

ULTRA-SR Challenge: Assessment of Ultrasound Localization and TRacking Algorithms for Super-Resolution Imaging

Marcelo Lerendegui¹, Kai Riemer², Georgios Papageorgiou³, *Member, IEEE*, Bingxue Wang⁴, Lachlan Arthur⁵, Arthur Chavignon⁶, Tao Zhang, Olivier Couture⁷, *Member, IEEE*, Pingtong Huang, Md Ashikuzzaman, *Member, IEEE*, Stefanie Dencks⁸, *Member, IEEE*, Chris Dunsby, Brandon Helfield⁹, *Member, IEEE*, Jørgen Arendt Jensen¹⁰, *Fellow, IEEE*, Thomas Lisson¹¹, *Student Member, IEEE*, Matthew R. Lowerison¹², Hassan Rivaz¹³, *Senior Member, IEEE*, Anthony E. Samir, Georg Schmitz¹⁴, *Senior Member, IEEE*, Scott Schoen Jr.¹⁵, Ruud van Sloun¹⁶, *Member, IEEE*, Pengfei Song¹⁷, *Senior Member, IEEE*, Tristan Stevens¹⁸, *Graduate Student Member, IEEE*, Jipeng Yan¹⁹, Vassilis Sboros²⁰, and Meng-Xing Tang²¹, *Senior Member, IEEE*

Manuscript received 16 January 2024; revised 7 March 2024; accepted 27 March 2024. Date of publication 12 April 2024; date of current version 1 August 2024. The work of Marcelo Lerendegui, Kai Riemer, Bingxue Wang, Chris Dunsby, Jipeng Yan, and Meng-Xing Tang was supported in part by the U.K. Engineering and Physical Sciences Research Council (EPSRC) under Grant EP/T008970/1 and Grant EP/V04799X/1, in part by the Chan Zuckerberg Foundation under Grant 2020-225443, in part by the EPSRC Impact Acceleration Account Funding under Grant EP/R511547/1, and in part by the Medical Research Council (MRC) Confidence in Concept Scheme at Imperial College under Grant MC-PC-18050. The work of Georgios Papageorgiou and Vassilis Sboros was supported in part by the Chief Scientist Office Scotland under Grant TCS/18/40, in part by the Medical Research Council Confidence in Concept under Grant MRC/CIC3/027, and in part by the Science and Technology Facilities Council Project under Grant ST/M007804/1. The work of Lachlan Arthur was supported by the British Heart Foundation under Grant FS/PhD/21/29113. The work of Arthur Chavignon and Olivier Couture was supported by the European Research Council (ERC) under the European Union Horizon H2020 Program/ERC Consolidator under Grant 772786-ResolveStroke. The work of Md Ashikuzzaman, Brandon Helfield, and Hassan Rivaz was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC). The work of Stefanie Dencks, Thomas Lisson, and Georg Schmitz was supported by Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Grant 233312120 and Grant 445703531. The work of Brandon Helfield was supported in part by the Canada Research Chair Program and in part by Burroughs Wellcome Fund under Grant 1018212.03. The work of Jørgen Arendt Jensen was supported by the European Research Council (ERC), Synergy, through SUPER Resolution ultrasound imaging using Erythrocytes (SURE), under Grant 854796. The work of Matthew R. Lowerison was supported in part by the National Institute of Biomedical Imaging and Bioengineering of the National Institutes of Health under Award R21EB030072 and in part by the Beckman Institute Postdoctoral Fellowship. The work of Scott Schoen Jr., was supported by the FV Hunt Postdoctoral Research Fellowship in Acoustics from the Acoustical Society of America. The work of Pengfei Song was supported by the National Institute of Biomedical Imaging and Bioengineering of the National Institutes of Health under Award R21EB030072. (Marcelo Lerendegui and Kai Riemer contributed equally to this work.) (Corresponding author: Meng-Xing Tang.)

This work involved human subjects or animals in its research. Approval of animal experiments were approved by the Institutional Committee C2EA-54 "Comité d'Éthique Normandie en Matière d'Expérimentation 628 Animale" under Protocol No. APAFIS 22544-2019093017523797 and the Human Lymph Node Acquisition was approved by the Zhejiang University Institutional Ethics Board.

This article has supplementary downloadable material available at <https://doi.org/10.1109/TMI.2024.3388048>, provided by the authors.

Please see the Acknowledgment section of this article for the author affiliations.

Digital Object Identifier 10.1109/TMI.2024.3388048

Abstract—With the widespread interest and uptake of super-resolution ultrasound (SRUS) through localization and tracking of microbubbles, also known as ultrasound localization microscopy (ULM), many localization and tracking algorithms have been developed. ULM can image many centimeters into tissue *in-vivo* and track microvascular flow non-invasively with sub-diffraction resolution. In a significant community effort, we organized a challenge, Ultrasound Localization and TRacking Algorithms for Super-Resolution (ULTRA-SR). The aims of this paper are threefold: to describe the challenge organization, data generation, and winning algorithms; to present the metrics and methods for evaluating challenge entrants; and to report results and findings of the evaluation. Realistic ultrasound datasets containing microvascular flow for different clinical ultrasound frequencies were simulated, using vascular flow physics, acoustic field simulation and nonlinear bubble dynamics simulation. Based on these datasets, 38 submissions from 24 research groups were evaluated against ground truth using an evaluation framework with six metrics, three for localization and three for tracking. *In-vivo* mouse brain and human lymph node data were also provided, and performance assessed by an expert panel. Winning algorithms are described and discussed. The publicly available data with ground truth and the defined metrics for both localization and tracking present a valuable resource for researchers to benchmark algorithms and software, identify optimized methods/software for their data, and provide insight into the current limits of the field. In conclusion, Ultra-SR challenge has provided benchmarking data and tools as well as direct comparison and insights for a number of the state-of-the-art localization and tracking algorithms.

Index Terms—Super-resolution, SRUS, ULM, CEUS, ultrasound, imaging, microbubble, localization, tracking, challenge.

I. INTRODUCTION

ULTRASOUND is a frontline imaging modality that is extensively used in the clinic, characterized by its excellent accessibility and affordability, high temporal resolution, patient friendliness, and free from ionizing radiation. The spatial resolution of clinical ultrasound, typically between hundreds of microns to millimeters, has been limited by a classical compromise between resolution and penetration

depth. On the other hand, the functioning of healthy tissue is regulated by the microvasculature and changes in its structure and flow dynamics are important diagnostic indicators, which are largely unavailable through existing modalities [1]. To be able to non-invasively image microvasculature *in-vivo* with both large penetration depth and microscopic spatial resolution can be valuable.

Microbubble (MB) contrast agents, micron sized bubbles typically with a lipid shell and inert gas core, have been widely used clinically for imaging blood flow and perfusion [2], [3], [4]. The advent of super-resolution ultrasound (SRUS) through the localization and tracking of MBs, also known as ultrasound localization microscopy (ULM) [5], [6], has circumvented the classical diffraction resolution limit and shown to visualize and quantify micro-vasculature in unprecedented resolution [7], [8]. ULM typically involves a number of steps. Firstly, an ultrasound image sequence of tissue containing flowing MBs is acquired and tissue motion corrected. Secondly, the tissue background is removed through either contrast specific sequences such as pulse inversion or amplitude modulation or spatial temporal filters such as singular value decomposition (SVD). Thirdly, signals from spatially isolated MBs on each image frame are identified and bubble positions localized, and such localizations are accumulated over a large number of imaging frames over time to form the ULM localization density image. These bubbles are then tracked over frames to generate flow velocity and direction maps. Besides its sub-diffraction resolution in vascular geometry, another notable feature of ULM is its ability to generate super-resolved flow velocity information within these micro-vessels by detecting and tracking these MBs, offering non-invasive assessment of tissue blood flow at microscopic level. ULM was first demonstrated *in-vivo* in a tumor model [9], a mouse ear [10], and a rat brain [11], [12], followed by a multitude of examples of 2D and 3D SR preclinical imaging of tumors [13], [14], the renal vasculature [15], [16], [17], lymph nodes [18], the brain functional activity [19], small animal heart [20]. ULM was also demonstrated in humans in the breast [14], [21], lower limb [22], brain [23], liver [24], lymph nodes [25] and myocardium [26].

Localization and tracking algorithms are central to ULM and have a direct impact on its spatial and temporal resolution. A significant present challenge for ULM is its low temporal resolution as long acquisition time is required to accumulate sufficient spatially isolated bubble locations over time. An algorithm which can isolate, localize and track MBs accurately in the presence of high bubble concentrations and noise would reduce the time to accumulate sufficient number of localizations and tracks required. Efforts have been made to reduce the time of an ULM acquisition for example by tracking MBs at higher densities through incorporating a motion model via a Kalman filter [27], [28], deconvolution and multi-feature tracking [29], by splitting the bubble signal in the 3D Fourier domain [30], or through the use of phase change contrast agents [31]. Besides the classic approaches, sparsity-based methods [32] and recent developments in deep learning [33], [34], [35] have challenged the existing localization and tracking methods. These current and continuous technological

advances to improve ULM and reduce acquisition time are essential for its clinical translation.

Despite the widespread study of ULM and the large number of ultrasound localization and tracking algorithms developed in the past decade, their performance has not yet been consistently evaluated and benchmarked. A recent study [36] has provided benchmarking tools and dataset for evaluating performance of seven localization methods. However, the assessment of tracking methods was not considered there. Moreover, the datasets with ground truth were generated without considering MB oscillation physics. Tracking is commonly used to generate the final ULM images, and thus motion models (e.g. implemented using the Kalman filter) may have significant influence on these images. Such effects underscore the need for metrics to assess the performance of such tracking algorithms. Furthermore, different tracking and localization combinations may generate different ULM results. Therefore, evaluation of both localization and tracking individually and in combination, together with a more realistic imaging simulator, is valuable for future development and benchmarking of ULM algorithms.

To rigorously evaluate the performance of state-of-art algorithms for both localization and tracking and provide a benchmark, we ran the first open challenge to evaluate Ultrasound Localization and TRacking Algorithms for Super Resolution [ULTRA-SR, part of the IEEE International Ultrasonics Symposium (IUS) 2022 in Venice, Italy]. We (a) generated simulation datasets of different ultrasound frequencies by combining vascular flow simulation and acoustic field simulation including the modeling of the MB dynamics, (b) proposed a simple set of performance evaluation metrics, and (c) assessed the performance of localization and tracking for different algorithms developed by the challenge participants. Additionally, the algorithms on two *in-vivo* data sets were assessed.

II. METHODS

A. Challenge Design

For the ULTRA-SR challenge, a broad committee (whose membership can be found at www.ultra-sr.com) was established comprising members of the SRUS community with expertise in both experimental data acquisition and algorithm and software development, to advise the challenge design, including the scope of the challenge, the datasets, and the evaluation metrics and process. The consensus was reached on the datasets, the evaluation metrics and processes as described in this section including:

- Low (2.84 MHz) and high (7.24 MHz) frequency ultrasound simulation datasets with randomly generated vascular trees and flowing MBs that satisfy the Poiseuille's law (with additional shorter simulation datasets for initial test / adjustment of algorithms).
- A simple set of localization and tracking evaluation metrics for simulation data.
- *In-vivo* high frequency low concentration dataset of a rat brain and high concentration dataset of a human cervical lymph node.

Participants' algorithms were assessed in three challenge categories (participation in each of the categories was optional):

- Localization only with simulation data.
- Localization and tracking with simulation data.
- Localization and tracking with *in-vivo* data.

For the first two categories using simulation data with ground truth, the evaluation was automated according to the proposed evaluation metrics. Expert panel evaluation of the results was performed for the *in-vivo* datasets. The details of evaluation of the challenge are described in section G.

The challenge website was published with a call in January 2022. 192 researchers from all over the world registered their interest in the challenge. Following a consensus for data generation and evaluation, in April 2022 two short test simulation datasets with ground truth locations and bubble parings were provided. Participants were given feedback on their performance to minimize any systematic offsets in their localization. The final datasets were released in July 2022. The datasets were then assessed based on three challenge categories, and participation in each of the categories was optional. A total of 38 submissions from 24 groups worldwide in all categories were submitted for assessment. The evaluation of the submitted results was managed and conducted anonymously by the Heriot-Watt University group who did not participate in the challenge. The finalists presented their work at the special ULTRA-SR Challenge session of the IEEE IUS2022, and the winners and runners-up for each of the three awards were announced during the closing ceremony of the symposium. The accepted challenge proceeding papers can be accessed from [this link](#), the winning team's results can be found on the [challenge's website](#), section ii.C and ii.D summarizes their techniques.

