Super-Resolution Ultrasound Localization Microscopy Through Deep Learning

Ruud J. G. van Sloun, *Member, IEEE*, Oren Solomon, *Member, IEEE*, Matthew Bruce, *Member, IEEE*,
Zin Z. Khaing, Hessel Wijkstra, Yonina C. Eldar, *Fellow, IEEE*,
and Massimo Mischi, *Senior Member, IEEE*

***Abstract*— Ultrasound localization microscopy has enabled super-resolution vascular imaging through precise localization of individual ultrasound contrast agents (microbubbles) across numerous imaging frames. However, analysis of high-density regions with significant overlaps among the microbubble point spread responses yields high localization errors, constraining the technique to low-concentration conditions. As such, long acquisition times are required to sufficiently cover the vascular bed. In this work, we present a fast and precise method for obtaining super-resolution vascular images from high-density contrast-enhanced ultrasound imaging data. This method, which we term Deep Ultrasound Localization Microscopy (Deep-ULM), exploits modern deep learning strategies and employs a convolutional neural network to perform localization microscopy in dense scenarios, learning the nonlinear image-domain implications of overlapping RF signals originating from such sets of closely spaced microbubbles. Deep-ULM is trained effectively using realistic on-line synthesized data, enabling robust inference *in-vivo* under a wide variety of imaging conditions. We show that deep learning attains super-resolution with challenging contrast-agent densities, both *in-silico* as well as *in-vivo*. Deep-ULM is suitable for real-time applications, resolving about 70 high-resolution patches (128 × 128 pixels) per second on a standard PC. Exploiting GPU computation, this number increases to 1250 patches per second.**

***Index Terms*— Ultrasound, deep learning, super resolution, ultrasound localization microscopy, neural network.**

Manuscript received June 9, 2020; revised September 12, 2020; accepted November 9, 2020. Date of publication November 16, 2020; date of current version March 2, 2021. This work was supported by the Dutch Research Council (NWO) through the Research Program Rubicon ENW 2018-3 under Project 019.183.EN.014. *(Corresponding author: Ruud J. G. van Sloun.)*

Ruud J. G. van Sloun and Massimo Mischi are with the Department of Electrical Engineering, Eindhoven University of Technology, 5612 AZ Eindhoven, The Netherlands (e-mail: r.j.g.v.sloun@tue.nl).

Oren Solomon was with the Department of Electrical Engineering, Techion – Israel Institute of Technology, Haifa 3200003, Israel. He is now with the Center for Magnetic Resonance Research, Department of Radiology, University of Minnesota, Minneapolis, MN 55455 USA.

Matthew Bruce is with the Applied Physics Laboratory, University of Washington, Seattle, WA 98195 USA.

Zin Z. Khaing is with the Department of Neurological Surgery, University of Washington, Seattle, WA 98195 USA.

Hessel Wijkstra is with the Department of Electrical Engineering, Eindhoven University of Technology, 5612 AZ Eindhoven, The Netherlands, and also with the Academic Medical Center, Department of Urology, University of Amsterdam, 1012 WX Amsterdam, The Netherlands.

Yonina C. Eldar is with the Faculty of Math and Computer Science, Weizmann Institute of Science, Rehovot 7610001, Israel.

Digital Object Identifier 10.1109/TMI.2020.3037790

I. Introduction

ROBUST, precise, fast and cost-effective in-vivo microvascular imaging is a cornerstone for clinical management of diseases that are hallmarked by impaired or remodeled microvasculature, such as angiogenesis in cancer [1]. Contrast-enhanced ultrasound is a cost-effective modality, which combines ultrasound imaging with enhancement of blood through the use of ultrasound contrast agents, inert gas microbubbles that are sized similar to red blood cells [2]. Nevertheless, the spatial resolution of conventional contrast-enhanced ultrasound imaging is bound by the diffraction limit of sound. Being primarily determined by the adopted wavelength, this limit in practice manifests itself as an inherent trade-off between resolution and penetration depth, since acoustic waves suffer from increasing amounts of absorption at higher frequencies.

Recently, this trade-off was circumvented through the introduction of super-resolution ultrasound imaging, where Nobel-prize-winning super-resolution concepts from optics (e.g. Photoactivation Localization Microscopy - PALM) are exploited and translated into the ultrasound imaging domain with the aim of achieving sub-wavelength resolution images of flowing microbubbles. [3], [4]. Today, the most common approach for achieving ultrasound super resolution is Ultrasound Localization Microscopy (ULM). In ULM, individual microbubbles are pinpointed from diffraction-limited ultrasound data across a large sequence of imaging frames with sparse microbubble population, i.e., using low contrast-agent concentration. By combining all these position estimates into one frame, a super-resolved image is produced [5]–[8].

To boost the amount of imaging frames in a shorter time span, Errico *et al.* acquired over 75,000 frames of a fixed rat brain using an ultrafast ultrasound imaging scheme across 2.5 minutes [9]. Yet, for adoption in a clinical environment, imaging for 2.5 minutes is still lengthy when considering inevitable artifacts and drift due to patient motion.

Beyond mapping the accumulated microbubble localizations, one can also track detections across multiple frames [8], [10]. This not only permits rendering of velocities and interpolation of the detections, but also removal of spurious clutter localizations. While nearest-neighbor data association schemes are common [8], [9], [11], Ackermann *et al.* showed that more robust assimilation can be obtained through dedicated motion models in a modified Markov chain Monte Carlo framework [12]. The latter was recently evaluated in a clinical setting [13], advocating the value of ULM for diagnostics in the context of vascular phenotyping of tumors. At the same time, the authors underline a major hurdle to overcome if ULM is to be broadly implemented in clinical practice: long measurements (even of about one minute) suffer from significant tissue motion. While mild in-plane motion can be compensated with sub-wavelength accuracy [14], registration errors for relatively large movements can be much larger than the localization precision of ULM. Moreover, out-of-plane components cannot be corrected in 2D.

ULM avoids the trade-off between resolution and penetration depth, but it gives rise to a new trade-off that balances localization precision, microbubble concentration and acquisition time. High image fidelity is attained when large amounts of bubbles are localized with high precision, posing a lower bound on the acquisition time of ULM. This bound can be relaxed significantly when high concentrations are used, with many high-precision localizations per frame. Moreover, the probability of actually filling all arterioles with microbubbles in a certain timespan increases with higher concentrations. Obtaining the required localization precision in data with such a dense population of microbubbles with overlapping signals is a challenging task however, yielding a scenario in which single-bubble localization algorithms break down. As such, standard ULM methods adhering to this microbubble-sparsity constraint still require long acquisition times, impairing broad translation in a clinical setting, where high contrast concentrations, limited time, significant organ motion and lower frame-rate imaging are common.

Recently, an acoustic counterpart of PALM, termed acoustic wave sparsely activated localization microscopy was proposed [15], [16]. This approach elegantly leverages acoustic waves to sparsely and stochastically activate nanodroplets. Unlike standard ULM, it allows for the use of a high concentration of contrast agents, thereby more effectively covering the full vascular bed. Such agents are however currently not clinically approved.

ULM algorithms based on deconvolution [17], [18] or sparse recovery [19]–[21] have been developed specifically to cope with the overlapping point spread functions (PSFs) of multiple microbubbles. Sparse recovery strategies pose the localization task as an inverse problem with a structural sparsity prior [19], in which bubbles with overlapping PSFs but distinct sparse locations on a dense grid can be resolved. Sparsity-based ultrasound super-resolution hemodynamic imaging (SUSHI) [20] expands upon this by considering the inherent temporal structure of the data, and performs sparse recovery in the temporal correlation domain. While successful localization of densely-spaced emitters has been demonstrated, even highly optimized fast recovery techniques involve a time-consuming iterative procedure. Fourier-domain fast iterative shrinkage-thresholding [20] improves dramatically over an image domain formulation, but computational time grows significantly with the field of view. In addition, the optimal settings of sparse-recovery methods can vary across frames due to e.g. time-varying microbubble densities.

