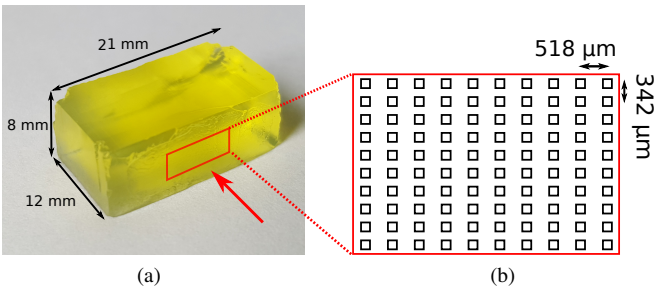


Fig. 2. Fabricated 3-D scatterer phantom: (a) photograph of the phantom and (b) 100 scatterers placed in a $10 \times 10$ grid.

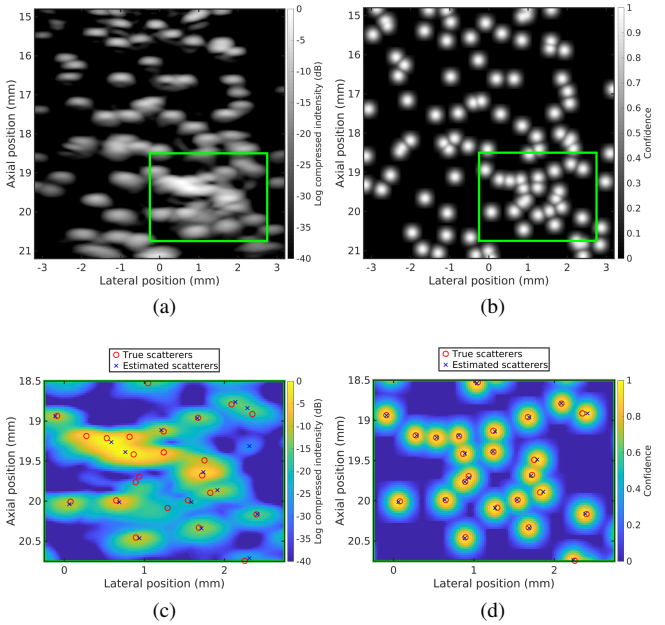


Fig. 3. Comparison of point target detection between DAS beamforming and CNN on a simulated test data using three steered plane wave transmissions. (a) DAS beamformed B-mode image, (b) confidence map returned from CNN, (c) true and estimated scatterer positions in the green square region of (a), and (d) true and estimated scatterer positions in the green square region of (b)

a precision and recall of 0.999 and 0.960, while DAS beamforming achieved a precision and recall of 0.986 and 0.756.

Localization uncertainties in the lateral and axial position were calculated using the positive detections, and is illustrated using a box-and-whisker plot in Fig. 4a. The bottom and top edges of the blue box indicate the 25th ($q_1$) and 75th percentiles ($q_3$) and the center red edge indicates the median. The vertically extended line from the box (whisker) indicates the range of inliers which are smaller than $q_3 + 1.5 \times (q_3 - q_1)$ and greater than $q_1 - 1.5 \times (q_3 - q_1)$. The inliers were within $\pm 46\,\mu\text{m}$ ($0.16\lambda$) in the lateral direction and $\pm 27\,\mu\text{m}$ ($0.09\lambda$) in the axial direction.

### *B. Phantom Experiment*

The CNN trained for the simulation experiment was not effective on the measured data because the scatterers in the phantom are not infinitesimally small point targets. The ultrasound beam is actually scattered twice at each scatterer in the phantom. Therefore, the RF data in the training dataset were simulated a second time by modeling a target using two points. In addition, the first scattering was phase reversed since the acoustic impedance is higher in the phantom than in the water inside the targets.

A new CNN was trained using the modified training dataset, and it successfully identified scatterers from the measured data as shown in Fig. 5. The achieved precision and recall were 0.976 and 0.998. The inliers were within $\pm 101\,\mu\text{m}$ ($0.33\lambda$) in the lateral direction and $\pm 75\,\mu\text{m}$ ($0.25\lambda$) in the axial direction, as illustrated in Fig. 4b.

## IV. CONCLUSION

A CNN-based ultrasound multiple point target detection and localization method was demonstrated. The CNN was trained to learn a mapping from RF channel data to non-overlapping Gaussian confidence maps, and point target positions were

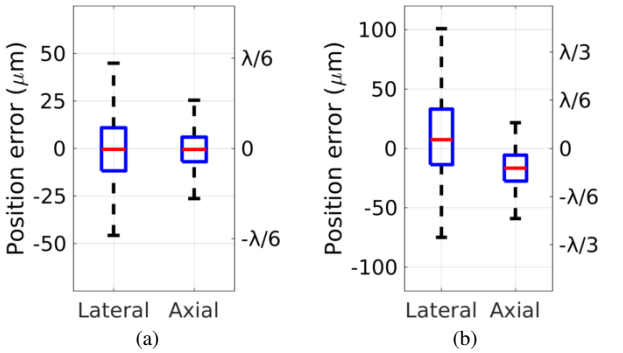


Fig. 4. Localization uncertainty in the lateral and axial direction measured (a) on the simulated test dataset and (b) on the measured phantom data.