B. Winning Teams

See Table I.

C. Simulation Data

Evaluation of super-resolution software on experimental data lacks the ground-truth required for rigorous quantification of software performance. Therefore, realistic simulated datasets are required. A challenge in simulating ULM data was the accurate modelling of ultrasound signals from MBs from a polydisperse population flowing in complex 3D microvascular networks. This was addressed through the use of a recently developed simulation framework, BUbble Flow Field (BUFF) [37], which incorporates 1) a 3D microvascular tree generator; 2) a Hagen–Poiseuille vascular flow model; 3) an acoustic field simulator (Field II [38], [39], [40]); and 4) a MB dynamics simulator coupled to the field simulator.

Simulation data sets, as illustrated in Fig. 7, were generated using the parameters of two clinical transducers, one cantered at 7.2 MHz for peripheral imaging, and one cantered at 2.4 MHz for cardiac imaging. Four randomly generated networks of MB seeded vessel trees (an example shown in supplementary video 1) were simulated using BUFF. The microvasculature was 3D and the simulation produced 2D

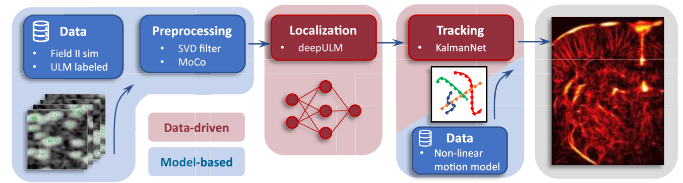


Fig. 1. Data processing diagram by EID12 TU/e group.

images from the 3D structures. This allowed generation of out of plane bubbles and motions in the 2D dataset. Through the designs of highly complex 3D vascular flow networks, the simulation image sequences had significant spatial variations and encompassed a wide range of flow velocities and bubble densities. The average numbers of bubbles per frame were 550 and 155, and the average distances between neighboring bubbles within each frame were 0.39 (1.9 wavelengths) ± 0.27 mm and 1.65 (3.1 wavelength) ± 1 mm, for the high and low frequency simulation data sets respectively. To minimize the predictability of the path of a MB the vessel tree generation parameters, e.g., bifurcation probability, angle between consecutive segments or MB seeding probability, were randomized. No tissue motion was included in the simulation data sets because the challenge was focused on localization and tracking. Gaussian white noise was generated, filtered using the transducer bandwidth, and beamformed. Time gain compensation was applied to the noise. In this way, the signal to noise ratio varied with the image depth. Table II and Table III contain the simulation parameters. The final challenge data has 500 frames for both low and high frequency data.

A test simulation dataset of only 20 frames with ground truth was generated using a simple microvascular network with a low MB concentration. The main purpose of this dataset was for participants to ensure their codes would work with the data files and would also optimize/train their algorithm, given the known ground truth.

The data released with the ULTRA-SR challenge is available to the public and aims to be a reference benchmark for future SR algorithm development for and in collaboration with the research community. Please visit the website (www.ultra-sr.com) for more detailed information.

D. Localization and Tracking Algorithms

1) **EID12 | Eindhoven University of Technology (TU/E):** A diagram of data processing is shown in Fig. 1.

a) **Localization: DeepULM:** For micro-bubble localization, we employed the DeepULM architecture presented in [33] and directly upsampled the input B-mode images to the final super-resolution grid. Training of the model was facilitated through a mean-squared error (MSE) between the predicted locations and the ground truth targets on input patches of 64×64 pixels. Furthermore, we applied a small Gaussian blur to the target labels to relax the MSE objective and favor detection accuracy over localization error. As a result, we could train the localization network in a regression style, such that detections that are off target, but in close vicinity, still contribute to the loss. The variance of the blurring kernel can be annealed over training time to regain localization accuracy. Finally,

TABLE I
WINNING PARTICIPANTS INFORMATION

ID	Affiliation	PI	Localization Method	Tracking Method	Localization Result	Overall Result	In-Vivo Result
EID1	Ruhr University Bochum (RUB)	Georg Schmitz	Gaussian fit	Markov Chain Monte Carlo Data Association	Winner	Runner-Up	
EID5	Harvard Medical School (HMS) Massachusetts General Hospital (MGH)	Anthony Samir	Morphological Reconstruction	Hungarian algorithm		Runner-Up	
EID6	University of Illinois Urbana-Champaign (UIUC)	Pengfei Song	Normalized cross-correlation	Kalman-filter & bipartite-graph-based pairing			Winner
EID12	Eindhoven University of Technology (TU/e)	Ruud van Sloun	DeepULM	KalmanNet			Runner-Up
EID16	Concordia University	Hassan Rivaz	NCC-based localization	Bubble-set registration	Winner		
EID22	Imperial College	Meng-Xing Tang	Sparsity-based deconvolution	Feature-motion-model framework		Winner	Runner-Up

TABLE II
SIMULATION PARAMETERS

Parameter	Low Freq	High Freq
Vessel branching probability from root to branch levels	0.05, 0.15, 0.2, 0.3, 0.4	
Angle range of a bifurcation	[-30°, 30°] in XZ, [-3°, 3°] in YZ	
Angle range of a single vessel	[-30°, 30°] in XZ, 0° in YZ	
Vessel step size	1mm	
Bubble flow velocities [min, mode, max]	[0.008, 5, 15 cm/s]	[0.002, 5, 14 cm/s]
Ambient pressure	101k Pa	
Liquid density	997 kg/m ³	
Liquid surface tension	0.073 N/m	
Liquid viscosity	0.001 Pa*s	
MB response sampling freq	1 GHz	
Transducer	GEM5ScD	L11-4
Transmitted Frequency	2.84 MHz	7.24 MHz
Sampling Frequency	80.0 MHz	
Framerate	50 Hz	
Image number of pixels	[512, 512]	
Image dynamic range	40 dB	
Plane wave angles	[-10°, 0°, 10°]	
Mechanical Index	0.1	
Element apodization	Tukeywin(0.25)	
Signal to noise ratio	23 dB	19 dB

different neural architectures and corresponding hyperparameters were explored through random sweeps. For inference, the best performing models were chosen and predictions were ensembled for optimal performance, by removing any outliers through majority voting. Peak detection on the output maps of the localization network was applied to obtain the final classifications of the bubbles, and the threshold value was chosen to maximize accuracy. For reference, inference on the rat brain dataset using a single 12GB NVIDIA GeForce RTX 3080 Ti took approximately 10 minutes.

We trained individual models on simulated datasets that match each challenge dataset to maximize performance. For

TABLE III
IN-VIVO RAT BRAIN DATA ACQUISITION PARAMETERS

Frequency	15.6 MHz
Frame rate	100 Hz
Continuous block size	80
Nb of block	100
Nb cycles	2
Image size (h x w)	166 x 126 pixels
Pixel size	0.050 x 0.049 mm
Compression	No, linear value

each of the challenge datasets, we generated training data using Field II ultrasound simulations. Subsequently, we empirically chose the acquisition and reconstruction parameters of the simulations that best match the PSF shapes in the challenge datasets. Furthermore, for the *in-vivo* datasets, we sampled noise from the challenge dataset itself, using patches containing exclusively noise and clutter, and randomly superimposed it onto the simulated point scatters to incorporate realistic noise artifacts into our training datasets. Since only post-envelope detected data was available, these could not be coherently added. Therefore, a (soft) masked addition of the point targets and clutter images was done for smooth transitions between signals and noise.

b) Tracking: KalmanNet: Our proposed tracking framework [41] for micro-bubble tracking combined the advanced motion estimation of KalmanNet [42] with the conventional data association of the Hungarian Algorithm (HA) [43]. KalmanNet was used for the motion estimation of the MBs and predicts the positions of the MBs in the next frame. Subsequently, HA was used to solve the linear sum assignment problem with a pairwise distance matrix as cost. The process of KalmanNet was similar to classic Kalman filtering; it also had prediction and update steps to estimate states. However, the Kalman gain computation was learned from the data through a recurrent neural network (RNN). Rather than relying upon faulty assumptions of the underlying state

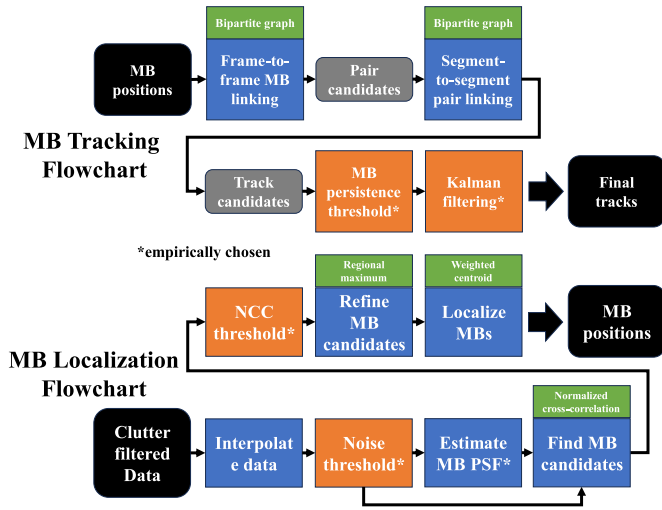


Fig. 2. Data processing diagram by the EID6 UIUC group. MB localization was performed using normalized cross-correlation with empirically selected thresholds. The MB candidate positions were then input into a Kalman-filter based bipartite-graph-based pairing algorithm to produce ULM tracks, which were then accumulated into ULM vasculature and velocity maps.

space model, learning the Kalman gain allowed us to use the powerful Kalman filter in a setting where we have little domain knowledge, particularly in the case of the highly non-linear paths of MBs.

The KalmanNet-based tracker consists of three steps that were executed for every frame, namely prediction, data association, and update. In the prediction step, KalmanNet estimated the next MB position on its corresponding track. Consequently, in the data association step a cost matrix, containing the Euclidean distances from predicted MB positions to localized MB positions, was constructed and used by the HA to solve the assignment problem. After applying the HA, the assignments were checked against a distance threshold to avoid physical implausible displacements. Lastly, in the update step, the assigned observations were used to update the state and Kalman gain of the corresponding tracks.

We trained KalmanNet on a dataset containing tracks that were generated with a nonlinear motion model (NMM) in state space format, as in [44]. The states included the lateral and axial coordinates, as well as the velocity, angle, and turning rate of MBs. By including the latter parameter, the motion model considered the curved physiology of vessels. In total, 25000 tracks of length 100 frames were generated at a frame rate of 100 Hz with a maximum flow velocity of 15 mm/s.

2) EID6 | University of Illinois Urbana-Champaign (UIUC): A diagram of data processing is shown in Fig. 2. All processing by UIUC was performed with MATLAB R2019a (MathWorks, Natick, Massachusetts, USA) on a Windows 10 workstation with 128 GB DDR4 2666 MHz RAM, an Intel Xeon Gold 2.00 GHz 6138 20 core CPU (Intel, Santa Clara, California, USA), and an NVIDIA GeForce RTX 2080 Ti GPU (NVIDIA, Santa Clara, California, USA).

a) Localization: normalized cross-correlation: For MB localization of the *in-vivo* challenge data, we implemented the ULM processing pipeline and methodologies presented in [30] and [45], [46]. Briefly, raw image data was spatially

interpolated to the desired resolution (10x interpolation in the axial and lateral directions) using a uniform 2-D spline upsampling. An intensity threshold was then applied to the interpolated data to remove noisy background and other artifacts (residual clutter, etc.). A point-spread function (PSF) specific to the challenge data was estimated by manually selecting isolated MBs in interpolated imaging frames and fitting a multivariate Gaussian function to the extracted signals. Following this, a 2-D normalized cross-correlation was performed between each frame of the interpolated image data and the MB PSF to generate a cross-correlation map of estimated MB locations. To further suppress unwanted signals and background, an empirically chosen threshold was applied to the cross-correlation map to remove pixels with a correlation coefficient less than 0.6. Rough candidate MB localizations were then estimated using a regional maximum search of the cross-correlation peaks in each frame. These candidate localizations were then input into a weighted centroid algorithm to further refine the localization estimates. This localization procedure took approximately 8 minutes to process the entire *in-vivo* rat brain dataset.

b) Tracking: kalman-filter bipartite-graph-based tracking: MB trajectories were estimated using a bipartite graph-based pairing and tracking algorithm to minimize the frame-to-frame MB centroid and track-end distances [45]. A minimum MB persistence of 5 frames was applied to remove low confidence tracks. Kalman filtering was then applied to improve MB speed estimation and to in-paint trajectories to the desired interpolated pixel grid [46]. The final ULM image was the temporal accumulation of all MB trajectories within each data acquisition. The tracking and accumulation processing took approximately 6 minutes for the entire *in-vivo* rat brain dataset.