Here we present Deep-ULM [22], an ultrasound localization microscopy strategy based on deep learning [23], designed and trained to cope with high-concentration contrast-enhanced ultrasound (CEUS) acquisitions. We harness a fully convolutional neural network for super-resolution image reconstruction from dense images containing many overlapping microbubble signals, and show that the method is robust to varying imaging conditions and microbubble concentrations. The output of the network is not a set of position vectors but rather a high-resolution image in which the pixel values reflect recovered backscatter intensities. Our approach shares similarities with a recently introduced deep learning technique for single molecule fluorescence microscopy [24], albeit in a completely different field and setting. Image recovery using Deep-ULM is fast, and can be applied to any CEUS acquisition in which the PSF can be estimated (e.g. from a-priori characterization or simply a few sparse frames), requiring minimal user expertise and no manual tweaking. Using controlled in-silico experiments we show that our approach outperforms both standard ULM as well as sparse recovery methods for high densities, and subsequently demonstrate that the detection process generalizes well to in-vitro and in-vivo applications.

II. Methods

A. Synthetic Training Data Generation

Deep learning typically relies on the exploitation of large, representative datasets that enable the training of a robust network that generalizes well when employed in practice. While measuring sufficiently diverse CEUS inputs along with their super-resolved outputs is not trivial, the generation of realistic synthetic training data is in fact rather simple. To this end, we sample the real system PSF from CEUS images using a tool that enables manual selection of a few individual microbubbles across a few frames. We then automatically fit a rotated anisotropic Gaussian PSF model to the data to extract the PSF parameters $\hat{\boldsymbol{\phi}}$. That is, for each acquisition setting, we selected a few sparse CEUS frames and fitted the PSF model to a set of well-isolated microbubbles. Using our imaging protocol, we did not observe significant deviation from Gaussian PSF shapes. For the in-vivo data, the median relative root mean squared error (RMSE) of the PSF fit used to generate training data was 6.2%

The generation of new synthetic data for training is straightforward, with each corresponding low-resolution CEUS input and super-resolved target represents the basis for a diverse training dataset involving a number of variations.

We generated target patches containing multiple microbubbles with various intensities on a high-resolution grid. A broad spectrum of contrast-agent-concentrations was simulated, randomly drawn from a uniform distribution between 0 and 260 microbubbles/cm^2. Relative backscatter intensities were also drawn randomly, reflecting the backscatter intensity variations of a polydisperse microbubble population imaged at various distances from the elevational beam, and ranged between 0.4 and 1 (a.u.). The set of microbubble locations was then converted to radiofrequency CEUS signals using the radiofrequency-modulated PSF. At this stage, the signatures of closely-spaced individual microbubbles create distinct interference patterns. These radiofrequency data were then envelope detected through the Hilbert transform, and subsequently down-sampled to an 8 times courser grid to yield the input image patches. Variance of the PSF and uncertainty in its estimate was incorporated in the training procedure by introducing variance in the PSF parameters $\boldsymbol{\phi}$ through a multiplicative random component, i.e.:

$$\boldsymbol{\phi} = \hat{\boldsymbol{\phi}}\left[1 + \mathcal{N}(\mu = 0, \sigma = 0.1)\right]. \tag{1}$$

While we here assume that tissue clutter is suppressed prior to ULM processing (e.g. through singular value filtering or the use of contrast-enhanced imaging sequences), in practice some clutter (and noise) may remain. Therefore, to increase the model's robustness after training, we added white and colored background noise with relative standard deviations of 2% and 5%, respectively. Colored noise was produced by spatially filtering white noise with a 2D Gaussian having a standard deviation of 1.2 pixels.

B. Deep Neural Network Architecture

A computational model should then be able to learn representations from this data through a hierarchy of nonlinear operations, having the capacity to perform an end-to-end mapping from diffraction-limited CEUS to super-resolved images. For this purpose, we adopted a fully convolutional network architecture based on an encoder-decoder structure [25]. The encoder is trained to optimally convert the ultrasound image space into a feature space that contains all relevant microbubble position information, through convolutions and down-sampling operations. The decoder is trained to transform this feature space into a high-resolution, super-resolved frame via up-sampling and transposed convolutions.

Specifically, the encoder follows a contracting path which consists of 3 layer-blocks, each block comprising two 3 × 3 convolution layers with leaky rectified linear unit (ReLU) activations, and one 2×2 Max-pooling operation. We use leaky ReLUs [26] rather than regular ReLUs across all convolution layers in the network to avoid inactive neurons/nodes that effectively decrease the model capacity. In addition, batch normalization is used before all activations to boost the network's trainability by enabling higher learning rates and requiring less-strict hyper-parameter optimization [27].

The subsequent latent layer includes two 3×3 convolutional layers, followed by a dropout layer (probability 0.5) which randomly disables about 50% of the latent features during training. This latent space is then transformed to a high-resolution localization image by the decoder. The decoder again consists of 3 blocks; the first two blocks encompassing two 5 × 5 deconvolution layers [28] of which the second has an output stride of 2 rather than 1, followed by a 2 × 2 nearest-neighbour up-sampling layer which simply repeats the image rows and columns. The last block consists of two deconvolution layers, of which the second again has an output stride of 2, preceding another 5 × 5 convolution which maps the feature space to a single-channel image through a linear activation function. The full network effectively scales the input image dimensions up by a factor 8.

C. Training Strategy

We used the Adam optimizer with learning rate 0.001, and trained the network on batches of 256 synthetic imaging frames across 20,000 iterations to minimize the following cost function, similar to the one proposed in [24]:

$$c(x, y|\theta) = \| f(\mathbf{x}|\theta) - G_{\sigma} * \mathbf{y}\|_2^2 + \lambda \| f(\mathbf{x}|\theta)\|_1 \tag{2}$$

where $\mathbf{x}$ and $\mathbf{y}$ are input CEUS and target super-resolved patches, respectively, $f(\mathbf{x}|\theta)$ is the nonlinear neural network function with parameters (weights and biases) θ, and λ is a regularization parameter that promotes network predictions that yield sparse images, and was (conservatively) set equal to 0.01. The operator G_{σ} denotes a 2D Gaussian filter of which the standard deviation was set to 1 pixel at the start of training. In practice, we observed that applying such a mild 2D filtering operation on the sparse target data strongly improved training stability; small localization errors are penalized less than large errors. This mean-squared-error-based regression strategy enables joint estimation of microbubble locations and their backscatter intensities. The latter is particularly useful to emphasize localizations near the elevational beam axis during image reconstruction.

As a unique batch of data is generated on-line for each iteration, the model's robustness and capacity to generalize to new cases is drastically improved. The latter is further supported by applying dropout during the training phase, randomly disabling features at the encoded latent space with a probability of 0.5.

Training (and inference) were run on a computation server, equipped with an NVidia Titan X Pascal that has 12 GB of video memory.

D. Reference ULM Implementations

1) Standard ULM: Standard ULM was implemented using a centroid localization approach, largely following the methodology described by Errico *et al.* [9] We first up-sample the images by a factor 8 (equal to the grid up-sampling of Deep-ULM) and deconvolve them with a Gaussian low-pass filter having a slightly lower standard standard deviation than the imaging PSF to avoid instability, and subsequently keep only values above 50% of the 98th percentile. We then perform a morphological opening operation to remove spurious peaks, after which we detect the local maxima. We finally select a small area of 24 × 24 pixels (i.e. 3 × 3 pixels on the original data) around the local maxima and therefrom compute the local image centroids. The code was written in Python, using the scikit-image and scipy modules.

