

Deep Learning-Based Microbubble Localization for Ultrasound Localization Microscopy

Xi Chen¹, Matthew R. Lowerison², Zhijie Dong³, *Graduate Student Member, IEEE*,
Aiguo Han¹, *Member, IEEE*, and Pengfei Song¹, *Senior Member, IEEE*

Abstract—Ultrasound localization microscopy (ULM) is an emerging vascular imaging technique that overcomes the resolution-penetration compromise of ultrasound imaging. Accurate and robust microbubble (MB) localization is essential for successful ULM. In this study, we present a deep learning (DL)-based localization technique that uses both Field-II simulation and *in vivo* chicken embryo chorioallantoic membrane (CAM) data for training. Both radio frequency (RF) and in-phase and quadrature (IQ) data were tested in this study. The simulation experiment shows that the proposed DL-based localization was able to reduce both missing MB localization rate and MB localization error. In general, RF data showed better performance than IQ. For the *in vivo* CAM study with high MB concentration, DL-based localization was able to reduce the vessel MB saturation time by more than 50% compared to conventional localization. In addition, we propose a DL-based framework for real-time visualization of the high-resolution microvasculature. The findings of this article support the use of DL for more robust and faster MB localization, especially under high MB concentrations. The results indicate that further improvement could be achieved by incorporating temporal information of the MB data.

Index Terms—Deep learning (DL), microbubble (MB), super-resolution, ultrasound imaging, ultrasound localization microscopy (ULM).

I. INTRODUCTION

AS A diffraction-limited imaging modality, the performance of ultrasound imaging has long been limited by the classical tradeoff between imaging resolution and penetration depth. For decades, there has been a long quest to break the diffraction limit of ultrasound [1]. For example, various attempts had been made to super-focus the acoustic beams through negative refraction generated by acoustic metamaterials [2]–[4]. In a medium with random and inhomogeneous echogenicity, the diffraction limit may be overcome by

Manuscript received January 18, 2022; accepted February 12, 2022. Date of publication February 16, 2022; date of current version March 28, 2022. This work was supported in part by the National Cancer Institute under Grant R00CA214523 and in part by the National Institute of Biomedical Imaging and Bioengineering of the National Institutes of Health under Grant R21EB030072. (Corresponding author: Pengfei Song.)

The authors are with the Beckman Institute for Advanced Science and Technology and the Department of Electrical and Computer Engineering, University of Illinois Urbana-Champaign, Urbana, IL 61801 USA (e-mail: songp@illinois.edu).

This article has supplementary downloadable material available at <https://doi.org/10.1109/TUFFC.2022.3152225>, provided by the authors.

Digital Object Identifier 10.1109/TUFFC.2022.3152225

retransmission and refocusing of received signal through a time-reversal mirror [5]. As an acoustic analogy to structured illumination microscopy (SIM), acoustical structured illumination was recently proposed to surpass the resolution limit by generating a series of known patterns with the transducer, which enables encoding of high-resolution information of the observed image [6]. The lateral resolution of ultrasound can also be improved by null subtraction imaging (NSI), a technique that applies multiple receive apodizations to reduce sidelobes and enhance the mainlobe [7]. Super-resolution of these techniques is achieved contrast-free and based on manipulation of the transmit and/or received point spread function (PSF) of the ultrasound imaging system.

Recently, a contrast microbubble (MB)-based technique named ultrasound localization microscopy (ULM) [8], [9] was proposed to overcome the diffraction limit of ultrasound based on the principle of localization microscopy (e.g., PALM [10], [11] and STORM [12]). MBs have a similar size to red blood cells, and they travel within the vasculature for several minutes' post-intravenous injection [13]. A super-resolved blood vessel density map can be constructed by localizing and accumulating the MB locations in a consecutive sequence of diffraction-limited image frames. Moreover, a high-fidelity blood flow speed map can be recovered from measuring the frame-to-frame displacement of MBs. Various studies have demonstrated multiple applications of ULM under clinical and preclinical settings, including tumor characterization [14]–[16], brain imaging [17], and abdominal organs such as liver, kidney, and pancreas [18].

At present, the clinical potential and utility of ULM are challenged by various limitations of ULM, including the time-consuming imaging acquisition, the high computational cost of postprocessing steps, the lack of ground truth to validate ULM *in vivo*, and other practical challenges such as tissue motion caused by breathing and scanning. These limitations and challenges remain active topics of research. In this study, we focus on addressing the issues of MB localization that contribute to the time-consuming data acquisition and high computational cost associated with postprocessing. The current standard practice for MB localization is based on using an estimated template of the ultrasound system impulse response (i.e., the PSF) to identify individual MB signals and measure the physical location of the MB (e.g., by estimating the centroid of the MB signal). However, an ultrasound imaging

system has a spatially variant PSF that is subject to many confounding factors, such as noise, phase aberration, multipath reverberation, and attenuation. Estimation of PSF is a challenging task, which leads to unreliable MB localization. In addition, as shown in [19] and [20], MB localization based on the centroid of the MB signal is subject to errors induced by the nonlinear MB response and the spatial sampling of the MB signal. Meanwhile, even with an ideal localization algorithm and MB data, the accuracy of MB localization is still subject to the Cramer–Rao lower bound, which is approximately one-tenth of the ultrasound wavelength [21]. Since MB localization occurs early in the ULM processing workflow, errors introduced in the localization step propagate downstream to affect the performance of subsequent operations, resulting in suboptimal ULM image quality. Therefore, improving the performance of MB localization is an essential task for improving the overall robustness of ULM.

Another challenge associated with MB localization in practice is the tradeoff among MB concentration, data acquisition time, and localization accuracy. A diluted concentration of MB creates more MB signal separation in space, which makes the MB locations less ambiguous and easier to estimate. However, dilution of MB also slows down MB perfusion, resulting in longer acquisition times for full vascular reconstruction [22]. On the other hand, although high MB concentration leads to faster vessel MB saturation, the amount of spatially overlapping MB signals also substantially increases. This leads to a significant amount of wasted MB signals because overlapping MB signals are difficult to localize with conventional localization techniques and subsequently discarded. Although MB signals with partial overlaps can be potentially separated and localized, the overlap may distort the individual MB signal and challenge the predefined PSF. In reality, especially in clinical settings for human imaging, it is difficult to modulate the MB concentration in the bloodstream, which is largely dictated by physiology. In addition, the clinical dose of MB administration and MB concentration is typically regulated, which is difficult to change. Therefore, developing a robust MB localization method that works for various MB concentrations with various amounts of MB overlaps (especially for high MB overlap) is critical for the successful preclinical and clinical translations of ULM.

Several methods have been proposed to address the issue of high MB overlap at high MB concentrations. Zhang *et al.* [23], [24] proposed the use of phase change nanodroplets for flow and concentration-independent localization microscopy with reduced acquisition time. Recently, we proposed an MB separation technique based on Fourier-based filtering to address the issue of MB overlap by decomposing high MB-count data into subsets of low MB-count data [25]. The MB separation was achieved by leveraging the different movement speeds and directions of MBs, which manifests as different slow-time frequency components in the Fourier space. The method greatly improved the amount of MB signals that can be localized and subsequently the overall quality of ULM imaging. However, the assumption of utilizing temporal frequency difference to separate MB signals may break down when MB concentration is high and/or flow hemodynamics

are complex. Also, by separating the data into subsets with identical dimensions as the original data, the MB separation technique exacerbates the issue of the high computational cost of ULM. Therefore, it may be challenging in practice to implement the MB separation technique, and a faster and better MB localization method remains to be developed.

Considering ongoing breakthroughs in deep learning (DL) and computer vision, DL-based techniques have gained popularity in medical image processing, especially for tasks requiring a high level of abstraction [26]. In ultrasound imaging, DL implementation has mostly been focused on improving beamforming [27]–[31]. For DL-based ULM, van Sloun *et al.* [32] and Liu *et al.* [33] validated the feasibility of DL-based MB localization using different classes of neural network (NN) architectures. Both studies have shown promising potential for applying DL for accelerated and robust MB localization. Liu *et al.* [33] also used *in vivo* vessel networks to generate more realistic training data. Youn *et al.* [34] performed convolutional neural network (CNN)-based localization on radio frequency (RF) channel data generated by the Field-II simulation, which showed improvement over local peak detection using envelope-detected signals on both simulation and phantom data. Lok *et al.* [35] used data from chicken embryo chorioallantoic membrane (CAM) for *in vivo* validation with optical images as the reference. They showed that DL-based MB localization outperformed conventional localization techniques under low-to-moderate MB concentrations. As an alternative to the common localize-and-track workflow for ULM, Milecki *et al.* [36] proposed a DL-based framework that recovers tracks directly from the MB signal without explicitly localizing individual MBs. Since velocity information is not retained in the recovered tracks, this alternative framework is suited for applications where blood flow dynamics information is not required. In our study, we will focus on the typical ULM framework with explicit localization of MBs.

Different from existing studies where low-to-moderate MB concentrations were typically used to develop DL-based MB localization, our objective in this study was to develop and test the performance of DL-based MB localization under high-to-very-high MB concentrations. Using high MB concentration increases the probability of detecting MBs in the bloodstream, especially for small vessels. The time it takes to detect a certain amount of MBs for reconstruction can be reduced, which shortens ULM data acquisition time that is significant for clinical implementations of ULM. To the best of our knowledge, whether DL-based localization can withstand the high MB concentration scenario is yet to be explored. In this study, we characterized the performance of various MB localization techniques under a wide range of MB concentrations by using Field-II ultrasound simulations as well as vascular graph models extracted from *in vivo* data for training the NN. Our training data consist of realistic vascular structures extracted from optical images of chicken embryo CAM surface vessels. We also explored the use of RF data to investigate localization with both amplitude and phase information. Our method was applied to several different scenarios: simulated test data ranging from very sparse to very dense MB

distributions, for the characterization of how the performance gain of DL-based MB localization scales with increasing MB concentration; simulated test data of high concentration MB within the vascular structure, for the study of the potential reduction in acquisition time using DL-based localization; and on *in vivo* CAM data with high concentration MB injection and high-resolution optical images as the ground truth. In addition to using DL-based MB localization for conventional ULM processing, we also proposed a high-resolution blood flow visualization method using DL-processed B-mode ultrasound MB data to demonstrate the feasibility of potential real-time ULM imaging.