3) EID1 | Ruhr University Bochum (RUB): A diagram of data processing is shown in Fig. 3.

a) Localization: gaussian fit: Depending on the overlap and shape of the MB's PSF and the background clutter, different localization methods were chosen. In general, Gaussian shaped PSF prototypes were fit to the data either by matched filtering or nonlinear Gaussian fitting. For simulation 1 (low frequency) and the *in-vivo* rat brain dataset, matched filters with a fixed Gaussian filter kernel were applied [21], [27] because of overlapping PSFs and strong background clutter. For this, the standard deviations of the Gaussians were determined by evaluating the PSF size of single MBs in different image regions. The positions of regional maxima of the matched filter output are the localization results. To improve precision the matched filters were applied to data interpolated to a 5 μm grid. For this data set, the matched filter technique was the most robust localization method for suppressing PSF artifacts. The localization threshold was chosen to aim for highest precision by avoiding multiple detections of single MBs but risking more missing localizations.

Processing was performed on a Ryzen 7 5700G processor and 32 GB RAM. Because processing time was not a key factor for evaluation in this challenge, parameter selection for localization and tracking was carried out independently of the computational complexity and run time since the highest result accuracy was our primary aim. For the low frequency

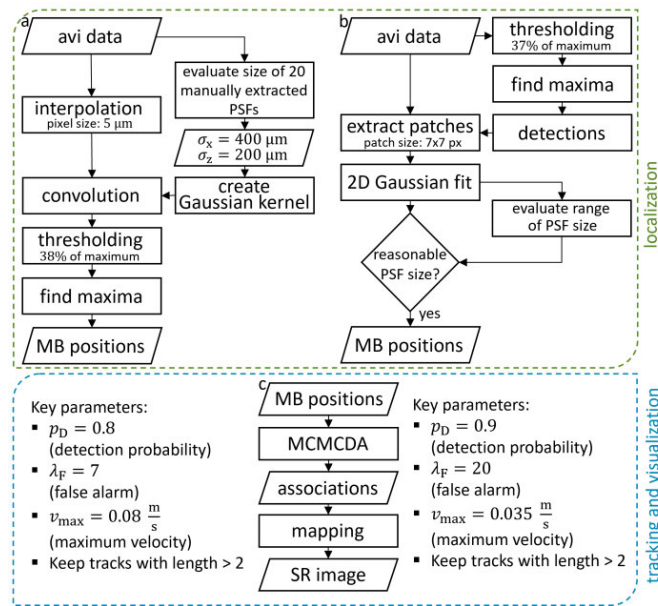


Fig. 3. Flowcharts of SR algorithm used for the synthetic datasets by EID1 RUB group. In subfigure a, the localization algorithm used for the low frequency (net_002), and in b, used for the high frequency dataset (net_003) are shown. The flowcharts only differ in the localization step. In subfigure c, the tracking algorithm is shown. Here, only the parameters were adapted to each dataset.

data set, the interpolation required for the matched filter was set to a relatively low pixel size. With an unoptimized MATLAB implementation this resulted in a parallelized processing time of 3728 s. Considering the localization mean error (see results), the interpolated pixel size could be increased, resulting in a drastically lower processing time.

For simulation 2 (high frequency) and the *in-vivo* lymph node dataset, the MBs were well separated and therefore localized using Gaussian least-squares fits with more degrees of freedom to adapt to the PSF shape. For this, first a thresholding and the detection of regional maxima had to be performed. Then, the Gaussians were fitted to patches centered around these maxima, large enough to contain the PSF. Fitting Gaussian functions, continuous MB positions are provided, so localization precision is not limited by the pixel grid. To ensure reliability, only localizations with an estimated standard deviation between 0 and two times the median standard deviation of all fits were accepted. For this data set, we aimed for a maximum detection rate at which the risk of false alarms is greater. The iterative Gaussian fit for the high frequency data set was computed on one core with an unoptimized MATLAB implementation, resulting in a processing time of 686 s.

b) Tracking: markov chain monte carlo data association: The MBs were tracked using mULM (motion model ultrasound localization microscopy) which is based on a Markov Chain Monte Carlo Data Association (MCMCDA) algorithm [14], [21], [27]. With this algorithm the detected positions are associated to tracks by a probabilistic optimization of the a-posteriori probability of these associations. In this, the track probabilities are evaluated by a Kalman filter with a linear motion model. Because of this motion model and assumptions about, e.g., the detection probability and localization precision,

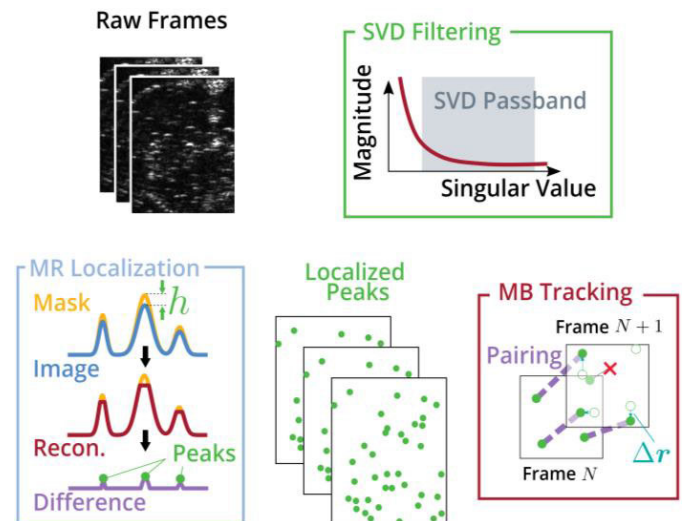


Fig. 4. Flowchart of localization and tracking method by the EID5 HMS/MGH group (after [47]). Raw US frames were passed through an SVD filter to remove static background signal. Morphological reconstruction was performed with empirically determined offset h , and peaks were taken to be centroids of regions in difference image. These localized peaks were then tracked between frames with the Hungarian algorithm, and their velocity defined as $v = \Delta r / \Delta t$, where Δr is the change in position between frames acquired at times separated by Δt .

this algorithm is robust even in case of high MB concentrations, complex vascular patterns, and low frame rates. For each dataset, the required parameters (e.g., maximum flow velocity) were adapted to the observed dataset properties. To improve reliability, only tracks containing at least 3 localizations were accepted.

Tracking was performed on a single core with C++ and the number of iterations was selected to be very high. This resulted in a processing time for the low frequency data set of 1278 s, and due to the high number of MBs in the high frequency data set, tracking needed 20313 s. Tracking is based on a Monte Carlo algorithm that in each iteration randomly performs changes to the track associations and accepts them dependent on a probability calculated using a linear motion model. Retrospective computations with one tenth of the iterations selected for the challenge indicated that the additional iterations do not have a major influence on the track assignments and computation time could be reduced by 90 %.

4) EID5 | Harvard Medical School (HMS) Massachusetts General Hospital (MGH): A diagram of data processing is shown in Fig. 4.

a) Localization: morphological reconstruction: MB localization was performed using an augmented version of the processing steps presented in [47] and [48]: First, the raw images were passed through a singular value decomposition filter [49] to remove static signals (largest singular values) and clutter (smallest). To further enhance bubble signal, the frames were then deconvolved with a position independent estimated PSF (Gaussian distribution with standard deviations of 0.5 and 1 wavelengths in the axial and lateral dimensions, respectively). Finally, grayscale morphological reconstruction [47] was performed on each normalized frame, with a constant empirically determined offset across all frames.

The morphological reconstruction process isolates local maxima with varying raw intensity values; applying an intensity threshold to the reconstructed image isolates regions with similar morphologies (i.e., those whose shape is similar to the PSF), and the centroids of each region may be designated as the MB locations. The precise efficiency of the morphological reconstruction operation is dependent on the features of each image [50], but in general is highly efficient: for the present case, localizations required less than 10 ms per frame (Intel Core i9, at 2.2 GHz 32 GB RAM) with the default MATLAB 2020b implementation.

The main parameter for the technique is the morphological offset $h \in (0, 1)$. Small values of h (e.g. $h < 0.1$) correspond to higher sensitivity and thus are better suited for high SNR datasets. Conversely, large offset values (e.g. $h > 0.5$) will detect only the strongest signals and are more appropriate for noisy data. While this value was selected manually based on trial values (a process whose duration was acceptable given the efficiency of the algorithm), results from the test datasets (for which ground truth was known) indicated that the detection rate varied slowly near the optimal value of h .

b) *Tracking: hungarian algorithm*: MB localizations in each frame were paired with those in subsequent frames with an implementation of the Hungarian algorithm (i.e., the set of pairings were chosen to minimize the distance between MBs in sequential frames) [51], [52]. To ensure separate individual MBs were not paired between frames, maximum and minimum velocities (2.5 cm/s and 0.1 cm/s, respectively) were set; pairings that yielded MB velocity magnitudes $v_{MB} = |\mathbf{r}_1 - \mathbf{r}_2|/\Delta t$ outside this range were excluded ($\mathbf{r}_i$ is the position of the MB in frame i , and Δt is the time between frames). Finally, to capture MB tracks that include a missing localization in a single frame, a gap closing threshold was set to 3 frames. The complexity of the algorithm as implemented was $O(N^3)$, where N is the number of points, and required on the order of minutes for the challenge datasets with unoptimized MATLAB implementation.

5) *EID22 | Imperial College*: A diagram of data processing is shown in Fig. 5.

a) *Localization: sparsity-based deconvolution*: Sparsity-based deconvolution, proposed in a previous work [29], was used for MB isolation. Before deconvolution, denoising was done with a median filter and a wiener filter. As SNR varies across image regions, adaptive thresholding with MATLAB ‘adaptthresh’ function was used to estimate noise levels for difference regions. The MATLAB function had two key parameters, ‘Sensitivity’ and ‘NeighborhoodSize’, which were empirically adjusted for each individual dataset. The final ‘Sensitivity’ parameter was set to 1 for both simulation datasets and ‘NeighborhoodSize’ parameter was set as 101 and 51 for the high- and low-frequency simulation data respectively. Pixel intensity below the estimated noise level in high-frequency data or below half of the estimated noise level in the low-frequency data were set to zero before localising MBs. Spatial moving average was used to smooth the edges of bubble signals after thresholding and reduce spectrum leakage that might happen during the sparsity-based deconvolution.

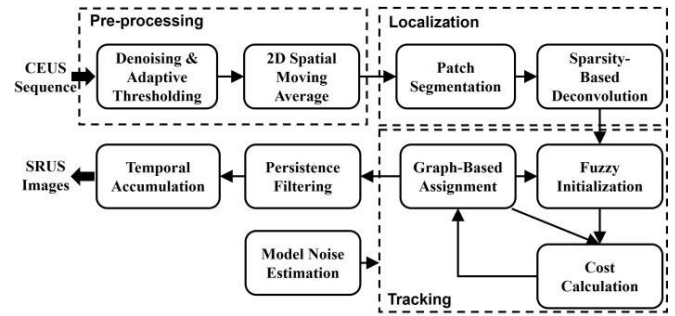


Fig. 5. Data processing diagram by EID22 Imperial Group. Noise in the CEUS sequence was reduced by pre-processing. Bubble image patches and estimated PSF were used for Sparsity-based deconvolution in order to isolate single MBs. Localized MBs were then tracked using a feature-motion-model framework. Kalman parameters were estimated from the data and velocities of new MBs were calculated by Fuzzy Initialization.