2) Sparse Recovery Based ULM: Sparse recovery based ULM [19] approaches the microbubble localization task as an inverse problem, by modelling each image frame as a superposition of translated and scaled PSFs according to microbubble locations and backscatter amplitudes on a high-resolution grid. Assuming that the microbubbles are smaller than a pixel and sparsely distributed across the image, the following regularized

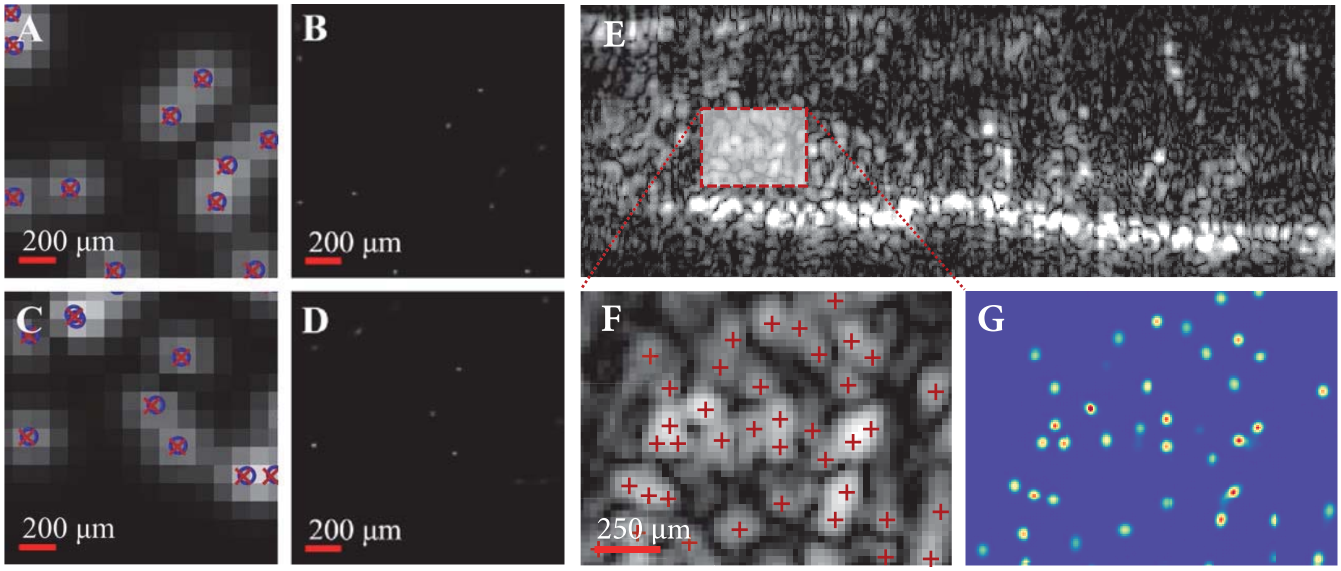

Fig. 1. Deep-ULM frame-based localizations for synthetic and *in-vivo* datasets. (A, C) Examples of synthetic datasets with different microbubble densities, generated using a point-spread-function model estimated from clinically acquired ultrasound data. (B, D) Corresponding Deep-ULM recoveries on a 15-μm spaced grid. The true locations of the microbubbles are marked as blue circles, and Deep-ULM predictions (on a discrete grid) as red crosses. (E) An illustrative example frame in an in-vivo CEUS loop, (F) localizations within an area of interest of (E), and (G) corresponding Deep-ULM recoveries from which these localizations are deduced.

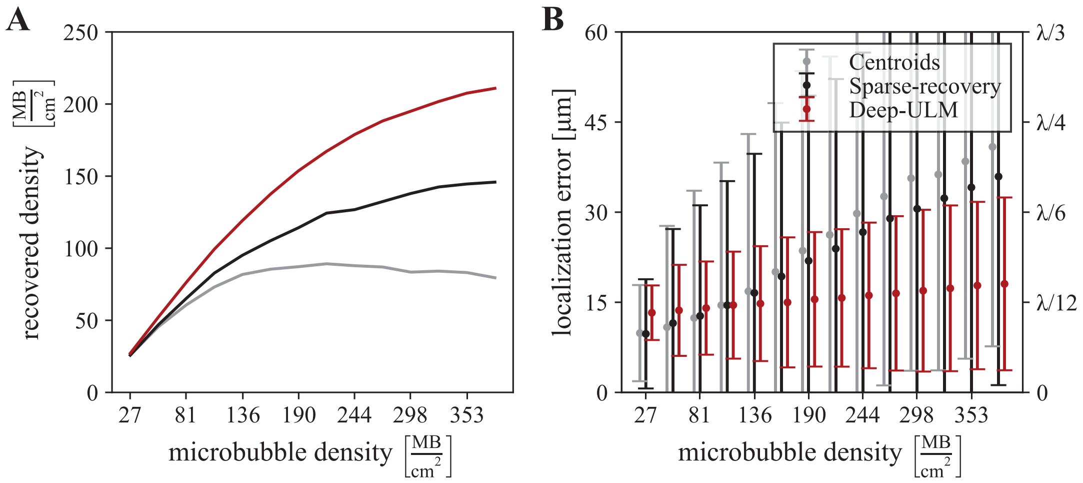

Fig. 2. Detection rate and localization precision of Deep-ULM (red) compared to ULM based on deconvolution and centroid estimation (gray) and sparse recovery (black). (A) Recovered density as a function of simulated microbubble (MB) density, and (B) corresponding median localization errors with bars representing the standard deviation. Note that Deep-ULM's localization errors are very close to the grid spacing (15 μm), and well below the wavelength (214 μm), even for high microbubble densities.

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

$$\mathbf{x} = \arg\min_{\mathbf{x}} \left(\|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 + \lambda \|\mathbf{x}\|_1 \right) \tag{3}$$

where $\mathbf{x}$ is the microbubble reflectivity vector on a high-resolution grid, $\mathbf{y}$ is the vectorized image frame, and $\mathbf{A}$ is the measurement matrix in which each column is a shifted version of the PSF. To solve (3), we employed a highly optimized Fourier domain implementation of the Fast Iterative Shrinkage-Thresholding Algorithm (FISTA). We used a grid up-sampling factor of 8, and λ was set to 0.01.

3) Sparse Recovery via SUSHI: Sparsity-based super-resolution hemodynamic imaging [20] aims at producing time-lapse super-resolved sequences of fast hemodynamic changes. It was implemented by first dividing the acquired CEUS clip into small movie segments and estimating the pixel-wise variance of each segment, resulting in a variance time-lapse movie of the MBs, which exhibits improved spatial resolution and background rejection. To further improve the spatial resolution beyond the acoustic diffraction limit, sparse recovery is then performed on the variance images, by using a similar formulation to (3), again with an up-sampling factor of 8.

III. Validation Methodology

A. In-Silico Experiments

Flow of microbubbles through an artificial vascular network was simulated by propagating particles along streamlines with