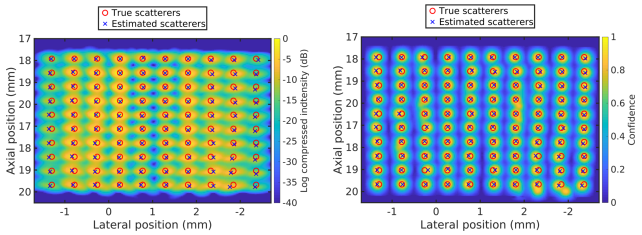


Fig. 5. Comparison of scatterer detection between DAS beamforming and CNN on phantom data using three steered plane wave transmissions. (a) DAS beamformed B-mode image and (b) confidence map returned from CNN with true and estimated scatterer positions

estimated from the confidence maps by identifying local maxima. The non-overlapping Gaussian confidence maps were introduced to relax the sparsity of binary confidence maps while maintaining local maxima as target positions. The CNN method resolved point targets closer than the diffraction limit, whereas DAS beamforming failed as shown in Fig. 3.

It is also shown that the CNN method is applicable to real-world data, as well as simulated data, through the phantom experiment. It is notable that the training was performed solely using simulated data because it is nearly impossible to obtain a large number of measurements with ground truth for these kinds of work. It was also imperative to employ the measured impulse response and model targets following realistic physical modeling in the simulation.

We expect that this method can be extended to MB detection and potentially shorten the data acquisition time of SRI by detecting a greater number of MBs in a shorter amount of time.

## 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. C. 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. 110 701–7, 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] F. L. Thurstone and O. T. von Ramm, "A new ultrasound imaging technique employing two-dimensional electronic beam steering," in *Acoustical Holography*, P. S. Green, Ed., vol. 5. New York: Plenum Press, 1974, pp. 249–259.

[7] M. L. Ommen, M. Schou, R. Zhang, C. A. V. Hoyos, J. A. Jensen, N. B. Larsen, and E. V. Thomsen, "3D printed flow phantoms with fiducial markers for super-resolution ultrasound imaging," in *Proc. IEEE Ultrason. Symp.*, 2018, pp. 1–4.

[8] 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.

[9] 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.

[10] ——, "Safety assessment of advanced imaging sequences, II: Simulations," *IEEE Trans. Ultrason., Ferroelec., Freq. Contr.*, vol. 63, no. 1, pp. 120–127, 2016.

[11] B. G. Tomov, S. E. Diederichsen, E. V. Thomsen, and J. A. Jensen, "Characterization of medical ultrasound transducers," in *Proc. IEEE Ultrason. Symp.*, 2018, pp. 1–4.

[12] 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.

[13] K. He, X. Zhang, S. Ren, and J. Sun, "Identity mappings in deep residual networks," in *Eur. Conf. Computer Vision*, 2016, pp. 630–645.

[14] R. Liu, J. Lehman, P. Molino, F. P. Such, E. Frank, A. Sergeev, and J. Yosinski, "An intriguing failing of convolutional neural networks and the coordconv solution," in *Neural Information Processing Systems*, 2018, pp. 9605–9616.

[15] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, pp. 1929–1958, 2014.

[16] W. Shi, J. Caballero, F. Huszár, J. Totz, A. P. Aitken, R. Bishop, D. Rueckert, and Z. Wang, "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network," in *IEEE Conf. Computer Vision and Pattern Recognition*, 2016, pp. 1874–1883.

[17] A. L. Maas, A. Y. Hannun, and A. Y. Ng, "Rectifier nonlinearities improve neural network acoustic models," in *ICML Workshop on Deep Learning for Audio, Speech, and Language Processing*, 2013.

[18] A. M. Saxe, J. L. McClelland, and S. Ganguli, "Exact solutions to the nonlinear dynamics of learning indeep linear neural networks," `arXiv:1312.6120v3 [cs.NE]`, 2013.

[19] D. Kingma and L. Ba, "Adam: A method for stochastic optimization," `arXiv:1412.6980 [cs.LG]`, 2015.

[20] J. A. Jensen, H. Holten-Lund, R. T. Nilsson, M. Hansen, U. D. Larsen, R. P. Domsten, B. G. Tomov, M. B. Stuart, S. I. Nikolov, M. J. Pihl, Y. Du, J. H. Rasmussen, and M. F. Rasmussen, "SARUS: A synthetic aperture real-time ultrasound system," *IEEE Trans. Ultrason., Ferroelec., Freq. Contr.*, vol. 60, no. 9, pp. 1838–1852, 2013.