For the deconvolution, each point spread function was estimated by averaging 10 manually segmented single bubble images from the data. The image was divided into 3×3 regions and PSF for each region was estimated to take the spatial variations of PSF into account. MB patches in each frame were cropped out by pixel connectivity. Sparsity-based deconvolution was done between each MB patch and its corresponding PSF to better isolate MBs by shrinking their images. MBs were then localized by the intensity-weighted centroid of the deconvoluted image and sum of pixel intensities in each region was taken as the image intensity feature for the corresponding MB.

b) *Tracking: feature-motion-model tracking framework*: This framework was based on a previous work to track bubbles using image feature and motion model [29]. MBs in two adjacent frames were paired by finding a minimum cost by the graph-based assignment. The cost for pairing two MBs was defined as the ratio of their image intensity difference to the probability obtained from the linear Kalman motion model established with localized MB positions. A 3-frame-based fuzzy initialization was proposed to estimate velocities of newly appeared MBs, instead of assuming these MBs to be static, further improving the performance of Kalman motion model in the two-frame pairing framework [26]. The covariance of observation noise in the Kalman filtering was set as a quarter of the wavelength and the covariance of model prediction noise was automatically estimated from the data to avoid human interaction and effort for tuning these parameters [26]. Maximum flow velocity, determined by the searching window size for pairing MBs, was 80mm/s and 120 mm/s for the high- and low-frequency data respectively. Note that the high frequency data has much higher concentration than the other. Only tracks containing at least 4 localizations were accepted to improve reliability.

The localization and tracking codes were developed in MATLAB R2021a and fully CPU-based. The processing time of localization and tracking increases with MB concentrations. When the high-frequency simulation data was processed by a desktop PC (CPU: 12-Core Ryzen 9 5900X (AMD, Santa Clara, California, USA), RAM 128 Gb), the localization processing was parallelized by processing each frame with one core and took around 325 second in total; the tracking processing was not parallelized and took around 173 seconds.

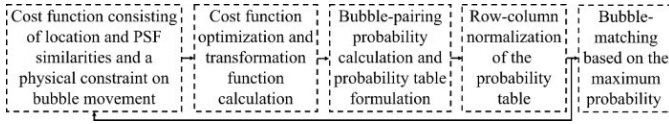


Fig. 6. Flowchart of the bubble-tracking algorithm proposed by EID16 Concordia group.

6) *EID16 | Concordia University*: A diagram of data processing is shown in Fig. 6.

a) *Localization: NCC-based localization*: A non-overlapping bubble signal is extracted from one of the contrast-enhanced ultrasound (CEUS) images to consider it as the tentative point-spread function (PSF) [53]. Then an amplitude thresholding-based clutter suppression step is performed on the CEUS images. Similarity maps are produced by calculating the normalized cross-correlation (NCC) between the tentative PSF and the clutter-suppressed CEUS image patches. The NCC values in the similarity maps are thresholded to prevent spurious localization of bubbles. The local maxima of the thresholded NCC maps ascertain the integer locations of the bubble centres. These rough estimates are refined by calculating the amplitude-weighted average of the relative grid coordinates surrounding the integer bubble centres. An unoptimized MATLAB implementation of this localization algorithm required 2.95 seconds to localize 124 bubbles in a $512 \text{ pixels} \times 512 \text{ pixels}$ CEUS image using a $9 \text{ pixels} \times 39 \text{ pixels}$ PSF as the reference on a 6th-generation Intel Core-i5 CPU, clocked at 3.50 GHz, with 32 GB RAM.

b) *Tracking: bubble-set registration algorithm*: The tracking algorithm in [53] casts the problem of MB temporal pairing as a bubble-set registration problem. Instead of binary pairing, the best matches of the bubble signals are assigned through probabilistic inference. A non-linear cost function enforcing position and point-spread function (PSF) similarities and imposing a physical constraint on the bubbles' movement is formulated. This cost function is minimized using the Gauss-Newton optimization technique to obtain the transformation function registering two bubble sets. Once the transformation function is calculated, the bubble-pairing probabilities are updated by multiplying the position and PSF similarity-based probabilities. Row-column normalization of the probability table is performed so that each row and column sum up to one. An iterative update of the transformation function and the probability table leads the algorithm to convergence, and the highest pairing probability finally matches a reference bubble to a target bubble.

The weights controlling the relative dominance between the bubble position-, PSF-, and the bubble movement-related constraints, in addition to the number of iterations, were all key parameters associated with this tracking algorithm. These parameters were empirically chosen for each dataset individually using one frame pair per dataset. The highest probability value in each row/column was observed carefully to determine the algorithm's confidence in bubble pairing and to select the optimal parameter set accordingly. In general, the bubble movement-related parameter was set to a substantially higher value than the other two parameters due to the small displacements of bubbles between two consecutive frames. In contrast, the PSF similarity weight was set to a small value owing to the non-linearity of scattered bubble signals.

The algorithm was allowed to run for a large number of iterations to facilitate convergence. An unoptimized MATLAB implementation of the proposed tracking algorithm required 74.42 seconds to track 124 bubbles with 30 iterative updates of the transformation function and the probability table and 200 iterations for normalizing the rows and columns of the probability table on a 6th-generation Intel Core-i5 CPU, clocked at 3.50 GHz, with 32 GB RAM.

E. Evaluation Metrics

The evaluation metrics for localization and tracking present estimations of both the fractions of correct events and the error. They are intended to be used with *in-silico* data with ground truth and are explained in the following.

1) *Localization Evaluation*: A search radius for determining correct localizations was used. Given that the spatial resolution of ultrasound systems is limited by diffraction to approximately half the transmitted wavelength, the search radius was set to $(\lambda/2)$. Localization results were then defined as following: a True Positive (TP) was defined as a localization inside a search radius surrounding a ground truth, a False Negative (FN) as a ground truth with no localization in its search radius, and a False Positive (FP) as a localization that is not inside the search radius of any ground truth. Only one localization inside the search radius can be associated with each ground truth position. This choice was selected so that it minimized the total accumulated distance. Precision and recall were used to evaluate the fractions of correct localizations. The Average Euclidean distance (AED) was used to indicate the average error between the ground truth bubble positions (B_i) and the participant's localization positions (L_i).

$$\text{Precision} = \frac{TP}{TP + FP} \quad (1)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (2)$$

$$\text{AED} = \frac{\sum_{i \in TP} \text{dist}(B_i \leftrightarrow L_i)}{N_{TP}} \quad (3)$$

2) *Tracking Evaluation*: Tracking is a process that analyzes localizations over consecutive frames and determines pairs of localizations that correspond to the same bubble. Tracking is strongly dependent on the localization performance. A bubble that was not localized will not be paired, and a FP localization should not be paired. If one were to consider all localizations for tracking, FN and FP localizations would generate a double penalty on the erroneous localizations. To separate the tracking evaluation from the localization evaluation, only true positive localizations were considered, i.e., only pairings of correctly localized bubbles in both frames are analyzed. This diminishes the total number of pairings to work with. As the total number of localizations increases, the impact of this systematic bias is reduced. The tracking or pairing process consists of associating localizations in consecutive frames by assigning identification numbers (IDs) to them. These results were then defined as following: A bubble whose localizations were assigned the same IDs were considered as a True Positive; A bubble whose localizations were assigned different IDs were considered as a False Negative; And localizations of two different bubbles

that were assigned the same ID were considered as a False Positive. Precision and recall as defined in Eqn. (1) & (2) were used to indicate the fractions of correct pairings.

To measure distance error for a pairing process, the pair distances, i.e., the distance between linked bubble pairs, can be used. Given that the total number of pairs is dependent on the localization process, the fraction of correct paired distance was used. This is equivalent to a weighted Jaccard index with the weights being the distance of each pair.

$$J = \frac{TP_d}{TP_d + FP_d + FN_d} \quad (4)$$

With: $TP_d = \sum_{i \in TP} \text{dist}(\text{pair}_i)$, $FP_d = \sum_{i \in FP} \text{dist}(\text{pair}_i)$, $FN_d = \sum_{i \in FN} \text{dist}(\text{pair}_i)$.

It is a fair metric, and bubbles that travel further away are more difficult to track, which would yield a better result in the metric. To provide better clarity, a remapping of this metric of the form $L = 2 * J - 1$ was used, so the range of the metric goes from $[0,1] \rightarrow [-1,1]$. The final metric has the following expression:

$$J_{map} = \frac{TP_d - FP_d - FN_d}{TP_d + FP_d + FN_d} \quad (5)$$

A negative number means more incorrectly paired distance than correctly paired distance, and a positive number the other way around. The metrics are summarized in Table II.

F. In-Vivo Data Generation

1) *Rat Brain Data*: The *in-vivo* rat brain dataset was obtained from the experimental data used in the article [36], Fig. 1, where the experimental procedures are detailed. The animal experiments were approved by institutional committee C2EA-54 “Comité d’éthique Normandie en matière d’expérimentation 628 animale” under the protocol APAFIS #22544-2019093017523797. Briefly, animals were anesthetized with 2% isoflurane before tail vein catheterization and craniotomy between Lambda and Bregma plane. Sonovue (Bracco, Milan, Italy) MBs were injected through the catheter at a rate of 80 microliter/min for a total of 4 minutes. A Verasonics Research scanner was used to drive a 15 MHz linear probe (Verasonics L22-14v High frequency imaging probe) with 128 elements (pitch of 0.1 mm) which was placed above the rat brain on a coronal plane. 2-cycles pulses were emitted at 15.625 MHz and 15 volts.

The continuous acquisition of 2 minutes comprised 150 blocks of 800 compounded frames (5 angles between -6 and +6 deg) acquired at 1000 Hz frame rate. Beamforming was performed posteriori with a conventional delay-and-sum processing (Verasonics) into $0.050 \times 0.049 \text{ mm}^2$ pixels.

For this challenge, data were filtered with a singular value decomposition: the first 14 singular vectors were removed to highlight only MBs. Finally, the data was cropped and downsampled from 1000 Hz to 100 Hz. The pre-processed data are summarized in Table III.

2) *Human Lymph Node Data*: The human lymph node CEUS data was acquired on a patient with suspected lymphadenopathy at the 2nd Affiliated Hospital of Zhejiang University.

Ethical approval was obtained from the University Institutional Ethics Board. Informed consent was given before the ultrasound examination. Ultrasound data acquisition was performed using a clinical scanner (Mindray Resona 7S) which provides live dual images of B-mode and CEUS. An L11-3U transducer was used with a transmit central frequency of 5.6 MHz. The patients were asked to breathe normally. B-mode and color Doppler ultrasound were first used to identify a long-axis view of the target node. Then, 1.2ml of SonoVue (Bracco, Milan, Italy) MBs was administered intravenously as a 2-second bolus followed by a flush of saline. Clinical CEUS data were acquired for 80 s after the injection, at a frame rate of 20 Hz in CEUS mode and a mechanical index (MI) of 0.085. The video data was offloaded from the clinical scanner, and then the CEUS data was motion corrected using the simultaneously acquired B-mode data.

G. Performance Ranking

All challenge entries in each category were ranked with every metric in the corresponding category, and then the ranks were summed to generate a final rank for each entry.

For the simulation localization category, the participants were asked to submit their results as a text file with a table containing one row per localization event, and, for each row, the frame number, bubble identifier, x, y, and z coordinates. With this standardized file format and an evaluation tool from the BUFF platform [37], the three metrics can be calculated automatically for each entry, based on which their rank for that metric was determined. The ranks for each entry were then summed to generate a final number, based on which the entry’s final rank in this category was determined (the lower the number, the higher the rank). For localization and tracking category, six metrics were calculated for each entry (three of which are the same as localization), and then similarly a final rank was determined based on summing up each entry’s individual rankings for the six metrics.

For the *in-vivo* data evaluation, participants were asked to submit their final ULM images with a uniform visualization using the MATLAB (MathWorks, USA) colormaps hot and jet for the structural (localization) and the functional (velocity) map respectively, together with a text file as that in the simulation. The expert panel judges were provided with the final ULM images from all participants and were asked to rank their top 6 participants for the rat brain and human lymph node separately according to their best knowledge of the specific organs and ULM data processing. Then the ranks for each dataset were summed in order to generate an overall rank for each entry. The judging was subjective, although the judges were provided with two image-based metrics to assist their judging: the imaging resolution, calculated using Fourier Ring Correlation measurements [54], and the image Gridding Index [36], [55].