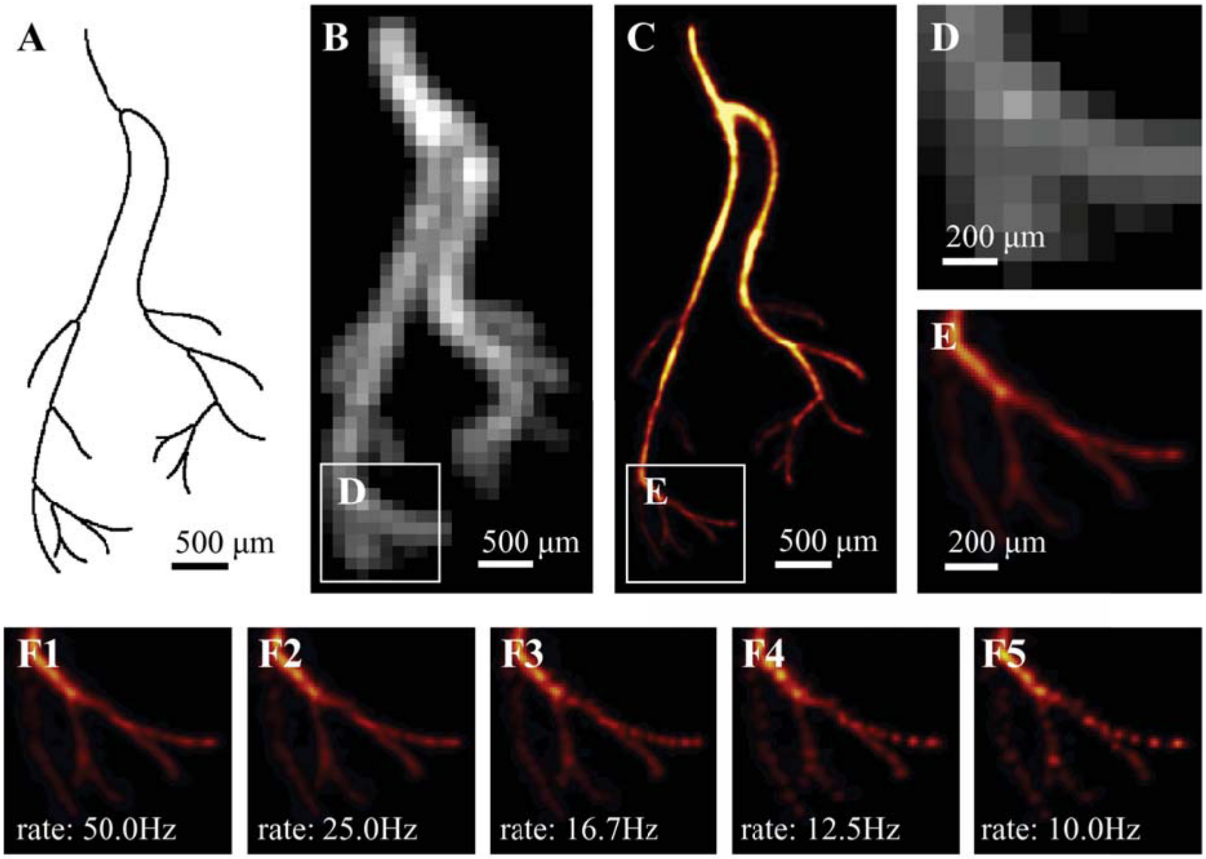

Fig. 3. Deep-ULM on *in-silico* flow data compared to diffraction limited imaging. (A) Simulated vascular skeleton, (B) diffraction-limited maximum intensity persistence image, (C) Deep-ULM super-resolution reconstruction and (D,E) zooms of (B,C). (F1-F5) Deep-ULM reconstruction across 12 seconds with decreasing frame rates, displaying how dense localization on high-concentration simulations maintains reasonable fidelity even when very limited imaging frames are available. The actual physiological requirement is that vessels are sufficiently filled by the agent within the imaging time, which is relaxed by the use of high concentrations.

a specific velocity, comprising a deterministic part, as well as a multiplicative random component, i.e.: $\boldsymbol{v}(x, y, t) = \max(0, \boldsymbol{v}_{det}(x, y, t)) \cdot \mathcal{N}(\mu = 1, \sigma = 1)$. 140 particles were infused at the injection point by randomly drawing particle injection times from a uniform distribution across a 12-second timespan, leading to the generation of approximately 12 particles per second. Ultrasound imaging of this process was simulated by modelling the scanner's PSF as a bivariate Gaussian, modulated by the transmit wavelength. The standard deviation in the axial and lateral direction were set to 0.14 and 0.16 mm, respectively. The modulation frequency was set to 7 MHz, the second-harmonic response of a nonlinearly resonating microbubble to an ultrasound transmit frequency of 3.5 MHz after fundamental mode suppression (e.g. bandpass filtering or pulse inversion). The image was formed by demodulating the radiofrequency scan lines originating from the summed contributions of all microbubble responses trough the Hilbert transform. Frames were constructed at a rate of 100 Hz, and the pixel dimensions were 0.12 × 0.12 mm. Note that through the above we do not explicitly simulate the complex microbubble physics, but rather directly model scanner's PSF.

B. In-Vitro Experiments

We then designed an in-vitro experiment by imaging microbubbles flowing through a 0.3-mm cross-channel flow phantom (20% Polyacrylamide). To that end, we performed CEUS imaging (power modulation with pulse inversion) using the Vantage ultrasound system (Verasonics, Seattle, WA), equipped with an L11-4v probe transmitting a single cycle burst at 3.5 MHz. The frame rate was 100 Hz and we acquired data across 10 seconds, starting from the moment when steady flow was observed. The phantom was infused with a 1/500 dilution of 1 mL SonoVue (Bracco). The pixel dimensions were 0.15 × 0.15 mm. We used a singular-value-decomposition (SVD) filter to suppress clutter and enhance microbubble signal.

C. In-Vivo Experiments

1) Data Acquisition and Pre-Processing: The animal experiments were performed at the University of Washington, Seattle, WA, USA, with prior approval from the University of Washington's Institutional Animal Care and Use Committee (IACUC). All appropriate guidelines from the University's Animal Welfare Assurance (A3464-01) as well as the NIH Office of Laboratory Animal Welfare (OLAW) were followed. A 250-grams female Sprague Dawley rat (Harlan Labs, Indianapolis, IN) was anesthetized using isoflurane (5 % to induce and 2.5 - 3 % to maintain), and the area overlying the T7/T8 vertebrae was shaved, cleaned and sterilized. After dissection of paraspinal muscles, a laminectomy was performed

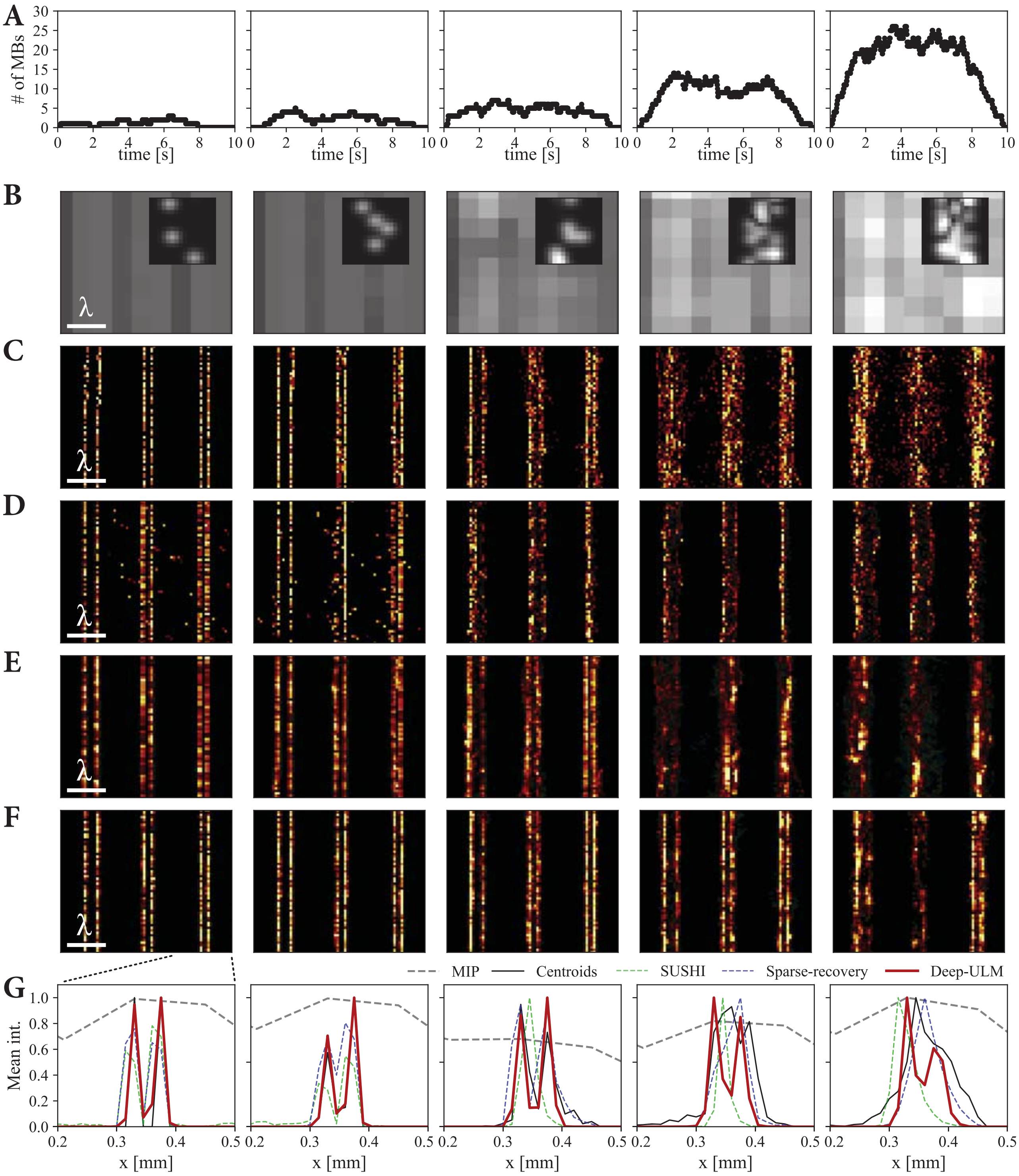

Fig. 4. In-silico evaluation of Deep-ULM compared to ULM based on centroids and sparse-recovery for parallel microbubble streams with varying densities and interfering point spread functions. (A) Number of microbubbles in the frame across time, (B) Maximum intensity persistence images (MIP), along with individual example frames (inset), (C) standard centroid ULM images, (D) SUSHI, (E) sparse recovery ULM, (F) Deep-ULM, and (G) mean lateral profiles of the two rightmost streams for all techniques. Note that Deep-ULM attains better sub-wavelength separation for higher densities than the other methods.