II. METHODS

A. Field-II Simulation for MB Data Synthesis

To study the effect of MB distribution in the training data, three groups of training datasets were generated based on Field-II [37], [38] and used to train the networks: group 1 uses a low MB concentration of up to 300 MBs/mm² ($1.8 \text{ MBs}/\lambda^2$), where λ is the wavelength of an ultrasound imaging system with 20-MHz center frequency, and group 2 and 3 both have higher MB concentration closer to *in vivo* experiments, where group 2 uses randomly distributed MB signal (in space) for training and group 3 uses the CAM vessel model for assigning MB positions. The hypothesis was that the added information of microvascular structure (e.g., vessel versus nonvessel regions and large versus small vessels) would facilitate DL performance in MB localization.

For each training group, we generated 9000 training images and 1000 validation images of MBs within a 1 mm \times 1 mm field of view (FOV), which corresponds to 203 \times 203 pixel images with a 4.9- μm pixel size. This FOV was selected to balance several factors, including being able to capture structural features of the vessel network on multiple scales, not exceed the computer memory constraints, and maintain the efficiency of training the NN. With confocal imaging [39], we were able to observe ~ 25 MBs in a 200 μm \times 200 μm FOV, which translates to ~ 625 MBs/mm². Therefore, we used a normal distribution with $\mu = 600$ and $\sigma = 600$ MBs/mm² to regulate MB concentrations in our high concentration simulations (groups 2 and 3).

For the group 1 and group 2 training set, the axial and lateral coordinates of the MBs were drawn from a uniform distribution within the dimension of the imaging FOV. For the group 3 training set, optical images (see Section II-D for optical imaging details) of the CAM surface vessels were used to place MBs in the vascular architecture (Fig. 1). Binary vessel maps were generated from the green channel (which provides the best contrast for vessels) of each optical image using MATLAB's `adaptthresh` function, which computes local threshold based on the Gaussian weighted mean of neighboring pixels [Fig. 1(b)]. A subregion (1 mm \times 1 mm) within the binary maps was randomly selected for the simulation of each training image. Training images with different MB concentrations were generated in a similar way as in groups 1 and 2.

The simulation sequence was configured according to the CAM ultrasound imaging settings described in Section II-D.

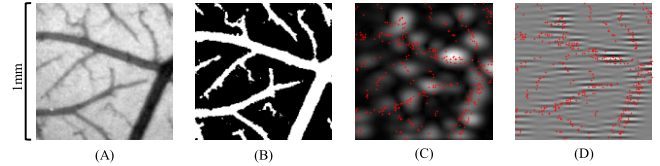


Fig. 1. Example of group 3 training data generation based on optical imaging of the CAM. (a) Green channel image of the CAM surface. (b) Binary vessel map after adaptive thresholding. (c) and (d) Simulated B-mode (i.e., envelope detected) image and RF data using Field II, respectively. Red crosses mark the true MB locations.

TABLE I
FIELD-II SIMULATION PARAMETERS

Variable name	Value
Transducer type	Linear array transducer
Number of transmit elements	128
Number of receive elements	128
Element height	80 μm
Element width	62 μm
Pitch	70 μm
Transmit apodization	Kaiser
Receive apodization	Rectangle
Center frequency	20 MHz
Sampling frequency	125 MHz
Speed of sound	1540 m/s
Number of compounding angles	9

Details of the simulation specifications are summarized in Table I. Simulated ultrasound data were interpolated to the resolution of 4.9 μm . Finally, Gaussian white noise with a level of $10\% \pm 2\%$ of the ultrasound signal peak amplitude was added to the ultrasound data. The noise level was selected based on measurements from our experimental CAM data.

B. DL Model Design

U-Net is a CNN architecture first proposed in 2015 for biomedical image segmentation [40]. The architecture and its variations have since been successfully applied to many different tasks in biomedical image processing. Here, we adopt the encoder-decoder structure of the U-Net for our CNN model. The model contains a feature extraction path (the encoder) that extracts high-dimensional feature maps from the input ultrasound data and a reconstruction path (the decoder) that extracts MB locations from the embedded feature map.

Fig. 2 shows a schematic of the network. Each down-sampling block contains two convolution-batch normalization-activation units. Each reconstruction block starts with a 2×2 -kernel convolution that halves the number of channels. The output of the corresponding feature extraction block is upsampled to match the spatial dimension of the output of the 2×2 convolution layer before they are stacked along the channel axis. The stacked data then go through convolution-batch normalization-activation units and are upsampled by a

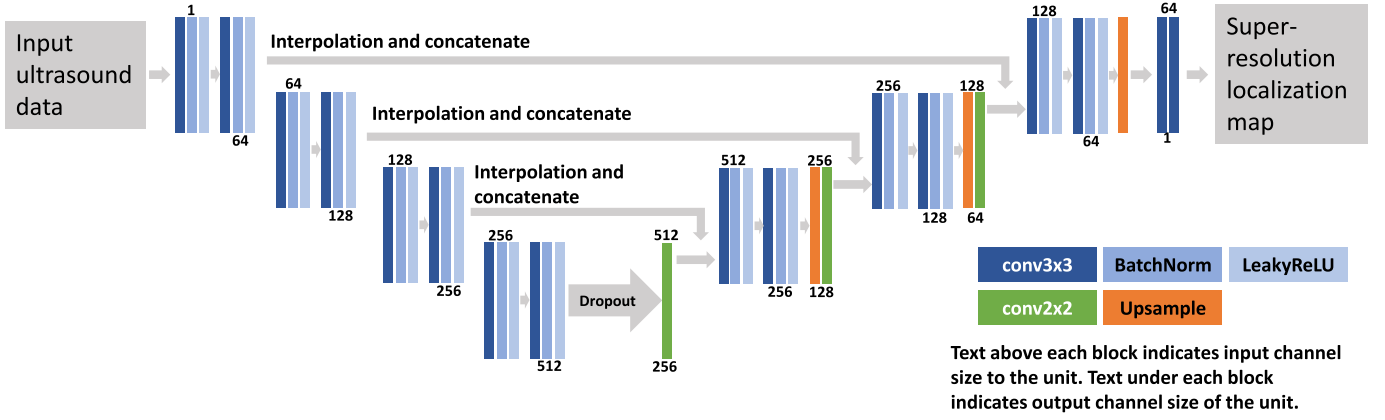


Fig. 2. Schematic of the NN. The color-coded blocks represent different types of layers in the NN. The NN contains a feature extraction path that extracts feature maps from the input ultrasound data and a reconstruction path that recovers a super-resolution localization map from the feature maps.

factor of 2. Dropout is implemented in the bottleneck layer to prevent overfitting [41]. The final output block contains two 2-D convolution layers. The network contains a total of eight convolution layers in the encoder part and eight convolution layers in the decoder part. We used leaky ReLU as the activation function, which is defined as

$$\text{LeakyReLU}(x) = \begin{cases} x, & x > 0 \\ 0.01 \times x, & x \leq 0 \end{cases} \quad (1)$$

where x is the output of a convolution layer after batch normalization. The model was trained with the Adam optimizer with a learning rate of 0.001 [42]. The loss function was defined as

$$L(m, y, \hat{y}) = \alpha \text{MAE}(G(y), \hat{y}) + \text{SSIM}(G(y), \hat{y}) + \beta \text{FPLoss}(m, \text{LM}(\hat{y})) \quad (2)$$

where y is the ground-truth binary localization map, m is the ground-truth binary vessel mask, \hat{y} is the predicted localization map, G is a Gaussian filter with $\sigma = 2$ pixels (the pixel size is $4.9 \mu\text{m}$), and MAE is the mean absolute error, which is defined as

$$\text{MAE}(y, \hat{y}) = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (3)$$

where y_i and \hat{y}_i denote the i th entry in the ground truth and the prediction, respectively. SSIM represents the structural similarity [43]

$$\text{SSIM}(y, \hat{y}) = \frac{(2\mu_y \mu_{\hat{y}} + c_1)(2\sigma_{y\hat{y}} + c_2)}{(\mu_y^2 + \mu_{\hat{y}}^2 + c_2)(\sigma_y^2 + \sigma_{\hat{y}}^2 + c_2)} \quad (4)$$

where μ_y and σ_y are the average and the standard deviation of the pixel values in the ground-truth image, respectively, $\mu_{\hat{y}}$ and $\sigma_{\hat{y}}$ are the average and the standard deviation of the pixel values in the prediction, respectively, $\sigma_{y,\hat{y}}$ is the covariance of the ground truth and the prediction, and c_1 and c_2 are the stabilizer terms in case of division by a very small denominator, which are defined as

$$c_1 = (0.01L)^2, \quad c_2 = (0.03L)^2 \quad (5)$$

where L is the dynamic range of the image. FPLoss is a term that penalizes false positive predictions (i.e., false MB localizations), defined as the number of MB localizations that are outside of vessels, divided by the total number of predicted positives

$$\text{FPLoss}(y, \hat{y}) = \frac{\sum (1 - m_i) \hat{y}_i}{\sum \hat{y}_i} \quad (6)$$

where m_i is the i th entry in the binary ground-truth vessel map. Since FPLoss is a metric for binary prediction and the output of the network is not necessarily binarized, a local maximum filter local maximum (LM) was used to convert the output to a binary localization map where the pixels of the local peaks were set to 1. α and β are weights for the MAE and FPLoss terms, respectively. α was set to 0.001 for all training sets. β was set to 1 for training sets that contain vessel structure and 0 for training sets without vessel structure. The loss function was designed so that the main objective function to reduce was still the SSIM. The addition of the MAE term helps the model to better preserve the intensity of the produced output, while FPLoss provides the network with additional information regarding the vascular structure that is not explicitly given by the map of the true MB locations. False MB localization within nonvascular regions will be penalized heavier than false MB localization within vascular regions.

After each epoch of model update using the 9000 training images, the training performance was evaluated on the validation set of 1000 images. Hyperparameter selection relies mostly on the validation set performance. Training samples were fed into the network as minibatches of 16 images. The final trained model generates “sharpened” MB signals (Fig. 3), reducing the dimension of the apparent PSF size, where overlapping MB signals in the original input are now separated.