H. Data Management and Communication

Following a registration process on the challenge website, the data was available on Dropbox (Dropbox, Inc., USA) via a personalized invitation. A README file containing detailed

instructions about the challenge was given and throughout the challenge a frequently asked questions document was updated. Participants also had access to a forum on the website to ask questions and updates were announced through a mailing list. On the day of the challenge and thereafter a survey was conducted. The survey was designed as a multiple-choice questionnaire and aimed to gather information about the participants' satisfaction regarding the challenge communication, data, data format and evaluation criteria. The level of satisfaction was measured from very dissatisfied (1) to very satisfied (5). The survey also sought to identify any gaps in the current challenge and gain insight into what participants believed should be included in a new challenge.

III. RESULTS

A. Synthetic and In-Vivo Data

The simulation data and an example of localization and tracking results are shown in Fig. 7. A segment of each simulation data set is also shown in the supplementary video 2-3. Additional results can be found on the [challenge's website](#). While the example ULM images in (e-g) picked up most of the vasculature as shown in the ground truth (a-c), some vessels are missing, particularly for the low frequency case. The *in-vivo* CEUS data from a rat brain and a human lymph node, together with examples of localization and tracking results, are shown in Fig. 8. A video segment of each *in-vivo* dataset is also shown in the supplementary video 4-5. Additional results can be found on the challenge's website.

B. Performance of Algorithms

Fig. 9 shows the ranking of the different algorithms in the two simulation categories as well as their performance measured using each of the six metrics. It can be seen from the individual metrics that while there is some spread in performance between the different methods, there is not a single method dominating the localization or the tracking performance. Indeed, most metrics for the top few performing methods are relatively closely spaced, and the winners often perform well in most categories although not necessarily topping every category. This is because some of the categories are linked, and an improved precision can be achieved at the cost of a lower recall. For example, for localization, the winner (RUB group) with a Gaussian fitting approach ranked top in localization recall, second in localization mean error, but only 6th in localization precision, while the joint winner (Concordia group) with a normalized cross-correlation approach ranked 2nd in localization recall, 3rd in localization precision, and 5th in localization mean error. For overall performance which include both localization and tracking performance, the winner (Imperial group) with a deconvolution and feature motion model-based tracking approach ranked 2nd in tracking precision, 3rd in tracking recall, and 3rd in tracking distance score. They also ranked 2nd in localization precision, 1st in localization mean error, but their localization recall is only ranked 10th at 43%, compared to the top performing RUB group (56%). The joint runner ups for the overall category include the Harvard group (who used morphological reconstruction for

localization and Hungarian data association algorithm), and the RUB group (who used a Gaussian fitting for localization and a Kalman motion model for tracking). It should be noted that RUB group has the highest recall in localization, which poses extra challenges in tracking as they have the highest number of bubbles to track. It should also be noted that some methods (e.g. EID20) provide the top performance in tracking, but did not make it to the winners' list, as they had a low localization performance. The individual localization recall (which corresponds to the total number of bubbles to be tracked by each contestant) impacts the tracking performance as fewer bubbles makes tracking easier. Therefore, the tracking performance in Fig. 9 should be interpreted with consideration of localization recall.

The rankings for the *in-vivo* rat brain dataset and for the human lymph node dataset are shown in Table IV & V. The final combined ranking for the top teams is shown in Table VI. It can be seen that besides the winning teams, the team EID8 also performed well in both categories, although it did not make it to the top three in the final ranking.

C. Survey

A total of 19 individuals responded to our survey regarding their satisfaction and limiting factors of the ULTRA-SR challenge (multiple answers possible: 16 participants, 1 advisory committee member, 1 judge, 7 researchers interested in ULM). 89% of respondents were satisfied or very satisfied with the overall challenge. The consensus was that the challenge data was appropriate with the rat brain data receiving the highest rating from 55% correspondents while the test data only receiving the highest rating from 38%. Two participants had difficulties processing data without bias, but the majority did not and 72% were satisfied or very satisfied with the realism of the simulated data. The participants were more in favour (satisfied or very satisfied) of the evaluation metrics (92%) over the expert panel (69%) and would support the evaluation of localization and tracking in a future challenge (100%) compared to 82% for the expert panel. A popular vote (42%) and use of other metrics (78%) were considered too. According to half of the respondents a future challenge could increase the complexity by including 3D data, motion, and some degree of artefacts. However, the type of artefact was debated. Access to the raw data, anonymous live ranking, centralised management (everything accessible through the website) and the ability to test algorithm repeatedly were the main points of criticism.

IV. DISCUSSION

In this paper both simulation data with ground truth and *in-vivo* data, together with simple performance metrics for both localization and tracking, useful for training and benchmarking ULM algorithms, are provided. These data and metrics were used in the ULTRA-SR challenge. One main finding from our results is that the quality of the resulting ULM images can vary significantly depending on the localization and tracking methods, and that effective performance can be attained with different methodologies. For localization at both frequencies,

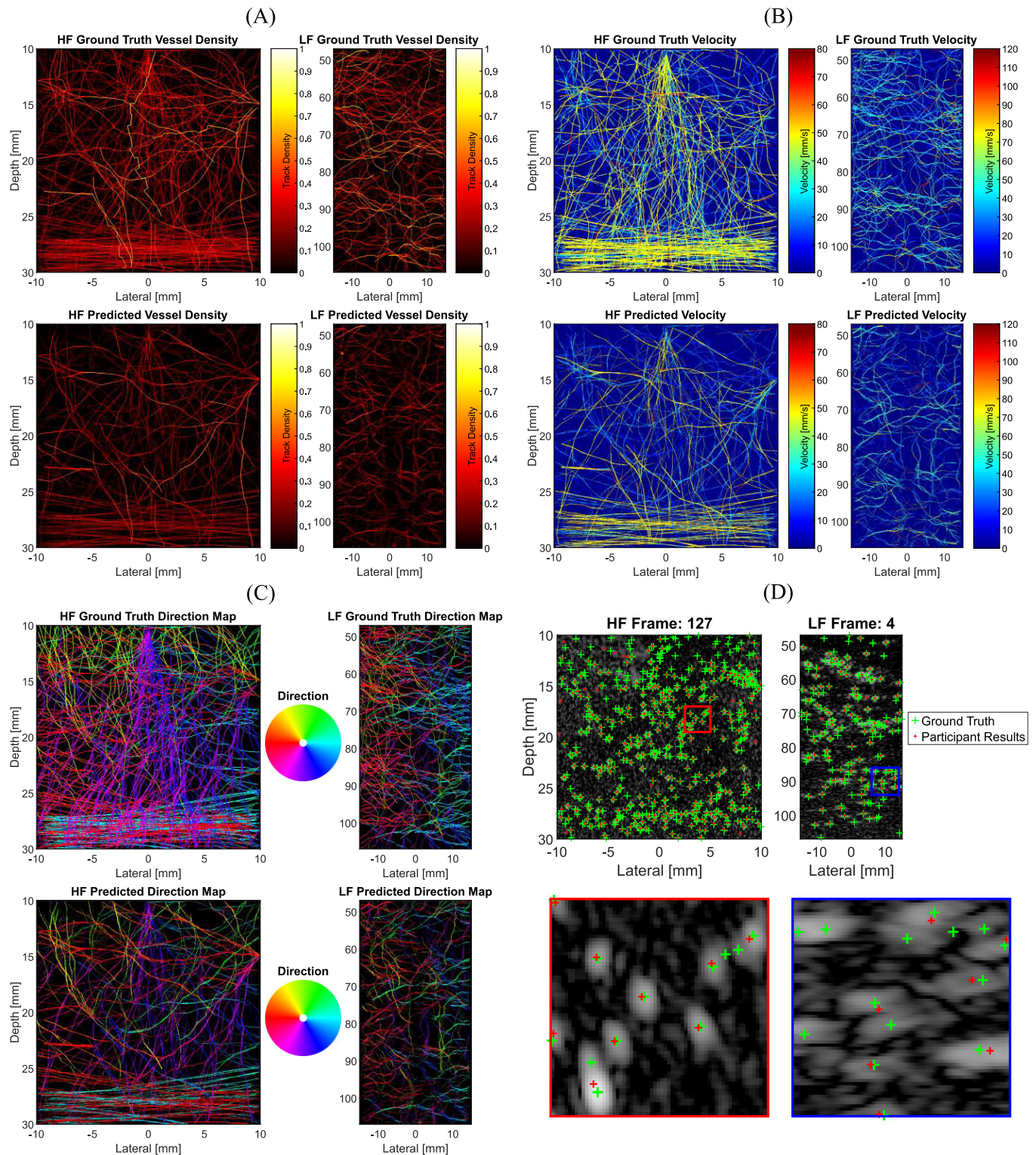


Fig. 7. Microvascular structures generated for the ULTRA-SR challenge and an example submission from winning team EID22 | Imperial College. (left) High frequency, (right) Low frequency; Top - ground truth, Bot - winning example. (A) Vessel Density Maps; (B) Flow Velocity Map; (C) Flow Direction Map (D) Individual frames of the simulation datasets with GT bubble locations (green +) and example localizations (red +) The datasets and participant's results are accessible from the [challenge website](#). Video clips of the data sets can also be found in supplementary video 2-3.

Gaussian fitting, normalized cross-correlation (NCC), and deconvolution-based methods, all established signal processing tools applied to ULM, have achieved top performance among the participating methods. For the overall category which includes tracking, the top performing methods all used some

form of Kalman filter to inform bubble tracking. Furthermore, it should be noted that specific implementation of a method holds significance and could result in different performance outcomes. For example, NCC-based localization approaches were used by multiple groups, but only one entry was ranked

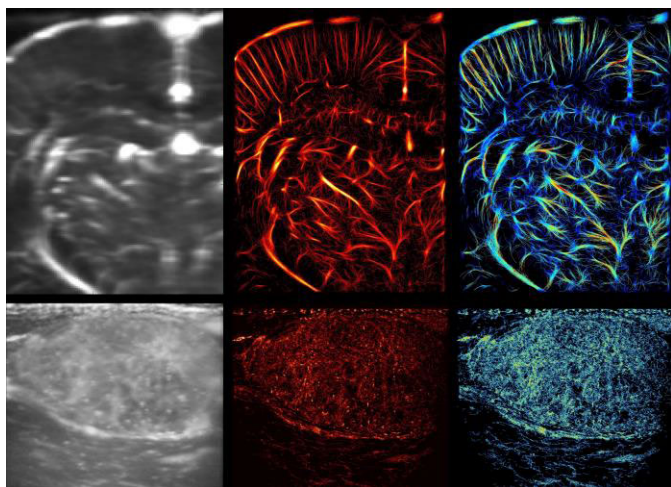


Fig. 8. In-vivo data and winning results for the brain data from EID 6 | UIUC group, and winning results for the lymph node data from EID12 | TU/e group. (Top) Left: a CEUS image of a rat brain averaged across the time sequences; Middle: a winning brain ULM localization map; Right: a winning velocity map. (Bottom) the same as top, but for a human lymph node. The datasets and participant's results are accessible from the [challenge website](#). Video clips of the data sets can also be found in supplementary video 4-5.

as a top 3 for both the high and low frequency datasets. Additionally, algorithm parameters (e.g., singular values to retain, tracking tolerances, etc.) can influence the outputs, and may need to be optimized for a particular dataset.

This highlights that understanding the functioning of specific methods is important for their optimal implementation, and that thorough adaptation of algorithms to each imaging sequence seems to be necessary. Thus, one future research aspect would be to make ULM less sequence/application dependent. Furthermore, while this manuscript has described for each algorithm the key parameters and their values, the sensitivity of these parameters to the outcome is also important and could be a subject of future evaluation.

The simulation datasets were generated using a recently developed simulation framework, BUbbly Flow Field (BUFF) [37], which incorporates 1) a 3D microvascular tree generator; 2) a Hagen–Poiseuille vascular flow model; 3) an acoustic field simulator, and 4) a MB dynamics simulator coupled to the field simulator. Particularly the last feature of incorporating MB oscillation physics offers improved realism to the datasets, as oscillating bubbles can result in altered PSFs e.g. through elongating the echo signals and distorting the signals with nonlinear components.