to expose the spinal cord from T6 to T10. A compression-type lesion was produced [29]. High-frame-rate CEUS acquisitions of the cord were performed with a Vantage ultrasound research platform (Verasonics, Seattle, WA, USA), using a linear array transducer (Vermon, Tours, France). Details of the acquisition are further described in [30]. The transmit center frequency was 15 MHz with 90% bandwidth in receive. An intravenous injection of 0.15-mL Definity®(Lantheus, New Jersey, USA) contrast agent followed by a 0.2-mL saline flush was administered via the tail vein using a catheter (BF-27-01, SAI Infusion

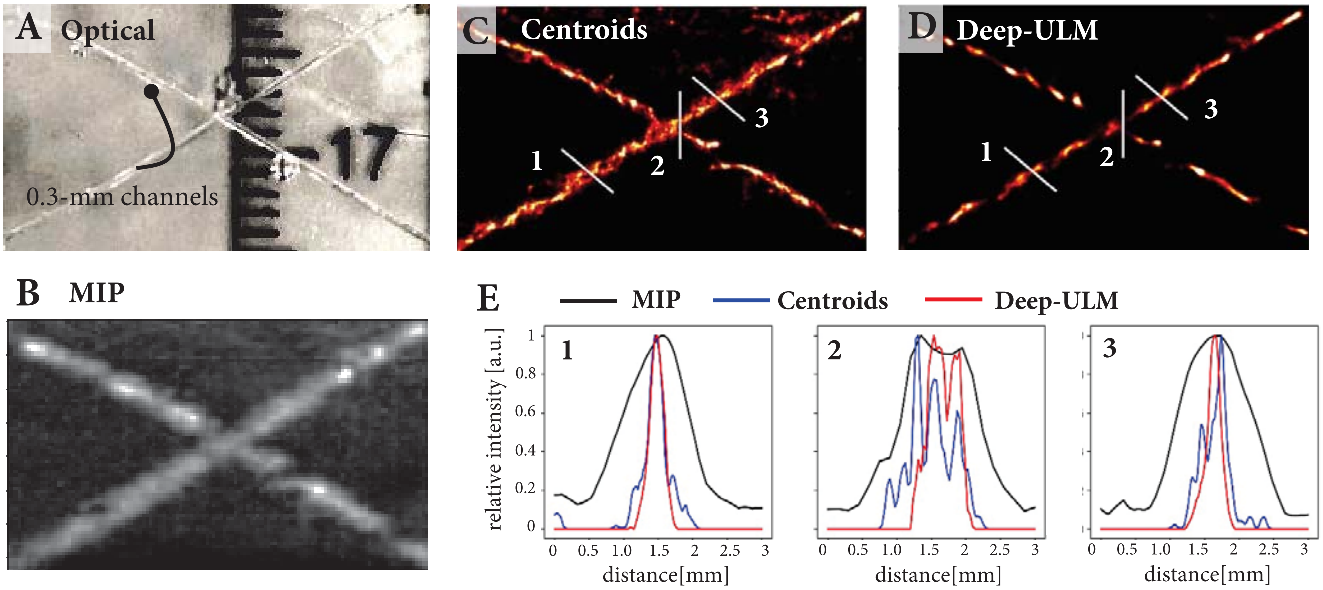

Fig. 5. *In-vitro* Deep-ULM in a 300-μm crossed-channel phantom. (A) Optical reference image (B) Maximum intensity projection image, (C) Centroid localization image, (D) Deep-ULM image, and (E) Intensity profiles of channels.

Technologies, Lake Villa, IL, USA). We then waited about 3 minutes for the concentration to drop. A 5-angle plane wave amplitude modulated sequence was adopted [30], using delay-and-sum beamforming in receive. The plane wave frame rate was 30 kHz to avoid motion artifacts when compounding the 5 angles. Compounded images were obtained at a rate of 400 Hz. The IQ data were then wall filtered (Butterworth high-pass of order 20 with a cutoff at 50Hz) and SVD-filtered to suppress tissue clutter and enhance the response to microbubbles, and subsequently envelope detected through the Hilbert transform. Details of the microbubble composition and concentration of Defininity can be found in the package insert [31].

2) Motion Compensation: For motion compensation, we first extracted the tissue signal from by performing a singular value decomposition on the space-time data (i.e. a Casorati matrix of which the columns are vectorized frames), and attributing the first few singular values (describing components with high spatiotemporal coherence) to tissue [19]. We then computed the required rigid transformations that map each resulting frame to the first frame in the loop. The maximum measured motion was about 75 μm for the rat spinal cord sequence.

IV. Results

A. Synthetic Data

Fig. 1 shows several examples of Deep-ULM applied to such synthetic datasets, with the reconstruction being on an 8 times up-sampled grid. Recovery of a high-resolution 128 × 128 patch using Deep-ULM takes less than 0.8 milliseconds on a GPU-equipped workstation, and about 14 milliseconds on a standard PC. The training and testing loss (a measure of resemblance between the network predictions and ground truth) monotonically decrease as a function of the number of iterations, showing no sign of overfitting. This can be attributed to the on-line synthetic data generation and dropout-based regularization. During testing, dropout is disabled, further pushing the loss down as a consequence of effective model ensemble averaging [32].

Fig. 2 displays the recovered density and localization precision of Deep-ULM compared to an optimized standard ULM method based on deconvolution and centroid localization [9] and ULM based on sparse recovery [19] as a function of simulated microbubble density. A microbubble is only considered detected if a localization was obtained close to its true location, within 30 μm (about 1/7th of the wavelength). To determine the localization precision, each identified microbubble is associated to the closest ground-truth microbubble position, and their Euclidian distance is calculated. For low densities, all methods perform similarly well, with Deep-ULM displaying a very slight increase in localization error. When the density increases however, sparse recovery adequately detects more microbubbles than standard ULM, while Deep-ULM significantly outperforms both sparse recovery and standard ULM in terms of detection rate and localization precision. Deep-ULM is moreover about 4 orders of magnitude faster than iterative sparse-recovery procedure.

B. In-Silico Flow Through a Branching Vessel

We then tested Deep-ULM on a simulated CEUS acquisition of microbubble flow through a realistic bifurcating vessel. Deep-ULM detects over 20000 microbubbles across the generated 1200 frames, finely delineating the vascular architecture as shown in fig. 3. Because the method detects many microbubbles per frame, reconstructions at lower frame rates and shorter timespans become feasible. The impact of using such a reduced amount of imaging frames is evaluated in fig. 3F, showing that reconstructions with as few as 120 frames already display good fidelity.