C. ULM Data Processing

For all ULM data except simulation, the tissue signal was first removed by singular value decomposition (SVD)-based clutter filtering [44]. For MB localization with conventional,

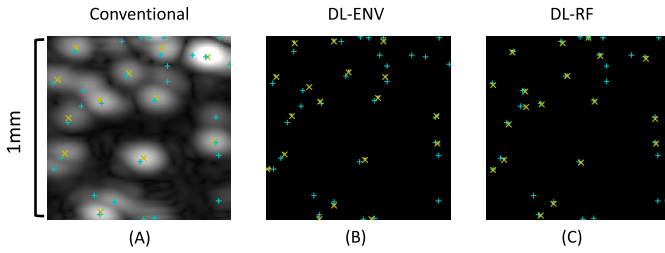


Fig. 3. (a) Simulated B-mode (envelope detected) image of the MBs and the results of conventional localization. (b) Output of the model trained with envelope-detected (ENV) data. (c) Output of the model trained with RF data. In all subfigures, measured locations are marked by yellow \times and true MB locations are marked by cyan $+$. DL enhances the input ultrasound data by sharpening the MB signal so that they can be better separated for easier localization.

non-DL MB images, a cross correlation-based approach as described in [45] was used with a synthesized multivariate Gaussian PSF. For the DL-processed MB data, the location of each MB was extracted by the blob detection function available in the scikit-image Python package using the difference-of-Gaussian method [46]. The extracted locations were accumulated across each imaging frame to generate the microvessel density map. To better evaluate the performance of MB localization across different methods, no MB pairing or tracking (i.e., raw accumulation only) was implemented in our processing.

In addition to conventional ULM processing, the DL-enhanced MB data can also be used for real-time, high-resolution blood flow display: a sliding window of size W in temporal direction was used to directly accumulate the DL-enhanced MB data to display blood vessels with high spatial resolution. A matrix keeps the latest W DL-processed frames of MB data. When a new frame is processed by the network, the matrix will be updated by removing the oldest frame with the new frame. A new frame of output display was obtained by summing all the W frames of the updated matrix in the time direction.

D. In Vivo ULM Data Acquisition

Fresh fertilized chicken eggs were provided by the University of Illinois Poultry Research Farm and housed in a tilting incubator (Digital Sportsman Cabinet Incubator 1502, GQF Manufacturing Inc., Savannah, Georgia). After four days, the eggshells were removed and embryos were transferred into plastic weigh boats. The *ex ovo* CAMs were then placed into a humidified incubator (Darwin Chambers HH09-DA) until the day of imaging, 13 days after eggshell removal.

In preparation for contrast MB injection, a glass capillary needle was produced by pulling a borosilicate glass tube (B120-69-10, Sutter Instruments, Novato, CA, USA) with a PC-100 glass puller (Narishige, Setagaya, Japan). The glass needle was attached to a 1-mL syringe using Tygon R-3603 laboratory tubing. The surface vasculature of the CAM was punctured with the glass capillary needle and 50 μ L of the Definity solution (Lantheus Medical Imaging Inc., Billerica, MA, USA) was injected into the embryo immediately prior to imaging.

Optical images were acquired using a Nikon SMZ800 stereomicroscope (Nikon, Tokyo, Japan) with a mounted DS-Fi3 digital microscope camera (5.9-Mpixel CMOS image sensor, Nikon). Optical data were recorded using the Nikon NIS-Elements software platform and exported for offline analysis.

Ultrasound data were acquired using a Vantage 256 system with an L35-16vX high-frequency linear array transducer (Verasonics Inc., Kirkland, WA, USA). The transducer was placed on the side of the plastic weigh boat to image the surface vasculature of the CAM. Imaging was performed with a center frequency of 20 MHz using nine-angle plane-wave compounding (1° increment) with a post-compounding frame rate of 1000 Hz. Ultrasound data were saved as in-phase and quadrature (IQ) datasets of 1600 frames each for a total acquisition length of 32 s (32 000 frames). RF data were obtained from the IQ data based on IQ demodulation [47].

III. RESULTS

A. Individual MB Localization Performance for Low-to-Moderate MB Concentration

The performance of the DL-based localization was quantified on a set of Field-II simulation data with low-to-moderate MB concentration using models trained with group 1 low concentration data. The evaluation focuses on the ability to correctly identify individual MB signal. The testing set consists of 1 mm \times 1 mm samples with MB concentration ranging from 1 to 100 MBs/mm² (0.0059–0.59 MBs/ λ^2), and 50 samples were generated for each concentration. DL-based localization was performed on each of the simulated data. The DL localized MB locations and true MB locations were paired by using the algorithm proposed in [45]. True MB locations that were not paired, as well as paired true MB locations where the pairwise distance was greater than a threshold tol , were considered false negatives (FN_{MB}). Similarly, predicted MB locations that were not paired, as well as paired predicted MB locations where the pairwise distance is greater than tol , were considered false positives (FP_{MB}). The remaining paired MB locations were considered true positives (TP_{MB}) and subject to localization error measurement.

An evaluation metric that includes three criteria was developed to quantitatively measure MB localization performance: MB false discovery rate (FDR), MB miss rate, and mean localization error that are defined as

$$FDR = \frac{FP_{MB}}{FP_{MB} + TP_{MB}} \quad (7)$$

$$MissRate = \frac{FN_{MB}}{FN_{MB} + TP_{MB}} \quad (8)$$

$$Mean\ Localization\ Error = \frac{\sum_{i=1}^N |\hat{y}_i - y_i|_2}{N} \quad (9)$$

where \hat{y}_i is the measured true positive MB locations and y_i is the corresponding ground truth. The performances of ENV-trained network (DL-ENV), RF-trained network (DL-RF), and conventional localization were averaged across all the 50 testing images for each MB concentration. Fig. 4 plots the average performance of DL-based localization and

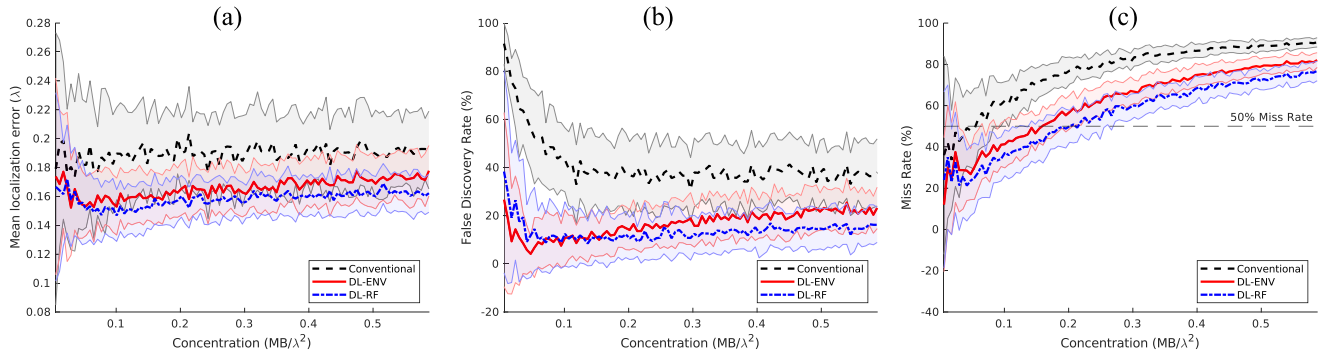


Fig. 4. Average localization performance on the low concentration simulation test dataset. (a) Mean localization error. (b) False discovery rate. (c) Miss rate.

conventional localization against increasing MB concentration using the three metrics described above with tol set to 5 pixels (0.32λ). The shaded regions represent the standard deviation of the performance within each concentration. Both DL-based localization methods outperformed conventional localization under this setting. DL-RF outperforms DL-ENV in all criteria except for the low concentration FDR, which can be a result of DL-RF having the overall tendency of localizing more MBs than DL-ENV. Conventional localization has particularly poor performance compared to DL-ENV and DL-RF for low concentration in FDR because it lacks the ability to distinguish noise and artifact from the actual MB signal. The gap between DL-ENV and DL-RF increases in the mean localization error and the FDR as concentration increases, which indicates that the addition of phase information in the RF data helps to better maintain the robustness of localization as the amount of overlapping MB signal increases.

We are specifically interested in the case of missing localizations, as it is closely related to whether increasing MB concentration is still profitable in terms of reducing the acquisition time required for full reconstruction of the vascular structure. We identify the point where a localization method hits above 50% miss rate as the indicator that it stops effectively localizing individual MBs. The 50% miss rate was marked on the performance graph as horizontal dashed lines. Fig. 5 plots the concentrations at which each localization method hits above 50% miss rate against different tol values chosen to define a missed localization (FN_{MB}). Both figures showed that DL-based localization can withstand higher concentration before accurate localization of individual MB becomes a task too challenging, which can be useful for data acquired with high MB concentration as well as for complex or hierarchical vasculature, where large vessels are present in the imaging plane. Specifically, DL-RF localization is able to maintain less than 50% miss rate using four times higher MB concentration compared to conventional localization.

However, reaching the “breakpoint” for individual MB localization does not necessarily mean the complete breakdown of ULM algorithm. Localization of the centroids of “blobs” of MBs can still provide useful information for reconstruction of the vascular structure. Moreover, as the blobs often tend to stay clustered for several consecutive frames, motion

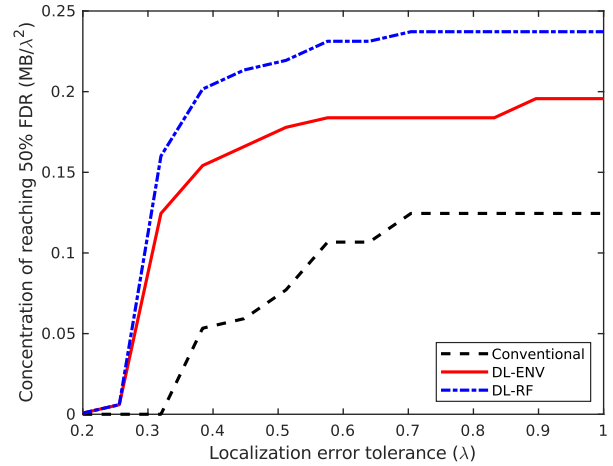


Fig. 5. Point where each localization method hits above 50% miss rate versus different tolerance values for localization error for the simulation dataset.

tracking on the entire blobs can still be used for estimation of flow speed. Therefore, we will further extend the concentration range to study the performance of each localization method under more extreme scenarios that are likely to be closer to real experimental environment.