The datasets in this challenge are purposely designed to be difficult, with relatively low frame rate, limited number of image frames, and significant variations of some key parameters including frequency, flow velocity, bubble concentration, and image SNR. Two simulation datasets were generated, each with a typical clinical frequency. Significant variations in flow velocity were built into each dataset through having vessels of very different flow velocities. Note that varying flow velocity has a similar effect as frame rate for tracking purposes. Significant variations in bubble concentration were built into the datasets through randomizing the spatial distribution of the vessels and bubbles, so that there will be a mixture of both

high and low bubble concentration in local areas. Furthermore, significant variations of SNR were built into our data through a varying SNR over the depth of the image.

The metrics used to evaluate the algorithms include precision, recall, and mean error for both localization and tracking. The average precision and recall of the algorithms in this challenge for the top teams was 0.821 and 0.502 for localization, and 0.766 and 0.691 for tracking. The average localization error for the top teams was $31.557 \mu\text{m}$ for the HF dataset, and $141.055 \mu\text{m}$ for the LF dataset. The purpose of the challenge is to compare the different algorithms, and while such numbers allow benchmarking, it would also be interesting to discuss how the numbers impact final results and what value would be desirable for generating an “acceptable” ULM image. Firstly, all the metrics heavily depend on the data, e.g. higher SNR and contrast in the initial CEUS image data would certainly improve localization metrics, and high frame rate data of low bubble concentration could generate excellent tracking metrics even for simple algorithm such as nearest neighbor. Therefore, while such metrics are entirely suitable for comparing algorithms when the same dataset is used, their absolute numbers should be interpreted with caution as they are dataset-dependent. Localization errors, if not a systematic bias, directly impact the best image resolution achievable. The errors are frequency dependent and should be less than half a wavelength in order to achieve super-resolution. Note that in this challenge only those localizations within half a wavelength of a true bubble position were counted, as the purpose was for comparing the different algorithms. The errors larger than this were accounted for in the precision metric. It should be noted that median errors may be a better metric than mean errors as it is likely that the distributions of errors may not be normally distributed. If looking at the challenge data sets specifically, as many bubble signals are below noise level in the initial CEUS images due to their sizes, a localization recall of over 50% would mean the majority of visible bubbles are detected. Localization precision determines the SNR in the final ULM images, as incorrect localization (or localization errors larger than half a wavelength) would contribute to noise in the final image. For example, precision of 75% or greater would mean an equivalent final ULM image SNR of ~ 10 dB (or perhaps greater, given that some wrong localization could still contribute to signals if within a vessel). For tracking, precision is important as we want to minimize wrong tracks. However often only tracks longer than a number of frames are kept in the final ULM maps, and wrong tracks have less chance to extend multiple frames if e.g. a motion model is used [14]. Tracking recall will determine the vascular saturation in the final ULM image, and 50% means they miss half of the tracks. Furthermore, the final image quality is also affected significantly by how good the motion correction is, which is a separate topic of its own and not included in this challenge.

For evaluating ULM imaging, other possible metrics include image resolution, saturation curves, gridding index, and image similarity to ground truth images. As the evaluation metric should be effective but also kept simple, the consensus of the challenge advisory committee was that how well bubbles are

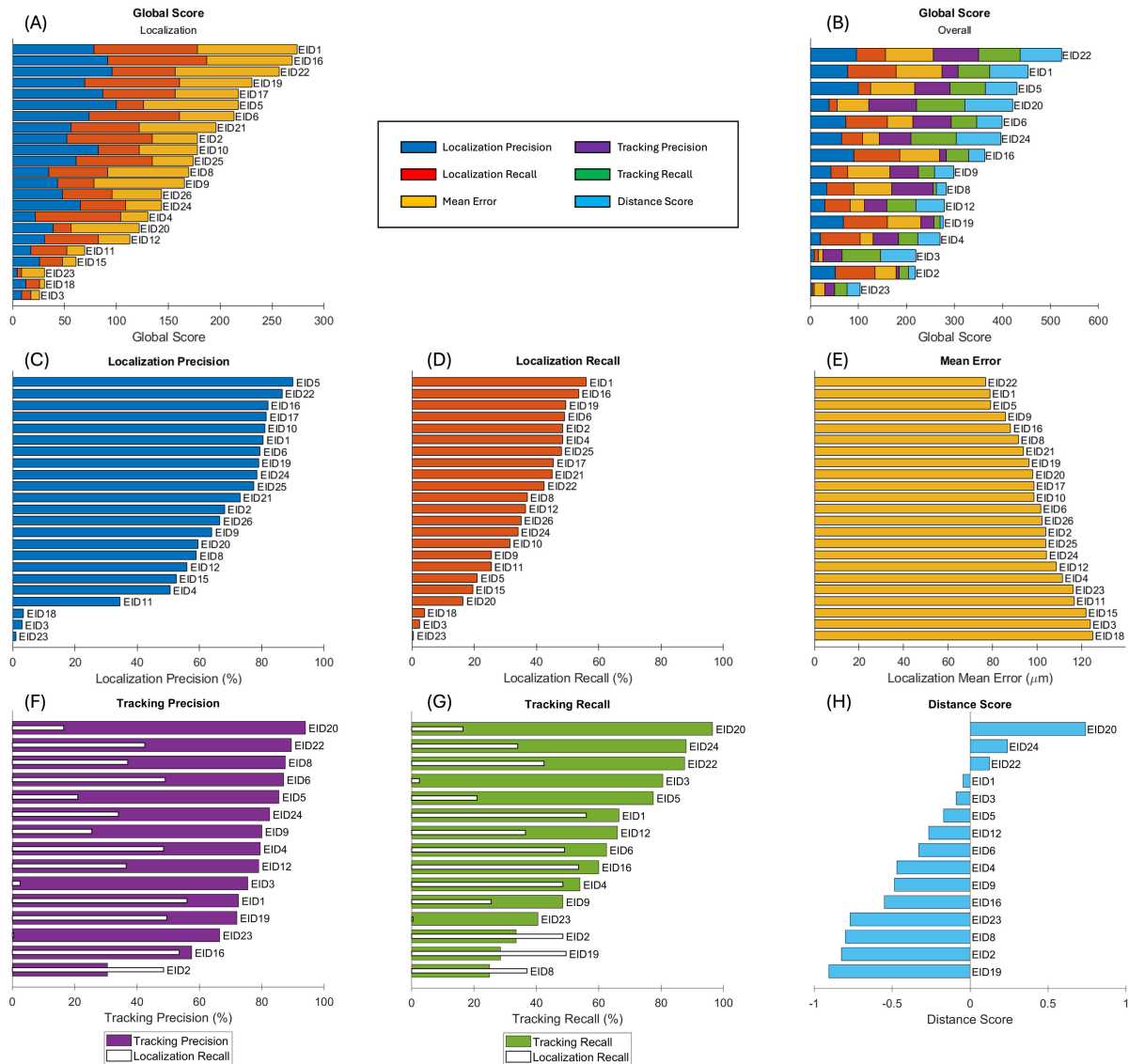


Fig. 9. Evaluation and ranking of participants based on simulation data. The bar graphs depict the competitors' scores ranked from the highest to lowest (top-down). (Top row) Global ranking of all contestants averaged across the high and low frequency data sets. Each colored segment corresponds to a ranking score between 0 and 100. (A) Localization Category; (B) Overall Category. (Middle row) evaluation of the performance for all contestants using the three localization metrics. Each bar in represents a number calculated from the formulas in Methods section. (C) Localization Precision; (D) Localization Recall; (E) Localization Mean Error. (Bottom rows) evaluation of the performance for all contestants using the three tracking metrics. (F) Tracking Precision; (G) Tracking Recall; (H) Tracking Distance Score. The overlaid thin white bar in (F) and (G) indicate the corresponding localization recall. Note that tracking precision and recall cannot be assessed without considering localization recall, since perfect tracking can be obtained if there are very few bubbles to track (i.e. extremely low localization recall). A higher localization recall means a larger number of bubbles and available bubble pairs to track, which would increase the difficulties for tracking. EID1: RUB group, joint winner of the localization category, and joint runner-up in the overall category; EID5: Harvard group, joint runner up of the overall category. EID16: Concordia Group, joint winner of the localization category; EID22: Imperial group, winner of the overall category.

localized and tracked, as we proposed, are the most direct metrics for evaluating localization and tracking for ULM when ground truth is available.

There are a number of deep learning (DL) based methods for localization and also some for bubble tracking. Compared to the non-DL methods, DL ones usually heavily rely on the quality of training data. In this challenge only very limited test datasets consisting of 20 image frames with ground truth were published in advance. This could be a reason why most DL-based algorithms did not work well in the simulation category, although one of them, 'EID12 – TU/e DeepULM', was ranked among top performers in the *in-vivo* category. Supervised learning methods usually rely on huge datasets

—in the order of thousands of sets with ground truth, thus, the amount of data should then be increased at least 50-fold to significantly improve the performance of these methods in simulation.

There are some limitations of this challenge, which should be considered for future competitions. Firstly, the simulation data did not consider tissue nonlinear signal breakthrough, large vessel flow dynamics, and speed of sound aberration. Secondly, tissue motion was not included, although this was deliberate as we want to focus on evaluating localization and tracking algorithms, and motion in the data may confound the evaluation. Motion correction could be the subject of a future challenge. For both simulation data sets, some MB

TABLE IV

IN-VIVO RAT BRAIN JUDGE PANEL RANKINGS. EACH JUDGE WAS ASKED TO RANK THEIR TOP 6 ENTRIES. *ONLY ENTRIES RANKED AS TOP 6 FOR AT LEAST 6 PANEL MEMBERS ARE CONSIDERED FOR THE FINAL RANKING. (GREEN: 1st, YELLOW: 2nd RED: 3rd)

ID	Panel #1	Panel #2	Panel #3	Panel #4	Panel #5	Panel #6	Panel #7	Panel #8	Panel #9	Panel #10	Panel #11	Panel #12	Panel #13	Panel #14	Panel #15	Panel #16	Sum	Votes	AVG
EID1	3	2	4	5	4	6	6	5	6		2	4	6	1	5	5	64	15	4.267
EID2																	0	0	-
EID3																	0	0	-
EID4																	0	0	-
EID5	4				6					4	3	6			6	1	30	7	4.286
EID6	1	1	1	1	3	1	1	3	2	1	1	2	1	1	1	3	24	16	1.500
EID8	6	2	3	3		3	2	6		2	4	5	3	2	3		44	13	3.385
EID9				6					3		6						15	3	5.000*
EID12		5	5	4	5	2		1	4	6		1	2	5	2	6	48	13	3.692
EID14																	0	0	-
EID16	5		6		2		5		5							2	25	6	4.167
EID19													4			4	8	2	4.000*
EID20		6		6		5		4									21	4	5.250*
EID22	2	4	1	2	1	4	3	2	1	3	5	3		4	4		39	14	2.786
EID23							4			5			5	6			20	4	5.000*

TABLE V

IN-VIVO LYMPH NODE JUDGE PANEL RANKINGS. EACH JUDGE WAS ASKED TO RANK THEIR TOP 6 ENTRIES. *ONLY ENTRIES RANKED AS TOP 6 FOR AT LEAST 6 PANEL MEMBERS ARE CONSIDERED FOR THE FINAL RANKING. (GREEN: 1st, YELLOW: 2nd RED: 3rd)

ID	Panel #1	Panel #2	Panel #3	Panel #4	Panel #5	Panel #6	Panel #7	Panel #8	Panel #9	Panel #10	Panel #11	Panel #12	Panel #13	Panel #14	Panel #15	Panel #16	Sum	Votes	AVG
EID1	5	6		6				5	6							3	31	6	5.167*
EID2							6	3									9	2	4.500*
EID3					6		5										11	2	5.500*
EID4																	0	0	-
EID5	6																6	1	6.000*
EID6		1		5		2		6		6	3		5	6		4	38	9	4.222
EID8			5	3	4	3	3		3			6	2	3	2		34	10	3.400
EID9	3		4		2				5	5	1	3	3	5	5	5	41	11	3.727
EID12	1	4	3	4	1	1	4	2	1		2	2	1	4	3		33	14	2.357
EID14													6				6	1	6.000*
EID16	2						2	4		4	6	5				1	24	7	3.428
EID19		3				4					4	4					15	4	3.750*
EID20		5	2	2	5	6			4	3					4	6	37	9	4.111
EID22	4	1	1	1	3	5	1	1	2	2	5	1	4	1	1	2	35	16	2.187
EID23										1				2			3	2	1.500*

positions of the ground truth appear repeatedly on the exact same positions somewhat reducing the realism of the simulations. Furthermore, the assessment of algorithms did not consider the different algorithms' computing requirements, including processing time and hardware (CPU, GPU and memory) requirements. The fact that some methods could be much more time-consuming than others should be con-

sidered when choosing a suitable approach for any specific application. Additionally, for simulation data we only had two challenge categories, localization, and overall (including both localization and tracking). We did not have a challenge category to evaluate tracking only, as some algorithms do not have a clear cut between localization and tracking, and such a tracking evaluation would require a large amount of image