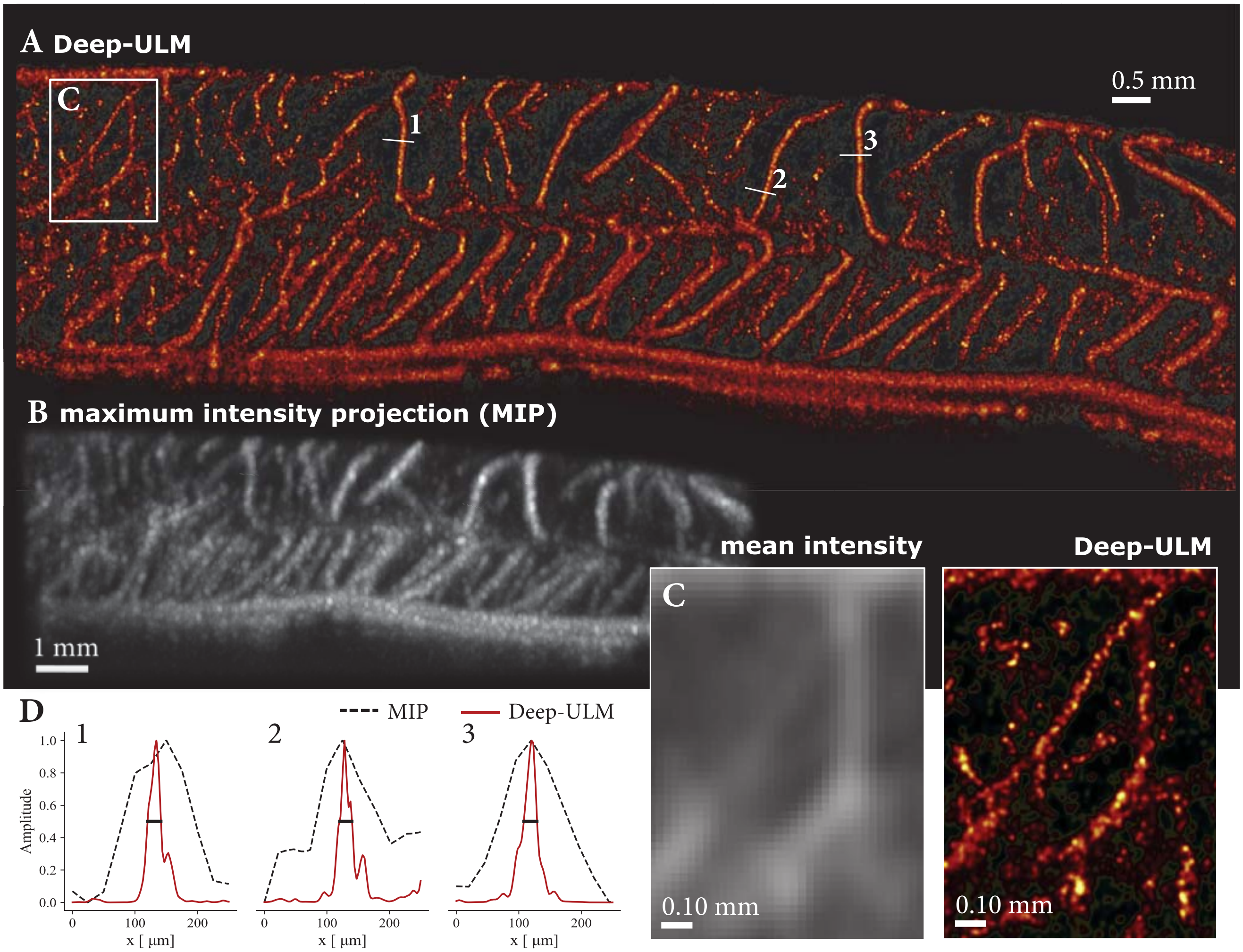

Fig. 6. *In-vivo* Deep-ULM in a rat spinal cord. (A) Deep-ULM across 8-seconds acquired at a frame rate of 400 Hz, along with a (B) Maximum intensity persistence (MIP) image and (C) zoomed region of interest for Deep-ULM and the corresponding mean intensity image. (D) Intensity profiles of vessels, with their full-width-half maxima indicated by black horizontal lines, being 21 μm, 19 μm, and 20 μm for profiles 1, 2, and 3, respectively.

We then compare standard ULM [9], frame-by-frame sparse-recovery [19], SUSHI [20], and Deep-ULM on sub-diffraction spaced parallel streams. To that end, we simulated a 10-second ultrasound acquisition of microbubbles moving at 1 mm/s through 3 pairs of parallel vessels (separated by $\lambda/3$, $\lambda/4$, and $\lambda/5$, respectively) for increasing densities. From fig. 4C, we can observe that standard ULM again performs very well for low densities, but yields many false localizations when the number of microbubbles per area increases. SUSHI however displays good performance across all densities (fig. 4D); although it does not detect the most closely spaced vessels for the higher densities, it delineates the $\lambda/3$- and $\lambda/4$-separated vessels across all experiments. Sparsity-driven ULM (fig. 4E) remains more robust than standard ULM up to higher densities, but is less stable then SUSHI for the densities used in the two rightmost panels. Despite the use of a highly-optimized Fourier-domain implementation [20], the sparsity-based methods are about four orders of magnitude slower than inference with GPU-accelerated Deep-ULM (6 seconds/frame compared to 0.6 milliseconds/frame on our system). In this specific experiment, Deep-ULM outperforms the sparse recovery methods, which we attribute to learning of the image-domain implications of overlapping RF signals from closely-spaced microbubbles (figs. 4F-G).

C. In-Vitro Crossed-Channel Phantom

A comparison of the optical image, maximum-intensity-persistence CEUS, centroid ULM, and deep-ULM of the cross-channel phantom is given in Figure 5.

Qualitatively, Deep-ULM reaches the resolution required to adequately represent the physical channel dimensions. This is also evident from the intensity profiles given in Figure 5E. Compared to standard centroid localization, Deep-ULM appears less sensitive to noise, which we attribute to the denoising prior of the autoencoder structure.

From Figures 5 A and D one can observe that the dimensions of the tubes by optical imaging and Deep-ULM are similar. The intensity profiles displayed in Fig 5E-2 are taken where the two channels physically start to overlap in the imaging plane. At that point, the spacing between the centerlines of the two peaks in the deep-ULM line profile is

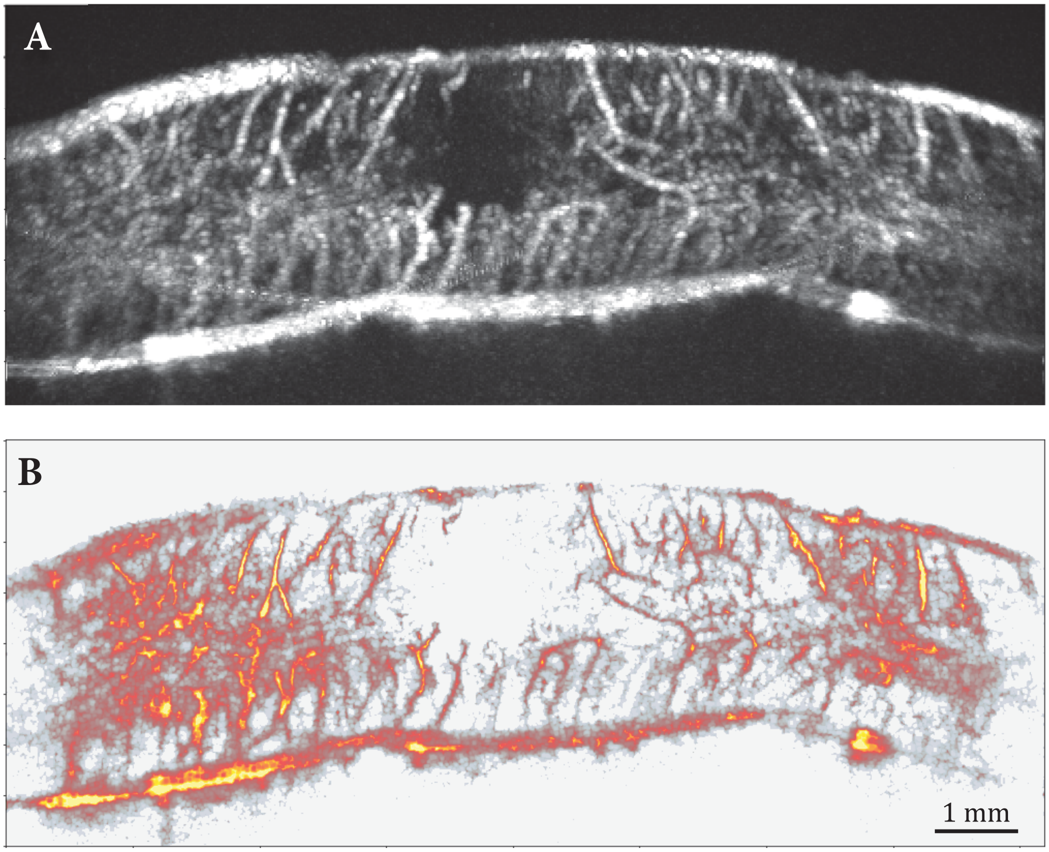

Fig. 7. *In-vivo* Deep-ULM of an injured rat spinal cord. (A) Maximum intensity persistence image and (B) Deep-ULM image, clearly depicting the vascular deficit originating from a compression injury.

only 280μm. This visible separation, despite physical channel overlap, is likely due to the larger concentration of microbubbles in the center of the channels.