B. Individual MB Localization Performance for High Concentration MB Data

The concentration of simulation data was extended to 1000 MBs/mm^2 ($5.9 \text{ MBs}/\lambda^2$) with a 10-MBs/mm^2 step size, with 50 samples generated for each concentration. The DL model was retrained using groups 2 and 3 high concentration training data. Evaluation of each method was carried out using the same method described in Section II-A. The behavior of the trained model changes with significantly increased MB concentration. Fig. 6 compares the performance of models trained with simulation with and without the CAM vessel structure against conventional localization. Fig. 7 compares the MB localization performance between conventional and DL-based methods based on CAM vessel training data only. A similar performance as the low-to-moderate concentration was observed for the extended concentration ranges. DL-based localization methods, in general, showed better

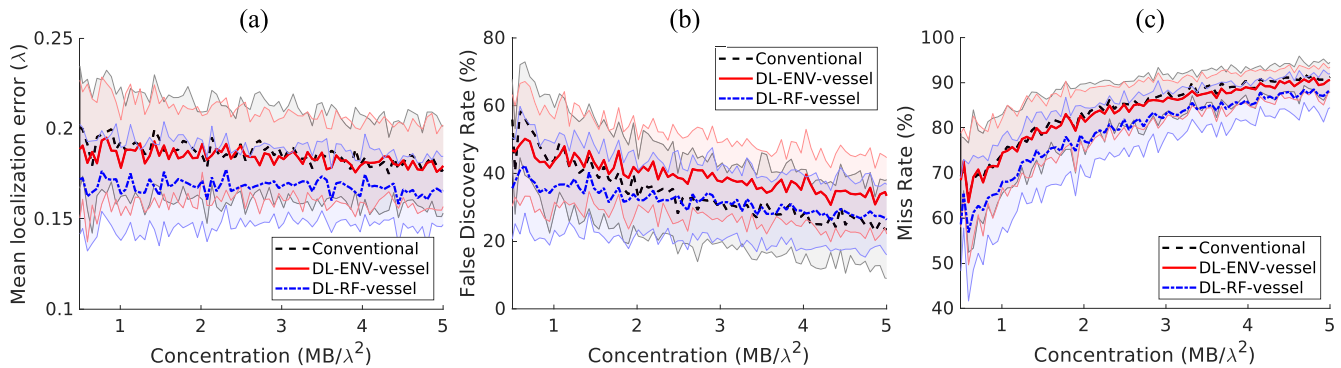


Fig. 6. Localization performance on the simulation testing set, comparison between conventional localization, localization using DL-ENV, and localization using DL-RF. (a) Mean localization error. (b) False discovery rate. (c) Miss rate.

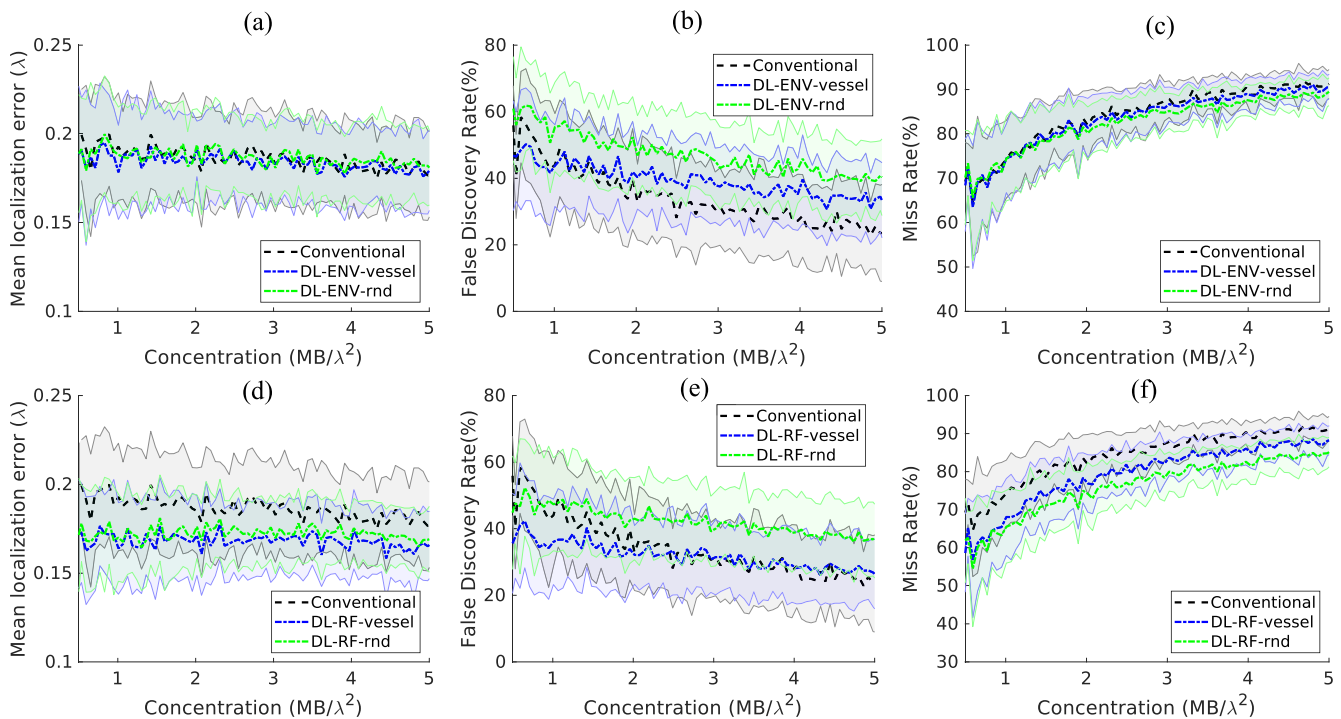


Fig. 7. Localization performance on the simulation testing set, comparison between simulation training set with random MB distribution and MB within vessels. (a)–(c) Compares the mean localization error, false discovery rate, and the miss rate between conventional localization and DL-ENV methods. (d)–(f) Compares the mean localization error, false discovery rate, and the miss rate between conventional localization and DL-RF methods. Note that the results presented in Fig. 6 were also included in Fig. 7. The plots in Fig. 6 were rearranged to show the comparison between conventional localization, localization using DL-ENV, and localization using DL-RF.

performance than conventional localization across the board, with DL-RF showing the best localization performance overall. DL-based localization with the CAM vessel structure incorporated in simulation had better MB localization performance than without. FDR was significantly reduced, while minor improvements were observed for mean localization error and miss rate. The significantly lower false positives for both ENV and RF trained networks indicate that the spatial information associated with vessels was indeed helpful for the network to differentiate vessel versus nonvessel regions to facilitate better MB localization. DL-RF localization was able to reduce the MB miss rate while keeping the FDR at a similar level with conventional localization. This is not surprising because, as hypothesized, RF data have both the amplitude and the phase information of MBs, which facilitates MB separation

and localization. The performance gain of DL-ENV localization was worsened compared to the low concentration case, which indicates that without phase information, the localization method is more sensitive to increased overlapping MB signal.

Of note, beyond the concentration of $3 \text{ MBs}/\lambda^2$ or $500 \text{ MBs}/\text{mm}^2$ (example shown in Fig. 8), the localization performance of all methods plateaued. The gain in MB localization performance using DL becomes very small compared to conventional localization. This is not surprising because, at such high concentration, the task of recovering MB locations from ultrasound signal turned into an ill-posed problem. As it is not feasible to correctly localize individual MBs using a single frame of MB data with spatial information alone, regardless of the localization techniques being used,

the previously introduced performance metric becomes less meaningful.

C. Resolving Synthetic Simple Vascular Structure

When evaluated as a part of the ULM processing chain, the individual MB localization quality translates to the ability to reconstruct vessel structures with high fidelity. Therefore, we continued to investigate the quality of vessel maps reconstructed using each localization method. The evaluation was first performed on a set of synthetic data consisting of two vertical or horizontal pairs of vessels with a diameter of 5 pixels. Each pixel is $4.9 \mu\text{m}$ (approximately 0.064λ). The separation between vessels was 2 pixels (0.13λ) in the horizontal case and 6 pixels (0.38λ) in the vertical case. MBs were randomly placed within vessels to generate 200 frames of ultrasound data. ENV-trained network (DL-ENV), RF-trained network (DL-RF), and conventional localization were applied to each frame. Accumulating location of localized MBs over 200 frames results in the vessel maps shown in Fig. 9 (the horizontal case) and Fig. 10 (the vertical case). The profile of the reconstructed vessels averaged along the direction of the vessels was also shown in Figs. 9 and 10. For the horizontal vessels with a separation of 2 pixels (0.13λ), DL-RF was able to produce a clean separation of the vessels. DL-ENV was also able to separate the vessels, but the separation was not as clear and only observed from the profile. In the vessel map generated using conventional localization, the two vessels were merged, shown by the absence of two separate peaks in Fig. 9(c). When the separation was increased to 10 pixels (0.64λ), all three methods were able to resolve the two vessels. However, DL-RF was able to produce the clearest boundary and cleanest nonvascular space in between vessels [Fig. 9(b) and (d)]. Both DL-based localization methods were able to separate the two horizontal vessels with separation of 6 pixels (0.38λ), where conventional localization failed to resolve two vessels. When the separation was increased to 10 pixels, all methods were able to resolve the vessels, with both DL-based methods producing cleaner boundaries than the conventional method and DL-RF slightly outperforming DL-ENV, which is consistent with the horizontal case. To summarize, both DL-based localization improved the ability to resolve closely positioned structures of ULM. Moreover, DL-RF was able to further improve the axial resolution due to its access to the information contained in the axial oscillation of the MB signal.

D. MB Localization Performance Using CAM Vessel Structures as Testing Data

In this section, we tested the MB localization performance on simulation based on realistic CAM vessel structures. Only DL networks trained with group 3 high concentration data (with CAM vessel structures) were used in this section. For test data generation, four $1.5 \text{ mm} \times 1.5 \text{ mm}$ regions of interest (ROIs) were sampled from the CAM optical images and converted to binary vessel maps using adaptive thresholding. For each vessel, frames of ultrasound MB images were simulated using the same method described in

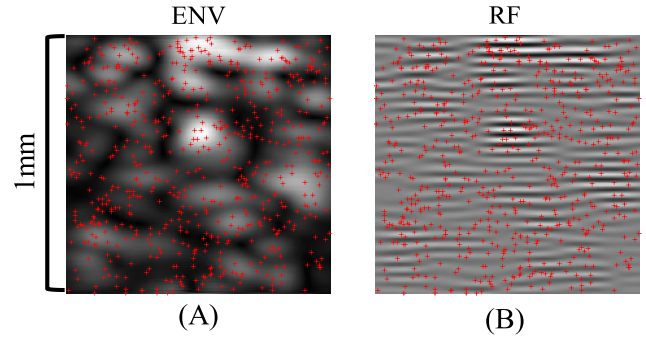


Fig. 8. Ultrasound MB simulation where 625 MBs were present in the $1 \text{ mm} \times 1 \text{ mm}$ FOV. “x” marks the true MB locations. (a) ENV data. (b) RF data.