TABLE VI

COMBINED JUDGE PANEL RANKINGS FOR THE TOP 7 TEAMS (THESE TEAMS ARE LISTED HERE AS THEY WERE RANKED AMONG THE TOP 6 FOR AT LEAST 6 PANEL MEMBERS). EID6: UIUC GROUP, WINNER OF THE *IN-VIVO* CATEGORY; EID12: TU/e GROUP, JOINT RUNNER-UP. EID22: IMPERIAL GROUP, JOINT RUNNER-UP. THE WINNING DECISION WAS BASED ON THE TEAM (EID6) ACHIEVING BY FAR THE OVERALL BEST RANKING FOR THE BRAIN. * FRC VALUE CANNOT BE CALCULATED

Participant's ID	Average Brain Ranking	Average Lymph Ranking	Combined Ranking	FRC Resolution [um] (Brain / Lymph Node)	Ranking
EID1	4.267	5.167	9.433	34 / 60	
EID5	4.286	6.000	10.286	*/316	
EID6 UIUC	1.500	4.222	5.722	28 / 37	1 st
EID8	3.384	3.400	6.785	*/34	
EID12 TU/e	3.692	2.357	6.049	35 / 42	2 nd
EID16	4.167	3.429	7.595	30 / 49	
EID22 Imperial	2.786	2.188	4.973	28 / 37	2 nd

data with bubble location ground truth, interfering with the localization competition and potentially favor those with e.g. a deep learning approach. It was decided that evaluating both localization, and localization and tracking, was a simple and sufficient solution to this complication. Though for the future, a tracking-only challenge would be useful to fully evaluate the different tracking algorithms.

It can be seen from the *in-vivo* results that the quality of the lymph node results are inferior to those of the rat brain. This is expected mainly for two reasons: firstly the data acquisition systems are very different. The rat brain data were acquired by a Verasonics research platform at an order of magnitude higher frame rate than that used for the clinical lymph node acquisition. Most existing clinical systems are optimized for image quality and often deliberately use low frame rate to avoid bubble destruction. However, such low frame rates pose.

Since the *in-vivo* part of the challenge was evaluated by an expert panel ranking, it was subjective in nature and may factor images that are visually appealing although it might not be important at all when it comes down to diagnostic values of imaging. Generally, *in-vivo* results are difficult to evaluate, as there is no ground truth nor techniques available to provide reference images of the microvasculature. Micro-CT, histology, optical imaging, etc. all have limitations as a reference. A number of metrics can be used to assist the evaluation, although each of the metrics has their limitations [36]. A consensus on the evaluation and presentation of *in-vivo* data would be desirable since ULM images are intended to provide detailed and reliable information in addition to just the images. In addition, a common problem in the presentation of super-resolution images is that they are displayed in resolution-limited media, so that the visual impression depends strongly on the respective display parameters. That is, an image looks good at first, but details may be questionable when zoomed in, or, conversely, is very good in detail, but loses expressiveness when zoomed out (e.g., small vessels are no longer visible). This effect can be greater with larger ROI. One suggestion for evaluating *in-vivo* ULM images is that, in addition to evaluating the whole images, zoomed-in regions should be used as part of the evaluation.

Furthermore, while more quantitative measures such as FRC and gridding index were provided to the panel to potentially

assist their judgement, it was up to the panel members as to whether or how they use such information. In future evaluations for *in-vivo* data, such quantitative evaluation of the quality of the final reconstructed vessels, in addition to task-specific metrics (e.g., sensitivity, specificity for early detection and diagnosis of diseases), could be used directly as part of the final ranking scores.

The outcome of the survey showed high satisfaction with the ULTRA-SR challenge. The data was deemed appropriate and relevant for the challenge and the evaluation metrics were met with satisfaction. This was the first attempt to create an open challenge with the ultrasound super-resolution community. A future challenge could increase the complexity of the challenge by including 3D data, potentially with access to not only image data but raw channel radio frequency data, tissue motion, and more complex image artefacts.

Some further aspects of the winning algorithms, including their strength and weakness, are discussed below:

A. TU/e

Our method was the joint runner-up in the *in-vivo* category. We suspect that our method performed well in the *in-vivo* domain due to its data driven nature. Heuristic ULM methods work in well-defined settings, with clear PSF shapes, a low concentration and limited number of overlapping bubbles. In the case of the lymph node *in-vivo* dataset, due to the bolus injection, the concentrations of bubbles are very high. Furthermore, the addition of noise artifacts limits the performance of conventional algorithms heavily. Incorporating these noise artifacts in our training data helped our data driven localization method detect more bubbles.

Our method struggled a bit more with the simulation method. The main cause was the lower sensitivity in deeper regions of our localization method. The lower SNR and larger PSF shapes proved to be a challenge. We could address this by making our algorithm depth dependent, allowing it to focus more on those regions. Furthermore, for our method to perform well, the data distribution between the training set and test set needs to be close. However, for the simulation evaluation set, not all the information regarding the ultrasound imaging scheme was provided, so some of imaging parameters had to

be estimated. This brought the discrepancy between the training and evaluation sets, which made our method struggle more.

B. UIUC

Our group used normalized cross-correlation for MB localization using an empirically chosen MB PSF template. This technique has several advantages, including relatively intuitive parameter selection, high computational efficiency, and parallelization potential. However, the procedure does suffer from operation dependence for tuning, and a fixed PSF template will not adequately capture variable MB PSFs arising under diverse conditions (e.g.: spatial PSF variance, polydisperse MB populations, nonlinear scattering, etc.). This combination of features may explain the relative underperformance on simulation datasets, which exhibited a high degree of spatial variance and MB concentrations, in comparison to the *in-vivo* high frequency dataset. Future work will focus on deep-learning solutions to MB localization to address the domain mismatch between simulation/training datasets and *in-vivo* MB signals, with evidence that this approach better captures *in-vivo* vasculature at high concentration, thereby reducing data acquisition time [56].

Our Kalman-filter based bipartite-graph-based pairing algorithm is conceptually similar to conventional Hungarian (Kuhn–Munkres algorithm) pairing solutions. The process is robust and computationally efficient, and the addition of the Kalman-filter motion model can correct for noisy localization samples through consideration of the entire MB trajectory. However, the *a priori* motion model can be too aggressive, obscuring some of the more erratic MB trajectories – especially for slower flowing MBs in microvasculature. Like all bipartite-graph-based solutions, the maximum pairing distance and MB persistence parameters need to be finely tuned to the experimental dataset, which introduces an element of operator dependence.

C. RUB

From our experience, 2D Gaussian fit and matched filtering with a Gaussian kernel work best for different data sets as these algorithms are robust when dealing with different noise levels [36], [54], [57]. Generally, the localization in the low frequency data set performed better because the PSF covers more pixels [57]. In contrast, the PSF of the high frequency data set spans fewer pixels. Also, the noise in the high frequency data set made localization and therefore tracking more difficult. Particularly in the lower image area and at the upper corners, the SNR is poor, so that it is difficult to visually distinguish between noise and MBs. In addition, the low frequency dataset seems to have a more realistic vascular structure, since long straight vessels only occur in certain sectional views of well-organized organs (e.g., brain, kidney). Depending on the choice of parameters (especially the detection threshold), each performance metric score can be influenced, so that an appropriate optimization is necessary. The motion model incorporated in the tracking algorithm leads to the ability of tracking MBs in difficult situations like crossing trajectories, even when the concentration is high, and the frame rate is low.

D. HMS/MGH

The high overall score of our MR algorithm was largely due to its high precision. While the recall of our method did not rank very highly (i.e., there were many missed localizations), the precision and accuracy of the peaks that were detected in turn allowed efficient tracking, and thus a favourable overall score.

The principal benefit of the MR algorithm is that it enables a large increase in the effective dynamic range of MB signals that can be detected (see Fig. 4). Thus, signals that are weaker due to MB's position in the frame, or its potentially being slightly outside the elevational focus, are more likely to be detected than if a constant threshold were used. Additionally, as a binary operation, the final map centroids are reliably close to the intensity distribution's peak (i.e., the MB position). While very effective for isolated MB signals, MR is still limited in that it cannot distinguish two MBs separated by less than the PSF size – such sources would be localized as a single MB the centroid of their combined distribution.

E. Imperial

The deconvolution method used can separate close MBs. Some MBs with their images partly overlapped can be isolated as a single MB after a simple thresholding but can be further isolated as individual MBs after being shrunk by the deconvolution, as demonstrated in a previous study [29]. However, deconvolution is more sensitive to SNR than the normalized cross-correlation and matching filter, as the latter two can smooth the image with a kernel. A low-intensity main lobe with relatively high noise was sometimes too different from the estimated PSF and thus deconvolved into multiple MBs. Therefore, we chose to use relatively strong thresholding to discard the weak signals, which could have reduced the recall of localization. Besides, the weight for the sparsity regularization and the stopping criteria we used for our Fast Iterative Shrinkage-Thresholding Algorithm can also affect the performance of deconvolution.

Using a Kalman motion model to predict MB movement, estimating Kalman parameters from the data and estimating velocity of new MBs made the feature-motion-model framework outperforming. Combining Kalman motion model and MB image features in tracking made tracking more robust, as demonstrated in our previous study. Parameters for linear Kalman filtering, i.e., covariance of the process noise and covariance of the observation noise, were estimated from the data, as presented in [26], to improve tracking performance and reduce human input. MBs that moved into the imaging plane were given velocity at their first frame by Fuzzy Initialization before pairing which can enable motion model for newly appeared MBs. While MBs move in curved vessels, nonlinear motion model in tracking could further improve the tracking performance [58].

F. Concordia

The idea of formulating the bubble-pairing problem as an iterative registration problem was the key ingredient in the performance of our tracking algorithm. The transformation

function and the probability table iteratively drive each other towards the correct solution of the bubble-set registration problem. This complementary update of the transformation function and the probability table, which ensures robust MB tracking, was the main strength of our technique. Nevertheless, this algorithm pairs the bubbles in two consecutive frames. Imposing temporal consistency on the velocities of bubbles in multiple frames might improve the accuracy of MB tracking.

V. CONCLUSION

In conclusion, Ultra-SR challenge has provided benchmarking data and tools as well as direct comparison and insights for a number of the state-of-the-art localization and tracking algorithms, and is well received by the participants.

The survey results also revealed that the ULTRA-SR challenge was well-received and met its goals in terms of engagement and participant contentment. A significant majority was content with the data, and a commanding majority (92%) were in favor of the evaluation metrics. There was unanimous support for incorporating localization and tracking assessments in future iterations of the challenge.

ACKNOWLEDGMENT

The authors would like to thank Prof. M. Bell, Prof. H. Liebgott, Prof. S. Chen, Dr. K. Christensen-Jeffery, Prof. P. Dayton, Prof. Y. Eldar, Dr. S. Harput, Prof. J. Luo, Prof. M. Misch, Prof. J. Provost, Prof. M. O'Reilly, Prof. M. Tanter, and Prof. M. Wan for their valuable advice on the organisation of the ULTRA-SR Challenge.

Marcelo Lerendegui, Kai Riemer, Bingxue Wang, Jipeng Yan, and Meng-Xing Tang are with the Department of Bioengineering, Imperial College London, SW7 2AZ London, U.K. (e-mail: m.lerendegui@imperial.ac.uk; kr616@ic.ac.uk; bingxue.wang18@imperial.ac.uk; j.yan19@imperial.ac.uk; mengxing.tang@ic.ac.uk).

Georgios Papageorgiou, Lachlan Arthur, and Vassilis Sboros are with the Institute of Biological Chemistry, Biophysics and Bioengineering, Heriot-Watt University, EH14 4AS Edinburgh, U.K. (e-mail: G.Papageorgiou@hw.ac.uk; la25@hw.ac.uk; V.Sboros@hw.ac.uk).