D. In-Vivo Rat Spinal Cord With High-Frame-Rate CEUS

We proceed to apply Deep-ULM *in-vivo*, using a high-frame-rate (400 Hz) CEUS scan of a rat spinal cord acquired with a Verasonics Vantage ultrasound research scanner [29]. We retrained the neural network based on an estimate of the PSF parameters of this system (obtained using the tool described earlier), and performed Deep-ULM on an 8-second acquisition to obtain a super resolved image. Recovery on the 8 times up-sampled (4096 × 1328) grid took 100 milliseconds per complete imaging frame. In total, over 600,000 localizations were attained. Localizations for an illustrative example frame are shown in fig. 61E-G. The method robustly resolves individual microbubbles on dense data with significant overlaps in their PSFs. Fig 6 shows how the method achieves super-resolution image recovery, resolving vessels beyond the diffraction limit. A spatial resolution of about 20-30 μm was achieved, estimated by measuring the full-width-half-maxima of several profiles of arterioles that carried a sufficiently large amount of microbubbles (see fig. 6D). This was a 4-5 fold improvement with respect to the image resolution of the maximum intensity projection image for these profiles. Comparing the full-width-half-maxima (21, 19 and 20 μm) to the predominant diameter of vessels in the spinal cord as identified by μCT (around 20 μm), we observe good agreement [33]. Fig. 7 shows the application of Deep-ULM on a 12-second CEUS acquisition of a rat spinal cord following an injury [29], clearly depicting the resulting vascular deficit.

V. Conclusion and Discussion

Ultrasound localization microscopy (ULM) has enabled researchers to achieve extraordinary and unprecedented resolution in vascular ultrasound, no longer hindered by the diffraction limit of sound. Yet, its harsh limitations in terms of allowable contrast-agent concentrations lead to long acquisition times, and have spurred research in the direction of solving the high-density problem. Although recent methods exploiting sparse-recovery strategies do indeed allow for higher concentrations [19], [20], they come at a high computational cost. In this paper, we show how deep neural networks can learn how to perform efficient ULM in challenging high-density scenario's, requiring nothing more than an estimate of the local PSF of the image system. Notably, the network architecture, settings, and training procedure remained unchanged across the different in-silico and in-vivo experiments; the method was simply used "as is".

Deep-ULM uses a convolutional neural network that is trained using synthetic datasets that consist of ground truth microbubble backscatter amplitudes on a fine grid along with their corresponding CEUS ultrasound images. The method's performance depends on the capacity of the network to learn how to solve this sparse-recovery problem in an efficient manner, by learning a nonlinear function that maps low-resolution B-mode images to super-resolved localizations. On the other hand the quality and representability of the synthetic data for the actual acquisitions used during inference plays a major role. To improve robustness with respect to the latter, uncertainty in the estimated ultrasound scanner parameters is incorporated by introducing a variance in the adopted PSF model parameters across the dataset.

A polydisperse set of microbubble signals, with bubbles being on or off resonance, ringing, interacting and more, is likely causing a large degree of variability in the behavior of the received signals. We here aimed to model those variations by introducing stochasticity in the parameters of a frequency-modulated anisotropic Gaussian PSF. Although this displays promising results in-vivo, it likely does not capture the full extent of variations due to the above said physical properties. In future work, devising a more realistic physics-driven microbubble model for training Deep-ULM may further boost its performance.

The neural network was designed based on an encoder-decoder principle to perform the end-to-end mapping between the input images and their targets; an architectural approach that has been widely adopted for various segmentation and image enhancement problems [25], [34], [35]. The total number of convolutional layers in our deep net amounts to 15, which yielded sufficient capacity to perform the desired sparse-recovery functionality, while not overfitting. The latter thrives with our on-line training data generation and the use of a relatively thin bottleneck latent layer with 50% dropout, effectively exploiting an ensemble of trained encoder models at the inference stage. With this deep network, super-resolution recovery of low-resolution images takes about 14 milliseconds per patch on a regular PC, and even less than 0.8 milliseconds when exploiting GPU computation. One could push this number further down by using model compression techniques [36], such as learning less complex models to replicate the current model's functionality through knowledge distillation [37].

Deep-ULM effectively learns the significant nonlinear image-domain implications of overlapping RF signals originating from closely spaced microbubbles, which up till now have hampered high-density super-localization in ultrasound. Being trained to deal with concentrations as high as 260 microbubbles per cm^2, our experiments show that Deep-ULM indeed performs well for high densities; conditions in which single particle localization algorithms based on image centroids break down. While sparsity-driven algorithms [19], [20] improve upon centroid-based localization, Deep-ULM outperforms them in both localization precision and speed, being about 4 orders of magnitude faster. Nevertheless, also for Deep-ULM, higher densities pose greater challenges for the algorithm. Although the maximum admittable concentration given a desired precision is significantly boosted, it will inherently depend on the signal to noise ratio of the ultrasound acquisition. In addition, at some point bubble-bubble interaction and multiple scattering may play a role. The latter is however expected to be minimal, even at high concentrations [38].

For very low concentrations, localization by Deep-ULM was found to be slightly less precise than centroid-based localization. Since Deep-ULM is trained to perform well across a large range of concentrations, gradient-based optimization of the deep neural network parameters is naturally biased towards achieving low cost at high concentrations, as having more microbubbles in the field of view yields a stronger train signal.

The ability to handle such high microbubble concentrations has significant implications for translation into clinical applications. Alleviating the very demanding temporal constraints of standard ULM by faster coverage of the relevant arterioles is a necessity rather than a luxury in many diagnostic settings, where time is scarce and the impact of organ movement across the acquisition becomes significant. With ultrafast high-frame-rate ultrasound imaging architectures finding their way into clinical scanners, a super-resolution method requiring less than 1000 frames can achieve sub-second temporal resolution, thereby drastically improving real-time clinical utility while at the same time mitigating those severe motion artifacts.

Opacic *et al.* [13] showed that application of super-resolution ultrasound in a clinical setting is feasible when using dedicated motion compensation and frame clustering strategies. While this holds great promise, a significant amount of acquisitions were excluded due to severe motion artifacts that could not be compensated for. Combining the above methodology with methods that exploit higher densities to reduce the acquisition time, such as Deep-ULM, could potentially bridge this practical gap.

While the present method is implemented for 2D imaging, the ability to perform 3D ULM in a fast and data-efficient manner would be a cornerstone for many of its purposes. Operating in a low-concentration regime, traditional ULM would require acquisition, transfer and storage of an enormous amount of volumes, precluding its current use [39]. On the other hand, Deep-ULM efficiently deals with higher-concentrations, significantly lowering the required amount of acquisitions. Its future translation into 3D might therefore actually be possible.

As opposed to non-learned signal processing techniques, end-to-end deep learning methods such as Deep-ULM strongly rely on high-quality training data. We here generated new training data for each imaging system to allow Deep-ULM to exploit and learn scanner-dependent wave interference patterns. Application of Deep-ULM to new systems and experiments with different system settings therefore requires retraining.

In this work, Deep-ULM was trained to localize individual microbubbles, without incorporating any structural priors on the vascular architecture or microbubble dynamics. Including such priors in the model has the potential to further improve image fidelity, and is part of future work. We note that care should be taken to ensure that such models generalize well to pathological conditions by choosing appropriate priors, or exploiting training data that also represents these diseased cases.