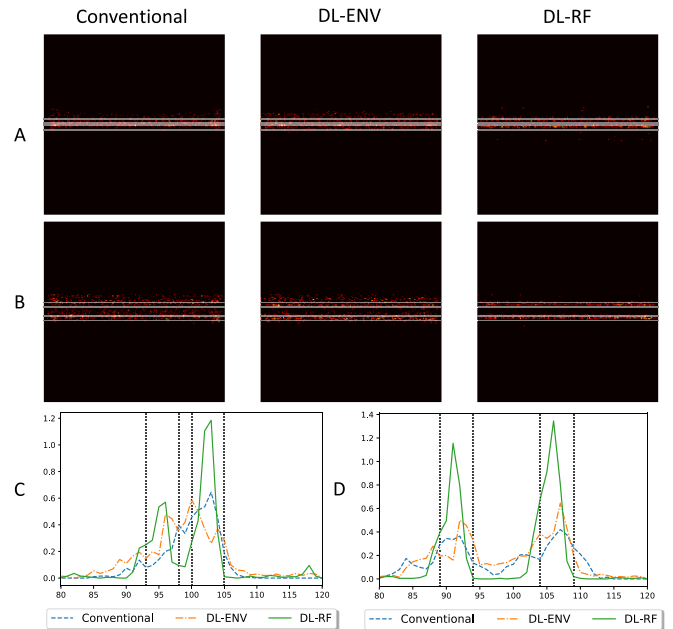


Fig. 9. ULM reconstruction of two closely spaced simulation vessels. (a) Vessels are axially separated by 2 pixels ($0.98 \mu\text{m}$, 0.13λ). (b) Vessels are axially separated by 10 pixels ($4.9 \mu\text{m}$, 0.64λ). (c) and (d) Profile of the reconstructed vessels in (a) and (b), respectively.

Section II-A. The average MB concentration of the testing data was around $400 \text{ MB}/\text{mm}^2$. Conventional and DL-based localizations were performed for each frame and accumulated across 1600 frames. Again, no MB tracking was conducted to better evaluate the localization performance for each method. Furthermore, the performance of localization methods was no longer evaluated based on the ability to precisely localize individual MBs. Two vessel-specific metrics were used: 1) the vessel FDR and 2) the vessel miss rate. Definitions for the vessel FDR and miss rate follow (7) and (8), where false negatives ($\text{FN}_{\text{vessel}}$) were defined as locations within vessels that were not covered by localization, false positives ($\text{FP}_{\text{vessel}}$) were defined as locations outside vessels that were detected as localization, true positives ($\text{TP}_{\text{vessel}}$) were defined as locations within vessels and covered by localization, and true negative ($\text{TN}_{\text{vessel}}$) were defined as locations outside the vessels that do not have localizations.

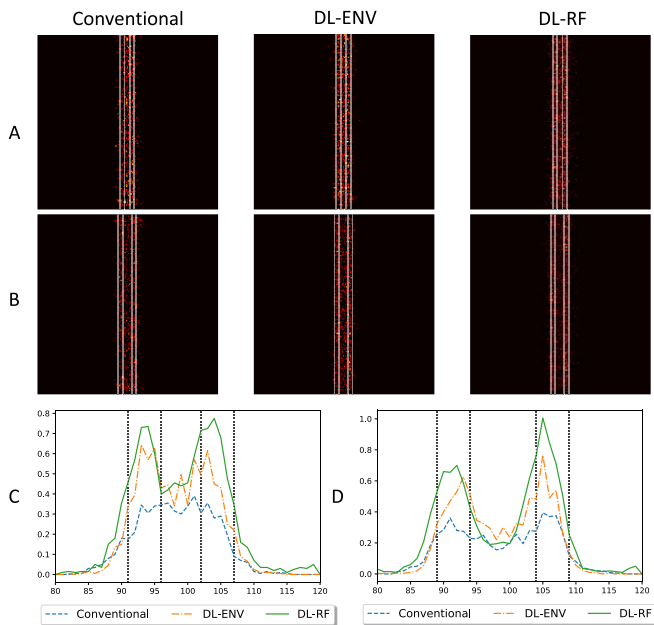


Fig. 10. ULM reconstruction of two closely spaced simulation vessels. (a) Vessels are laterally separated by 6 pixels ($3.0 \mu\text{m}$, 0.38λ). (b) Vessels are laterally separated by 10 pixels ($4.9 \mu\text{m}$, 0.64λ). (c) and (d) Profile of the reconstructed vessels in (a) and (b), respectively.

TABLE II
MB LOCALIZATION PERFORMANCE BASED ON
CAM VESSEL STRUCTURES

		Conventional	DL-ENV	DL-RF
ROI 1	Miss Rate (%)	26.52	22.67	18.11
	FDR (%)	36.93	44.42	36.95
ROI 2	Miss Rate (%)	25.83	23.33	18.54
	FDR (%)	38.77	44.02	35.61
ROI 3	Miss Rate (%)	26.36	21.05	18.31
	FDR (%)	39.90	45.54	39.28
ROI 4	Miss Rate (%)	26.89	22.86	19.65
	FDR (%)	37.11	43.81	36.89

Table II summarizes the localization results using 1600 frames of MB data. As shown in Table II, DL-RF localization was able to achieve both the lowest vessel miss rate and the lowest vessel FDR for all ROIs except for ROI 1 (although only 0.02% higher than the lowest FDR). DL-ENV had better miss rate performance than conventional localization (i.e., localize more MBs) but at the cost of elevated FDR.

Fig. 11 shows the final localization results against the ground-truth vessel map, where four different colors of pixels were used to represent $\text{FN}_{\text{vessel}}$ (white), $\text{FP}_{\text{vessel}}$ (blue), $\text{TP}_{\text{vessel}}$ (red) and $\text{TN}_{\text{vessel}}$ (black). Visually, it is apparent that the DL-RF method fills the vessel with the most $\text{TP}_{\text{vessel}}$ red pixels while retaining a reasonable amount of erroneous blue pixels, indicating relatively fewer $\text{FP}_{\text{vessel}}$ localizations than conventional DL-ENV. It is also worth noting that for both DL-based localization methods, $\text{FP}_{\text{vessel}}$ is most likely to occur just outside the edge of a vessel, while in the conventional

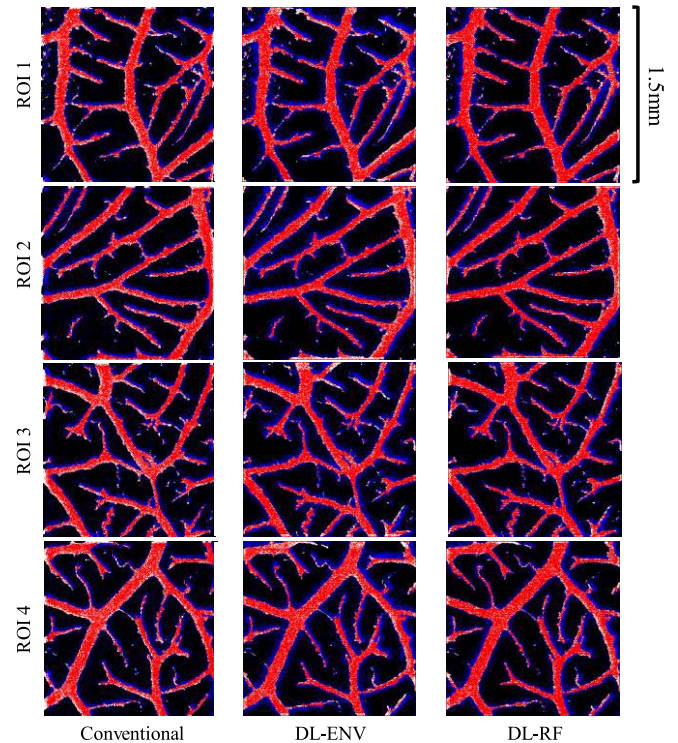


Fig. 11. Difference maps of accumulated localization and ground-truth vessel map for the simulation vessel dataset. Maps of same ROI were arranged in the same row, whereas maps created using the same localization method were arranged in the same column. Red pixels represent $\text{TP}_{\text{vessel}}$ localizations, blue pixels represent $\text{FP}_{\text{vessel}}$ localizations, white pixels represent $\text{FN}_{\text{vessel}}$ localizations, and black pixels represent $\text{TN}_{\text{vessel}}$ localizations.

localization, there are far more $\text{FP}_{\text{vessel}}$ pixels in nonvessel regions. This result indicates that DL-based localization methods are less susceptible to randomly distributed noise commonly mislocalized as MBs than conventional localization techniques.

E. In Vivo Study in CAM Surface Vessel

The proposed DL-based methods were then tested *in vivo* on CAM surface vessel data. The generated microvessel density maps were validated against an optical image of the same area. Selected ROIs (Fig. 12) that provide the best alignment between ultrasound and microscopy imaging were used for the comparison study. For each ROI, ground-truth vessel masks were generated based on the optical image. The performance of each localization method was evaluated in terms of vessel saturation percentage (i.e., localized vessel pixels over total number of vessel pixels). Table III summarizes the 90% vessel saturation time for all the localization methods, including the MB separation method proposed in [25]. DL-RF localization was over $2\times$ faster than the convention method to reach 90% vessel saturation and over 20% faster than MB separation. This result is significant since MB separation is much more computationally expensive than the DL-based methods (excluding training).

Fig. 12 compares all localization methods against the optical ground truth. Both DL-based localization methods filled the

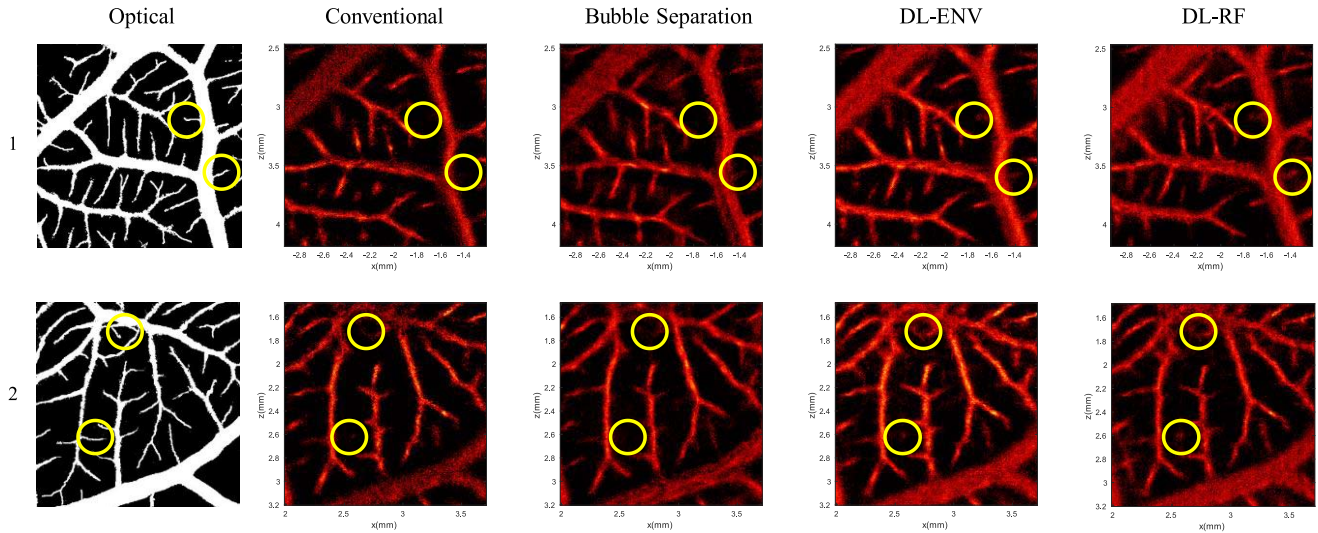


Fig. 12. Comparison of different MB localization methods with optical ground truth for experimental CAM data. The circles denote example thin vessel segments where DL-based localization was better at localizing MB signals than the conventional techniques. The top row and bottom row are of two different ROIs. The first column is the binarized optical ground truth. The rest of the columns are the microvessel density map obtained using different localization methods.

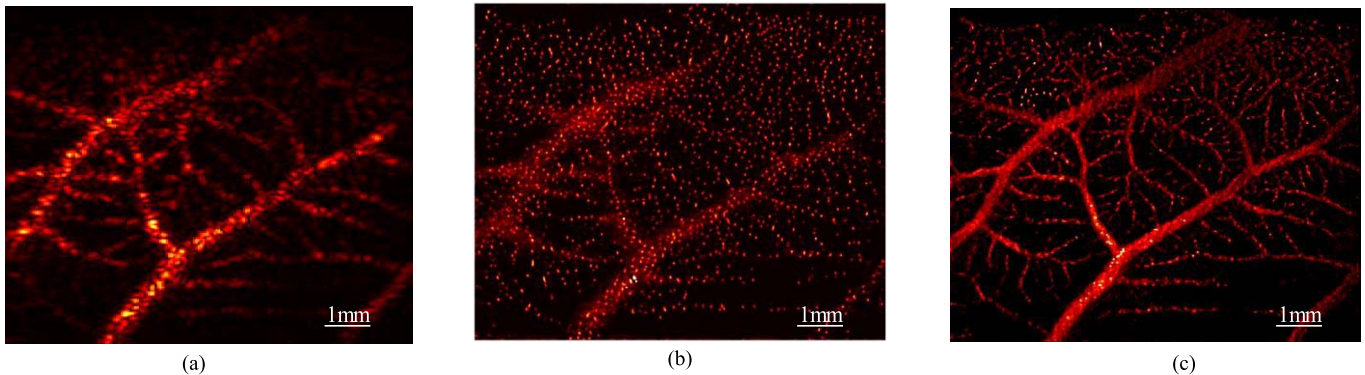






Fig. 13. Example workflow for real-time high-resolution blood flow visualization using experimental CAM data. (a) Original B-mode input image (a single frame) of MBs in the CAM (movie provided in Supplemental Video 1 ). (b) DL-processed MB image using (a) as input. (c) 30-frame accumulation of the DL-processed MB signals that demonstrates high-resolution blood flow visualization (movie provided in Supplemental Video 2 .



TABLE III
ESTIMATED AVERAGE SATURATION TIME

	Conventional	Bubble Separation	DL-ENV	DL-RF
$t_{99\%}$ (s)	4.46	2.40	2.53	1.96

large vessels better (e.g., higher MB count) than conventional localization and bubble separation, indicating a better MB isolation and localization performance at high MB concentrations. The circles in the figure highlight some thin vessel regions according to optical imaging, where DL-based methods had better performance than conventional methods. These results are consistent with the simulation results shown in Fig. 6 and Table II.

F. Real-Time High-Resolution Blood Flow Visualization

Fig. 13 (movie provided in Supplemental Videos 1  and 2 ) shows an example of using the proposed DL-based processing method to generate high-resolution, real-time

visualization of blood flow *in vivo*. Due to the significantly enhanced spatial resolution of MB signal [Fig. 13(b)], one can directly accumulate multiple frames [e.g., 30 frames as shown in Fig. 13(b) and (c)] of the enhanced MB signal to produce high-resolution blood flow maps at a very fast rate with a very low computational cost (i.e., forward processing in the NN). The accumulation alleviates the issue of sparse and weak MB signals within a single imaging frame and enhances the visualization of microvasculature. Longer accumulation time leads to better vessel delineation, at the cost of reduced temporal resolution. Fig. 14 (Supplemental Videos 3  and 4 ) demonstrates two small regions in sequential frames of the DL-enhanced display that captured some typical MB activities seen *in vivo*, including clusters of overlapping MB signal splitting and merging, which are difficult to visualize with conventional ULM.

G. Computational Cost

1) *ULM Processing*: Table IV summarizes the computational performance of different methods involved in this

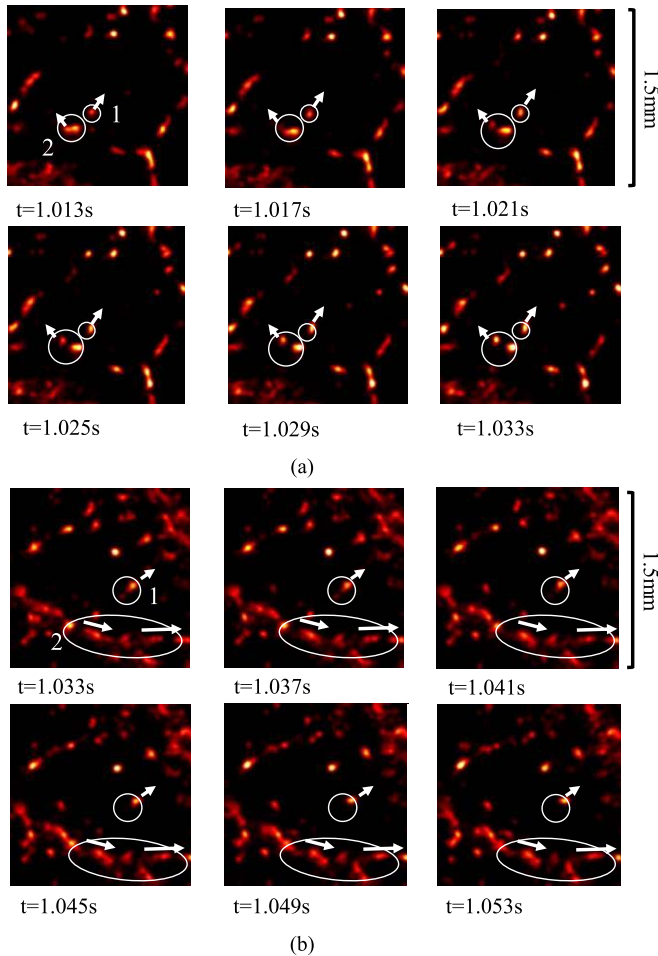


Fig. 14. Two cases demonstrating the utility of the real-time high-definition MB imaging by DL processing on experimental CAM data. (a) Case 1: in circle 1, an MB travels along the direction indicated by the arrow, and in circle 2, an MB splits from a cluster of MBs and travels along the direction indicated by the arrow. (b) Case 2: in circle 1, a weaker blob of MB merges with a stronger blob and travels along the direction indicated by the arrow, and in circle 2, consistent MB flows are observed in a vessel branch along the direction indicated by the arrows.

TABLE IV
COMPUTATIONAL COST FOR PROCESSING
A SINGLE FRAME OF CAM DATA

Method	Conventional (CPU)	Conventional (GPU)	Bubble Separation	DL-ENV	DL-RF
Time(ms)	601	130	304	77	168

study, measured by the time consumption of the localization processing time on a single frame of CAM acquisition with a size of 280×180 pixels ($6.88 \text{ mm} \times 8.83 \text{ mm}$). All algorithms were GPU-based, except for conventional localization that was implemented in both GPU and CPU for reference. The experiments were executed on a workstation running Ubuntu 18.04 operating system, with Intel Core i9-9820X @ 3.30GHz CPU (ten cores), NVIDIA GeForce RTX 2080 Ti GPU, and 64 GB of RAM. The conventional methods were implemented in MATLAB R2019b. The DL-based methods were implemented in Python 3.6 using PyTorch. DL-based methods, in general, have better computational performance

than conventional ones. DL-ENV was the fastest, but the MB localization performance associated with DL-ENV was not as good as with DL-RF. DL-RF was able to achieve around 40% acceleration compared to localization with bubble separation while delivering better MB localization performance. The difference in computation time between DL-ENV and DL-RF was due to the IQ-to-RF conversion process, where the axial dimension was interpolated to four times the pixel size of the original IQ data. The input to the DL-ENV has the same spatial dimension as the original IQ data, which is four times smaller than the input to the DL-RF method.

2) *Real-Time Imaging Feasibility of DL-Enhanced, High-Resolution Blood Flow Imaging:* The total time consumption for processing one frame of the MB data using a trained DL network and generating a new high-resolution vessel image is approximately $2.38 \mu\text{s}/\text{pixel}$. The time consumption of Verasonics beamforming is $\sim 40 \text{ ns}/\text{pixel}$ for each compounding angle. Therefore, for a sequence that uses nine compounding angles and acquires 100×100 pixel ultrasound data per frame, if ignoring the data transfer overhead, then the total processing time of each frame is $(2.38 \mu\text{s} + 9 \times 40 \text{ ns}) \times 10000 \text{ pixels} = 0.0274 \text{ s}$, which corresponds to a display frame rate of $\sim 36 \text{ Hz}$.