Deep-ULM enables high-fidelity super-resolution vascular ultrasound imaging under challenging conditions. It operates at a high recovery speed and does not require manual tweaking by an expert user, opening vast new possibilities for localization microscopy in ultrasound imaging.

Acknowledgment

The authors would like to thank the NVIDIA Corporation and its academic GPU program for donating a Titan X Pascal which greatly facilitated the research described in this work.

References

[1] D. Cosgrove, "Angiogenesis imaging–ultrasound," *Brit. J. Radiol.*, vol. 76, no. 1, pp. S43–S49, 2003.

[2] B. B. Goldberg, J.-B. Liu, and F. Forsberg, "Ultrasound contrast agents: A review," *Ultrasound Med. Biol.*, vol. 20, no. 4, pp. 319–333, 1994.

[3] M. Siepmann, G. Schmitz, J. Bzyl, M. Palmowski, and F. Kiessling, "Imaging tumor vascularity by tracing single microbubbles," in *Proc. IEEE Int. Ultrason. Symp.*, Oct. 2011, pp. 1906–1909.

[4] O. Couture, B. Besson, G. Montaldo, M. Fink, and M. Tanter, "Microbubble ultrasound super-localization imaging (MUSLI)," in *Proc. IEEE Int. Ultrason. Symp.*, Oct. 2011, pp. 1285–1287.

[5] O. M. Viessmann, R. J. Eckersley, K. Christensen-Jeffries, M. X. Tang, and C. Dunsby, "Acoustic super-resolution with ultrasound and microbubbles," *Phys. Med. Biol.*, vol. 58, no. 18, p. 6447, Sep. 2013.

[6] M. A. O'Reilly and K. Hynynen, "A super-resolution ultrasound method for brain vascular mapping," *Med. Phys.*, vol. 40, no. 11, Oct. 2013, Art. no. 110701.

[7] Y. Desailly, O. Couture, M. Fink, and M. Tanter, "Sono-activated ultrasound localization microscopy," *Appl. Phys. Lett.*, vol. 103, no. 17, Oct. 2013, Art. no. 174107.

[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, p. 499, 2015.

[10] O. Solomon, R. J. G. van Sloun, H. Wijkstra, M. Mischi, and Y. C. Eldar, "Exploiting flow dynamics for super-resolution in contrast-enhanced ultrasound," 2018, *arXiv:1804.03134*. [Online]. Available: http://arxiv.org/abs/1804.03134

[11] J. Foiret, H. Zhang, T. Ilovitsh, L. Mahakian, S. Tam, and K. W. Ferrara, "Ultrasound localization microscopy to image and assess microvasculature in a rat kidney," *Sci. Rep.*, vol. 7, no. 1, p. 13662, Dec. 2017.

[12] D. Ackermann and G. Schmitz, "Detection and tracking of multiple microbubbles in ultrasound B-Mode images," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 63, no. 1, pp. 72–82, Jan. 2016.

[13] T. Opacic *et al.*, "Motion model ultrasound localization microscopy for preclinical and clinical multiparametric tumor characterization," *Nature Commun.*, vol. 9, no. 1, p. 1527, Dec. 2018.

[14] V. Hingot, C. Errico, M. Tanter, and O. Couture, "Subwavelength motion-correction for ultrafast ultrasound localization microscopy," *Ultrasonics*, vol. 77, pp. 17–21, May 2017.

[15] 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.

[16] 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.

[17] J. Yu, L. Lavery, and K. Kim, "Super-resolution ultrasound imaging method for microvasculature *in vivo* with a high temporal accuracy," *Sci. Rep.*, vol. 8, no. 1, p. 13918, Dec. 2018.

[18] F. Foroozan, M. A. O'Reilly, and K. Hynynen, "Microbubble localization for three-dimensional superresolution ultrasound imaging using curve fitting and deconvolution methods," *IEEE Trans. Biomed. Eng.*, vol. 65, no. 12, pp. 2692–2703, Dec. 2018.

[19] R. J. G. van Sloun, O. Solomon, Y. C. Eldar, H. Wijkstra, and M. Mischi, "Sparsity-driven super-resolution in clinical contrast-enhanced ultrasound," in *Proc. IEEE Int. Ultrason. Symp. (IUS)*, Sep. 2017, pp. 1–4.

[20] A. Bar-Zion, O. Solomon, C. Tremblay-Darveau, D. Adam, and Y. C. Eldar, "SUSHI: Sparsity-based ultrasound super-resolution hemodynamic imaging," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 65, no. 12, pp. 2365–2380, Dec. 2018.

[21] A. Bar-Zion, C. Tremblay-Darveau, O. Solomon, D. Adam, and Y. C. Eldar, "Fast vascular ultrasound imaging with enhanced spatial resolution and background rejection," *IEEE Trans. Med. Imag.*, vol. 36, no. 1, pp. 169–180, Jan. 2017.

[22] R. J. G. van Sloun, O. Solomon, M. Bruce, Z. Z. Khaing, Y. C. Eldar, and M. Mischi, "Deep learning for super-resolution vascular ultrasound imaging," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2019, pp. 1055–1059.

[23] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, p. 436, 2015.

[24] E. Nehme, L. E. Weiss, T. Michaeli, and Y. Shechtman, "Deep-STORM: Super-resolution single-molecule microscopy by deep learning," *Optica*, vol. 5, no. 4, pp. 458–464, 2018.

[25] 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.* Munich, Germany: Springer, 2015, pp. 234–241.

[26] B. Xu, N. Wang, T. Chen, and M. Li, "Empirical evaluation of rectified activations in convolutional network," 2015, *arXiv:1505.00853*. [Online]. Available: http://arxiv.org/abs/1505.00853

[27] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," 2015, *arXiv:1502.03167*. [Online]. Available: http://arxiv.org/abs/1502.03167

[28] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 3431–3440.

[29] Z. Z. Khaing *et al.*, "Contrast-enhanced ultrasound to visualize hemodynamic changes after rodent spinal cord injury," *J. Neurosurg., Spine*, vol. 29, no. 3, pp. 306–313, Sep. 2018.

[30] C. Tremblay-Darveau, R. Williams, L. Milot, M. Bruce, and P. N. Burns, "Visualizing the tumor microvasculature with a nonlinear plane-wave Doppler imaging scheme based on amplitude modulation," *IEEE Trans. Med. Imag.*, vol. 35, no. 2, pp. 699–709, Feb. 2016.

[31] L. M. I. Inc. *Definity United States Federal Drug Administration Package Insert*. Accessed: Nov. 1, 2019. [Online]. Available: https://www.accessdata.fda.gov/drugsatfda_docs/label/2011/021064s011lbl.pdf

[32] 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, 2014.

[33] Y. Cao *et al.*, "Three-dimensional imaging of microvasculature in the rat spinal cord following injury," *Sci. Rep.*, vol. 5, no. 1, p. 12643, Oct. 2015.

[34] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017.

[35] H. Chen *et al.*, "Low-dose CT with a residual encoder-decoder convolutional neural network," *IEEE Trans. Med. Imag.*, vol. 36, no. 12, pp. 2524–2535, Dec. 2017.

[36] Y. Cheng, D. Wang, P. Zhou, and T. Zhang, "A survey of model compression and acceleration for deep neural networks," 2017, *arXiv:1710.09282*. [Online]. Available: http://arxiv.org/abs/1710.09282

[37] G. Hinton, O. Vinyals, and J. Dean, "Distilling the knowledge in a neural network," 2015, *arXiv:1503.02531*. [Online]. Available: http://arxiv.org/abs/1503.02531

[38] C. T. Chin, "Modelling the behaviour of microbubble contrast agents for diagnostic ultrasound," Ph.D. dissertation, Univ. Toronto, Toronto, ON, Canada, 2001.

[39] O. Couture, "Super-resolution imaging with ultrafast ultrasound localization microscopy (uULM)," in *Proc. Eur. Symp. Ultrasound Contrast Imag.*, Rotterdam, The Netherlands, 2017, pp. 155–156.