For the processing time to satisfy a 10-Hz frame rate for real-time imaging, the processing time of each frame cannot exceed 0.1 s. Therefore, the input MB data may have up to $0.1 \text{ s} / (2.38 \mu\text{s} + 9 \times 40 \text{ ns}) = 36496 \approx 191 \times 191$ pixels. For a center frequency of 5 MHz ($\sim 0.3\text{-mm}$ pixel size), this image size corresponds to a $58.83 \text{ mm} \times 58.83 \text{ mm}$ FOV. For a center frequency of 20 MHz ($\sim 0.077\text{-mm}$ pixel size), the image size corresponds to a $14.71 \text{ mm} \times 14.71 \text{ mm}$ FOV. Note that these calculations did not factor in costs associated with clutter filters (e.g., SVD), although nonlinear imaging (e.g., pulse inversion and amplitude modulation) could be used in lieu of clutter filtering to extract MB signals to be fed into the NN.

IV. DISCUSSION

In this study, we studied the performance of DL-based method for MB localization for ULM using both envelope-detected (ENV) and RF data under challenging high MB concentration scenario. We used Field-II to generate simulated training datasets both with random MB distributions and with realistic vascular structures obtained from *in vivo* CAM surface vessels. We found that having the vascular structure in the training data was beneficial for reducing the number of false positives of MB localization. Moreover, when underlying vascular structure information is available, imposing an additional structural constraint to further penalize localization in nonvascular regions can improve the quality of DL-based localization by producing cleaner nonvascular regions in the reconstruction. However, if the realistic vascular structure is not available to use for simulation, completely randomly distribute MB would also provide the serviceable result as the difference in mean localization error and miss rate was both minor. In general, DL-based localization methods showed better performance than conventional localization, including MB separation. The performance of localization using RF data in both simulation and *in vivo* showed better

performance over ENV since RF data contain both amplitude and phase information of the MB, consistent with the superior performance of RF than envelope reported for DL classification/regression [48]. We also identified three MB concentration ranges where the localization methods showed different types of behavior: for low concentration (under $0.2 \text{ MBs}/\lambda^2$), more than half of the individual MBs can be localized with high confidence. Under moderate concentration (i.e., up to $1 \text{ MBs}/\lambda^2$), localization of individual MBs becomes unfeasible as concentration increases, but the detection of the centroids of multiple clustered MBs can still be used for ULM reconstruction, and the increased concentration provides faster saturation. Finally, the performance of both conventional and DL-based localization plateaus as MB concentration reaches extreme highs (e.g., $>3 \text{ MB}/\lambda^2$). Increasing MB concentration is no longer profitable, even with a more robust localization mechanism. Under high MB concentrations, it is possible that information regarding individual MB locations no longer exists in the RF data due to significant MB signal overlap (i.e., ultrasound wave interference cannot be recovered using only spatial information). In order to overcome this limitation, modifying the DL model to also incorporate temporal information of MBs becomes necessary to further improve the DL-based MB localization performance.

For *in vivo* ULM imaging, adjusting the concentration of contrast agents in the bloodstream can be challenging, especially for imaging under clinical settings. Even if a lower MB concentration can be used to facilitate more robust MB localization, MB concentration can still be very high within large vessels and arteries. Moreover, *in vivo* animal and human imaging are in general more challenging than CAM due to deeper imaging depth, tissue motion, and other sources of noise such as multipath reverberation and phase aberration. Nevertheless, based on the promising results shown in this study, future studies targeting the development of a more robust DL-based MB localization method that can account for these challenges are warranted.

There are some limitations in this study. First, although optical images of the CAM surface vessel were used as ground truth for evaluating the performance of MB localization *in vivo*, it was difficult to achieve a perfect registration between ultrasound and optical imaging FOVs. The quality of the ground truth obtained from optical images was also affected by the quality of image segmentation as well as the resolution limit of the optical imaging system. As a result, the evaluation metric may not deliver an entirely accurate evaluation of the MB localization performance *in vivo*. Second, simulations such as Field II cannot perfectly account for the variability of PSF in the real image system. Also, Field II may not appropriately simulate nonlinear MB response subject to ultrasound. Therefore, the training of the NNs was not optimal and can be further improved. One solution to this issue is to carefully align ultrasound FOV with optical imaging and simultaneously acquire ultrasound and optical MB data (with MB fluorescently labeled), which can then be used for training. However, the amount of training samples that can be obtained using this method will be far less than from using simulation. Therefore, we posit that simulation will still serve

an important role to initialize DL training for MB localization, and experimental ground-truth data can be used for fine-tuning the network to boost its performance.

ULM is computationally expensive, and therefore, any improvement to any segment of the processing chain contributes to the reduction of overall processing time. In this article, we demonstrated the computational advantage of DL-based MB localization over conventional localization techniques. Our analysis of computational time did not include the clutter filtering and motion correction steps because they were performed for both DL-based and conventional localization techniques. The computational cost of these additional processing steps can significantly affect the real-time capability of the DL-based localization technique. The additional processing time can be optimized by utilizing more efficient algorithms, such as DL-based clutter suppression [49], [50]. Moreover, we did not conduct further downstream processing of the MB locations such as MB tracking in this study, which can be more computationally expensive than localization. Nevertheless, DL should also be capable of tracking MB locations using recurrent NN architectures such as the long short-term memory [51]. Finally, in this article, we proposed a real-time feasible high-resolution vessel visualization method that directly uses DL-enhanced MB signals to display vessels. Although the high-resolution display allows one to visually perceive the blood flow dynamics, the image does not contain quantitative flow speed information that can be obtained with conventional ULM. However, direct inference of MB velocity may be possible by training the DL network with spatial-temporal sequences of MB movement, which ultimately allows real-time display of blood flow speed at high spatial resolution.

V. CONCLUSION

In this article, we studied the performance of DL-based MB localization method for ULM for increasing MB concentration. A U-Net style CNN was trained with Field-II simulated ultrasound data (both envelope detected and RF data) based on *in vivo* CAM surface vessel structure. The NN was able to sharpen the ultrasound MB signal so that overlapping and distorted MB signal were able to be better localized. Under low concentration, DL-based methods were able to more precisely localize individual MBs. As concentration increased, the behavior of all localization methods changed to localizing speckles rather than individual MB. The DL-based method was still able to improve the localization accuracy and reduce saturation time by localizing more MB clusters in each frame of ultrasound acquisition. Furthermore, the DL-processed ultrasound data can be utilized to achieve real-time high-resolution visualization of blood flow dynamics. The proposed method can be further improved by developing a training dataset with better ground truth and incorporating temporal information of MBs in training.

ACKNOWLEDGMENT

The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

REFERENCES

- [1] K. Christensen-Jeffries *et al.*, "Super-resolution ultrasound imaging," *Ultrasound Med. Biol.*, vol. 46, no. 4, pp. 865–891, 2020.
- [2] A. Sukhovich, L. Jing, and J. H. Page, "Negative refraction and focusing of ultrasound in two-dimensional phononic crystals," *Phys. Rev. B, Condens. Matter*, vol. 77, no. 1, Jan. 2008, Art. no. 014301.
- [3] S. Zhang, L. Yin, and N. Fang, "Focusing ultrasound with an acoustic metamaterial network," *Phys. Rev. Lett.*, vol. 102, no. 19, May 2009, Art. no. 194301.
- [4] N. Kaina, F. Lemoult, M. Fink, and G. Lerosey, "Negative refractive index and acoustic superlens from multiple scattering in single negative metamaterials," *Nature*, vol. 525, no. 7567, pp. 77–81, Sep. 2015.
- [5] P. Blomgren, G. Papanicolaou, and H. Zhao, "Super-resolution in time-reversal acoustics," *J. Acoust. Soc. Amer.*, vol. 111, no. 1, pp. 230–248, Jan. 2002.
- [6] T. Ilovitsh, A. Ilovitsh, J. Foiret, B. Z. Fite, and K. W. Ferrara, "Acoustical structured illumination for super-resolution ultrasound imaging," *Commun. Biol.*, vol. 1, no. 1, pp. 1–11, Dec. 2018.
- [7] A. Agarwal, J. Reeg, A. S. Podkowa, and M. L. Oelze, "Improving spatial resolution using incoherent subtraction of receive beams having different apodizations," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 66, no. 1, pp. 5–17, Jan. 2019.
- [8] 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 Trans. Med. Imag.*, vol. 34, no. 2, pp. 433–440, Feb. 2015.
- [9] C. Errico *et al.*, "Ultrafast ultrasound localization microscopy for deep super-resolution vascular imaging," *Nature*, vol. 527, no. 7579, pp. 499–502, Nov. 2015.
- [10] E. Betzig *et al.*, "Imaging intracellular fluorescent proteins at nanometer resolution," *Science*, vol. 313, no. 5793, pp. 1642–1645, 2006.
- [11] S. T. Hess, T. P. K. Girirajan, and M. D. Mason, "Ultra-high resolution imaging by fluorescence photoactivation localization microscopy," *Biophys. J.*, vol. 91, no. 11, pp. 4258–4272, Dec. 2006.
- [12] M. J. Rust, M. Bates, and X. Zhuang, "Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM)," *Nature Methods*, vol. 3, no. 10, pp. 793–796, 2006.
- [13] M. Claudon *et al.*, "Guidelines and good clinical practice recommendations for contrast enhanced ultrasound (CEUS) in the liver—update 2012: A WFUMB-EFSUMB initiative in cooperation with representatives of AFSUMB, AIUM, ASUM, FLAUS and ICUS," *Ultrasound Med. Biol.*, vol. 39, no. 2, pp. 187–210, 2012.
- [14] T. Opacic *et al.*, "Motion model ultrasound localization microscopy for preclinical and clinical multiparametric tumor characterization," *Nature Commun.*, vol. 9, no. 1, pp. 1–13, 2018.
- [15] F. Lin, S. E. Shelton, D. Espíndola, 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.
- [16] J. Zhu *et al.*, "3D super-resolution US imaging of rabbit lymph node vasculature *in vivo* by using microbubbles," *Radiology*, vol. 291, no. 3, pp. 642–650, Jun. 2019.
- [17] C. Demené *et al.*, "Transcranial ultrafast ultrasound localization microscopy of brain vasculature in patients," *Nature Biomed. Eng.*, vol. 5, no. 3, pp. 219–228, Mar. 2021.
- [18] C. Huang *et al.*, "Super-resolution ultrasound localization microscopy based on a high frame-rate clinical ultrasound scanner: An in-human feasibility study," *Phys. Med. Biol.*, vol. 66, no. 8, 2021, Art. no. 08NT01.
- [19] K. Christensen-Jeffries *et al.*, "Microbubble axial localization errors in ultrasound super-resolution imaging," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 64, no. 11, pp. 1644–1654, Nov. 2017.
- [20] P. Song, A. Manduca, J. D. Trzasko, R. E. Daigle, and S. Chen, "On the effects of spatial sampling quantization in super-resolution ultrasound microvessel imaging," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 65, no. 12, pp. 2264–2276, Dec. 2018.
- [21] Y. Desailly, J. Pierre, O. Couture, and M. Tanter, "Resolution limits of ultrafast ultrasound localization microscopy," *Phys. Med. Biol.*, vol. 60, no. 22, pp. 8723–8740, Nov. 2015.
- [22] V. Hingot, C. Errico, B. Heiles, L. Rahal, M. Tanter, and O. Couture, "Microvascular flow dictates the compromise between spatial resolution and acquisition time in ultrasound localization microscopy," *Sci. Rep.*, vol. 9, no. 1, pp. 1–10, Dec. 2019.
- [23] G. Zhang *et al.*, "Fast acoustic wave sparsely activated localization microscopy: Ultrasound super-resolution using plane-wave activation of nanodroplets," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 66, no. 6, pp. 1039–1046, Jun. 2019.
- [24] G. Zhang *et al.*, "Acoustic wave sparsely activated localization microscopy (AWSALM): Super-resolution ultrasound imaging using acoustic activation and deactivation of nanodroplets," *Appl. Phys. Lett.*, vol. 113, no. 1, Jul. 2018, Art. no. 014101.
- [25] C. Huang *et al.*, "Short acquisition time super-resolution ultrasound microvessel imaging via microbubble separation," *Sci. Rep.*, vol. 10, no. 1, pp. 1–13, Dec. 2020.
- [26] H. Greenspan, B. V. Ginneken, and R. M. Summers, "Guest editorial deep learning in medical imaging: Overview and future promise of an exciting new technique," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1153–1159, Mar. 2016.
- [27] B. Luijten *et al.*, "Deep learning for fast adaptive beamforming," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2019, pp. 1333–1337.
- [28] Y. H. Yoon and J. C. Ye, "Deep learning for accelerated ultrasound imaging," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2018, pp. 6673–6676.
- [29] A. C. Luchies and B. C. Byram, "Deep neural networks for ultrasound beamforming," *IEEE Trans. Med. Imag.*, vol. 37, no. 9, pp. 2010–2021, Feb. 2018.
- [30] S. Vedula, O. Senouf, A. M. Bronstein, O. V. Michailovich, and M. Zibulevsky, "Towards CT-quality ultrasound imaging using deep learning," 2017, *arXiv:1710.06304*.
- [31] D. Hyun, L. L. Brickson, K. T. Looby, and J. J. Dahl, "Beamforming and speckle reduction using neural networks," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 66, no. 5, pp. 898–910, May 2019.
- [32] R. J. G. van Sloun, O. Solomon, M. Bruce, Z. Z. Khaing, Y. C. Eldar, and M. Misch, "Deep learning for super-resolution vascular ultrasound imaging," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2019, pp. 1055–1059.
- [33] X. Liu, T. Zhou, M. Lu, Y. Yang, Q. He, and J. Luo, "Deep learning for ultrasound localization microscopy," *IEEE Trans. Med. Imag.*, vol. 39, no. 10, pp. 3064–3078, Oct. 2020.
- [34] 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.*, vol. 39, no. 12, pp. 3855–3867, Dec. 2020.
- [35] U.-W. Lok *et al.*, "Fast super-resolution ultrasound microvessel imaging using spatiotemporal data with deep fully convolutional neural network," *Phys. Med. Biol.*, vol. 66, no. 7, Apr. 2021, Art. no. 075005.
- [36] L. Milecki *et al.*, "A deep learning framework for spatiotemporal ultrasound localization microscopy," *IEEE Trans. Med. Imag.*, vol. 40, no. 5, pp. 1428–1437, May 2021.
- [37] J. Jensen, "FIELD: A program for simulating ultrasound systems," *Med. Biol. Eng. Comput.*, vol. 34, no. 1, pp. 351–353, 1996.
- [38] J. A. Jensen and N. B. Svendsen, "Calculation of pressure fields from arbitrarily shaped, apodized, and excited ultrasound transducers," *IEEE Trans. Ultrason. Ferroelectr., Freq. Control*, vol. 39, no. 2, pp. 262–267, Feb. 1992.
- [39] M. R. Lowerison, C. Huang, Y. Kim, F. Lucien, S. Chen, and P. Song, "In vivo confocal imaging of fluorescently labeled microbubbles: Implications for ultrasound localization microscopy," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 67, no. 9, pp. 1811–1819, Sep. 2020.
- [40] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Intervent.*, 2015, pp. 234–241.
- [41] 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, no. 1, pp. 1929–1958, Jan. 2014.
- [42] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2015, pp. 1–15.
- [43] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
- [44] C. Demené *et al.*, "Spatiotemporal clutter filtering of ultrafast ultrasound data highly increases Doppler and fUltrasound sensitivity," *IEEE Trans. Med. Imag.*, vol. 34, no. 11, pp. 2271–2285, Nov. 2015.
- [45] P. Song *et al.*, "Improved super-resolution ultrasound microvessel imaging with spatiotemporal nonlocal means filtering and bipartite graph-based microbubble tracking," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 65, no. 2, pp. 149–167, Feb. 2018.

- [46] S. van der Walt *et al.*, “Scikit-image: Image processing in Python,” *PeerJ*, vol. 2, p. e453, Jun. 2014.
- [47] J. Kirkhorn, “Introduction to IQ-demodulation of RF-data,” IFBT, NTNU, 1999, vol. 15.
- [48] A. Han *et al.*, “Noninvasive diagnosis of nonalcoholic fatty liver disease and quantification of liver fat with radiofrequency ultrasound data using one-dimensional convolutional neural networks,” *Radiology*, vol. 295, no. 2, pp. 342–350, May 2020.
- [49] O. Solomon *et al.*, “Deep unfolded robust PCA with application to clutter suppression in ultrasound,” *IEEE Trans. Med. Imag.*, vol. 39, no. 4, pp. 1051–1063, Apr. 2020.
- [50] K. G. Brown, D. Ghosh, and K. Hoyt, “Deep learning of spatiotemporal filtering for fast super-resolution ultrasound imaging,” *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 67, no. 9, pp. 1820–1829, Sep. 2020.
- [51] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.



Xi Chen received the B.S. degree in computer engineering from the University of Illinois at Urbana–Champaign, Urbana, IL, USA, in 2019, where she is currently pursuing the Ph.D. degree with the Department of Electrical and Computer Engineering.

Her current research interest is super-resolution ultrasound localization microscopy.



Matthew R. Lowerison received the Ph.D. degree in medical biophysics from the University of Western Ontario, London, ON, Canada, in 2017.

He is currently with the Department of Electrical and Computer Engineering and the Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana–Champaign, Urbana, IL, USA. His current research interests include ultrasound microvessel imaging, super-resolution ultrasound localization microscopy, and ultrasonic characterization of tumor microenvironments.



Zhijie Dong (Graduate Student Member, IEEE) received the B.Eng. degree in information engineering from Southeast University, Nanjing, China, in 2017, and the M.S. degree in electrical and computer engineering from the University of Michigan, Ann Arbor, MI, USA, in 2018. He is currently pursuing the Ph.D. degree with the Department of Electrical and Computer Engineering, University of Illinois at Urbana–Champaign, Urbana, IL, USA.

His current research interests are ultrafast 3-D imaging and signal processing.

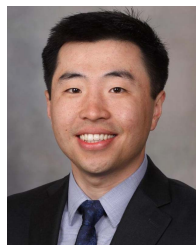


Aiguo Han (Member, IEEE) received the B.S. degree in acoustics from Nanjing University, Nanjing, China, in 2008, and the M.S. and Ph.D. degrees in electrical and computer engineering from the University of Illinois at Urbana–Champaign, Urbana, IL, USA, in 2011 and 2014, respectively.

He is currently a Research Assistant Professor with the Department of Electrical and Computer Engineering, University of Illinois at Urbana–Champaign. He is also a Faculty Member

with the National Center for Supercomputing Applications, University of Illinois at Urbana–Champaign. His current research interests include machine learning in ultrasound imaging and diagnostics, transcranial ultrasound imaging, quantitative ultrasound, and ultrasonic wave propagation in biological media.

Dr. Han is a Full Member of the Acoustical Society of America and a fellow of the American Institute of Ultrasound in Medicine. He was a recipient of the National Institutes of Health (NIH)/National Institute of Biomedical Imaging and Bioengineering (NIBIB) Trailblazer Award in 2022 and the American Institute of Ultrasound in Medicine New Investigator Award in Basic Science in 2016.



Pengfei Song (Senior Member, IEEE) received the B.S. degree in biomedical engineering from the Huazhong University of Science and Technology, Wuhan, China, in 2008, the M.S. degree in biological systems engineering from the University of Nebraska–Lincoln, Lincoln, NE, USA, in 2010, and the Ph.D. degree in biomedical engineering from the Mayo Clinic College of Medicine, Rochester, MN, USA, in 2014.

He is currently an Assistant Professor with the Department of Electrical and Computer Engineering, the Beckman Institute for Advanced Science and Technology, the Department of Bioengineering, and the Carle Illinois College of Medicine, University of Illinois Urbana–Champaign, Urbana, IL, USA. His research interests include super-resolution ultrasound imaging, ultrafast ultrasound, deep learning, 3-D ultrafast imaging, functional ultrasound, and ultrasound shear wave elastography.

Dr. Song is a fellow of the American Institute of Ultrasound in Medicine, a Senior Member of the National Academy of Inventors, and a Full Member of the Acoustical Society of America.