Arthur Chavignon and Olivier Couture are with Laboratoire d'Imagerie Biomedicale, CNRS, INSERM, Sorbonne Université, 75006 Paris, France (e-mail: arthur.chavignon.pro@gmail.com; olivier.couture@sorbonne-universite.fr).

Tao Zhang and Pingtong Huang are with the Department of Ultrasound, The Second Affiliated Hospital, Zhejiang University, Zhejiang 310027, China (e-mail: zhangtao-us@zju.edu.cn; huangpingtong@zju.edu.cn).

Md Ashikuzzaman was with the Department of Electrical and Computer Engineering, Concordia University, Montreal, QC H3G 2W1, Canada. He is now with the Department of Electrical and Computer Engineering, Johns Hopkins University, Baltimore, MD 21218, USA (e-mail: mashiku1@jh.edu).

Stefanie Dencks, Thomas Lisson, and Georg Schmitz are with the Faculty of Electrical Engineering and Information Technology, Ruhr University Bochum, 44801 Bochum, Germany (e-mail: stefanie.dencks@rub.de; thomas.lisson@ruhr-uni-bochum.de; georg.schmitz@rub.de).

Chris Dunsby is with the Department of Physics, Imperial College London, SW7 2AZ London, U.K. (e-mail: christopher.dunsby@imperial.ac.uk).

Brandon Helfield is with the Department of Physics and the Department of Biology, Concordia University, Montreal, QC H3G 2W1, Canada (e-mail: brandon.helfield@concordia.ca).

Jørgen Arendt Jensen is with the Department of Health Technology, Technical University of Denmark, 2800 Copenhagen, Denmark (e-mail: jaje@dtu.dk).

Matthew R. Lowerison and Pengfei Song are with the Department of Electrical and Computer Engineering and the Beckman Institute, University of Illinois Urbana-Champaign, Urbana, IL 61820 USA (e-mail: mloweri@illinois.edu; songp@illinois.edu).

Hassan Rivaz is with the Department of Electrical and Computer Engineering, Concordia University, Montreal, QC H3G 2W1, Canada (e-mail: hrivaz@gmail.com).

Anthony E. Samir and Scott Schoen Jr., are with the Harvard Medical School, Massachusetts General Hospital, Boston, MA 02114 USA (e-mail: asamir@mgh.harvard.edu; sschoenjr@mgh.harvard.edu).

Ruud van Sloun and Tristan Stevens are the Department of Electrical Engineering, with Eindhoven University of Technology, 5612 AZ Eindhoven, The Netherlands (e-mail: r.j.g.v.sloun@tue.nl; t.s.w.stevens@tue.nl).

REFERENCES

- [1] M. Dewey et al., "Clinical quantitative cardiac imaging for the assessment of myocardial ischaemia," *Nature Rev. Cardiol.*, vol. 17, no. 7, pp. 427–450, Jul. 2020, doi: [10.1038/s41569-020-0341-8](#).
- [2] J. R. Lindner, "Microbubbles in medical imaging: Current applications and future directions," *Nature Rev. Drug Discovery*, vol. 3, no. 6, pp. 527–533, Jun. 2004, doi: [10.1038/nrd1417](#).
- [3] S. R. Wilson and P. N. Burns, "Microbubble-enhanced U.S. in body imaging: What role?" *Radiology*, vol. 257, no. 1, pp. 24–39, Oct. 2010, doi: [10.1148/radiol.10091210](#).
- [4] M.-X. Tang et al., "Quantitative contrast-enhanced ultrasound imaging: A review of sources of variability," *Interface Focus*, vol. 1, no. 4, pp. 520–539, May 2011, doi: [10.1098/rsfs.2011.0026](#).
- [5] K. Christensen-Jeffries et al., "Super-resolution ultrasound imaging," *Ultrasound Med. Biol.*, vol. 46, no. 4, pp. 865–891, 2020, doi: [10.1016/j.ultrasmedbio.2019.11.013](#).
- [6] O. Couture, V. Hingot, B. Heiles, P. Muleki-Seya, and M. Tanter, "Ultrasound localization microscopy and super-resolution: A state of the art," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 65, no. 8, pp. 1304–1320, Aug. 2018, doi: [10.1109/TUFFC.2018.2850811](#).
- [7] 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, pp. 6447–6458, Sep. 2013, doi: [10.1088/0031-9155/58/18/6447](#).
- [8] 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, doi: [10.1063/1.4826597](#).
- [9] 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, doi: [10.1109/ULTSYM.2011.0476](#).
- [10] 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, doi: [10.1109/TMI.2014.2359650](#).
- [11] C. Errico et al., "Ultrafast ultrasound localization microscopy for deep super-resolution vascular imaging," *Nature*, vol. 527, no. 7579, pp. 499–502, Nov. 2015, doi: [10.1038/nature16066](#).
- [12] 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, doi: [10.1016/j.ultras.2017.01.008](#).
- [13] 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, doi: [10.7150/thno.16899](#).
- [14] T. Opacic et al., "Motion model ultrasound localization microscopy for preclinical and clinical multiparametric tumor characterization," *Nature Commun.*, vol. 9, no. 1, p. 1527, 2018, doi: [10.1038/s41467-018-03973-8](#).
- [15] S. B. Andersen et al., "Super-resolution imaging with ultrasound for visualization of the renal microvasculature in rats before and after renal ischemia: A pilot study," *Diagnostics*, vol. 10, no. 11, p. 862, Oct. 2020, doi: [10.3390/diagnostics10110862](#).
- [16] Q. Chen, J. Yu, B. M. Rush, S. D. Stocker, R. J. Tan, and K. Kim, "Ultrasound super-resolution imaging provides a noninvasive assessment of renal microvasculature changes during mouse acute kidney injury," *Kidney Int.*, vol. 98, no. 2, pp. 355–365, Aug. 2020, doi: [10.1016/j.kint.2020.02.011](#).
- [17] J. Foiret, H. Zhang, T. Ilvitsk, 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, Oct. 2017, Art. no. 1, doi: [10.1038/s41598-017-13676-7](#).
- [18] J. Zhu et al., "3D super-resolution U.S. imaging of rabbit lymph node vasculature in vivo by using microbubbles," *Radiology*, vol. 291, no. 3, pp. 642–650, Jun. 2019, doi: [10.1148/radiol.2019182593](#).

- [19] N. Renaudin, C. Deme  , A. Dizeux, N. Ialy-Radio, S. Pezet, and M. Tanter, "Functional ultrasound localization microscopy reveals brain-wide neurovascular activity on a microscopic scale," *Nature Methods*, vol. 19, no. 8, pp. 1004–1012, Aug. 2022, doi: [10.1038/s41592-022-01549-5](https://doi.org/10.1038/s41592-022-01549-5).
- [20] O. Demeulenaere et al., "Coronary flow assessment using 3-Dimensional ultrafast ultrasound localization microscopy," *JACC, Cardiovascular Imag.*, vol. 15, no. 7, pp. 1193–1208, Jul. 2022, doi: [10.1016/j.jcmg.2022.02.008](https://doi.org/10.1016/j.jcmg.2022.02.008).
- [21] S. Dencks et al., "Clinical pilot application of super-resolution U.S. imaging in breast cancer," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 66, no. 3, pp. 517–526, Mar. 2019, doi: [10.1109/TUFFC.2018.2872067](https://doi.org/10.1109/TUFFC.2018.2872067).
- [22] S. Harput et al., "Two-stage motion correction for super-resolution ultrasound imaging in human lower limb," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 65, no. 5, pp. 803–814, May 2018, doi: [10.1109/TUFFC.2018.2824846](https://doi.org/10.1109/TUFFC.2018.2824846).
- [23] C. Deme   et al., "Transcranial ultrafast ultrasound localization microscopy of brain vasculature in patients," *Nature Biomed. Eng.*, vol. 5, no. 3, pp. 219–228, Mar. 2021, doi: [10.1038/s41551-021-00697-x](https://doi.org/10.1038/s41551-021-00697-x).
- [24] 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, Apr. 2021, Art. no. 08NT01, doi: [10.1088/1361-6560/abef45](https://doi.org/10.1088/1361-6560/abef45).
- [25] J. Zhu et al., "Super-resolution ultrasound localization microscopy of microvascular structure and flow for distinguishing metastatic lymph nodes—An initial human study," *Ultraschall der Medizin Eur. J. Ultrasound*, vol. 43, no. 6, pp. 592–598, Dec. 2022, doi: [10.1055/a-1917-0016](https://doi.org/10.1055/a-1917-0016).
- [26] J. Yan et al., "Transthoracic super-resolution ultrasound localization microscopy of myocardial vasculature in patients," 2023, *arXiv:2303.14003*.
- [27] 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, doi: [10.1109/TUFFC.2015.2500266](https://doi.org/10.1109/TUFFC.2015.2500266).
- [28] I. Taghavi et al., "Ultrasound super-resolution imaging with a hierarchical Kalman tracker," *Ultrasonics*, vol. 122, May 2022, Art. no. 106695, doi: [10.1016/j.ultras.2022.106695](https://doi.org/10.1016/j.ultras.2022.106695).
- [29] J. Yan, T. Zhang, J. Broughton-Venner, P. Huang, and M. Tang, "Super-resolution ultrasound through sparsity-based deconvolution and multi-feature tracking," *IEEE Trans. Med. Imag.*, vol. 41, no. 8, pp. 1938–1947, Aug. 2022, doi: [10.1109/TMI.2022.3152396](https://doi.org/10.1109/TMI.2022.3152396).
- [30] C. Huang et al., "Short acquisition time super-resolution ultrasound microvessel imaging via microbubble separation," *Sci. Rep.*, vol. 10, no. 1, Apr. 2020, Art. no. 1, doi: [10.1038/s41598-020-62898-9](https://doi.org/10.1038/s41598-020-62898-9).
- [31] K. Riemer et al., "Fast and selective super-resolution ultrasound in vivo with acoustically activated nanodroplets," *IEEE Trans. Med. Imag.*, vol. 42, no. 4, pp. 1056–1067, Apr. 2023, doi: [10.1109/TMI.2022.3223554](https://doi.org/10.1109/TMI.2022.3223554).
- [32] 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, doi: [10.1109/TUFFC.2018.2873380](https://doi.org/10.1109/TUFFC.2018.2873380).
- [33] R. J. G. van Sloun et al., "Super-resolution ultrasound localization microscopy through deep learning," *IEEE Trans. Med. Imag.*, vol. 40, no. 3, pp. 829–839, Mar. 2021, doi: [10.1109/TMI.2020.3037790](https://doi.org/10.1109/TMI.2020.3037790).
- [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, doi: [10.1109/TMI.2020.3006445](https://doi.org/10.1109/TMI.2020.3006445).
- [35] 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, doi: [10.1109/TMI.2021.3056951](https://doi.org/10.1109/TMI.2021.3056951).
- [36] B. Heiles, A. Chavignon, V. Hingot, P. Lopez, E. Teston, and O. Couture, "Performance benchmarking of microbubble-localization algorithms for ultrasound localization microscopy," *Nature Biomed. Eng.*, vol. 6, no. 5, pp. 605–616, Feb. 2022, doi: [10.1038/s41551-021-00824-8](https://doi.org/10.1038/s41551-021-00824-8).
- [37] M. Lerendegui, K. Riemer, B. Wang, C. Dunsby, and M.-X. Tang, "BUbble flow field: A simulation framework for evaluating ultrasound localization microscopy algorithms," 2022, *arXiv:2211.00754*.
- [38] J. A. Jensen, "A model for the propagation and scattering of ultrasound in tissue," *J. Acoust. Soc. Amer.*, vol. 89, no. 1, pp. 182–190, Jan. 1991, doi: [10.1121/1.400497](https://doi.org/10.1121/1.400497).
- [39] 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, Mar. 1992, doi: [10.1109/58.139123](https://doi.org/10.1109/58.139123).
- [40] J. A. Jensen, "Field: A program for simulating ultrasound systems: 10th Nordic-Baltic conference on biomedical imaging," *Med. Biol. Eng. Comput.*, vol. 34, no. 1, pp. 351–353, 1997.
- [41] T. S. W. Stevens et al., "A hybrid deep learning pipeline for improved ultrasound localization microscopy," in *Proc. IEEE Int. Ultrason. Symp. (IUS)*, Oct. 2022, pp. 1–4, doi: [10.1109/IUS54386.2022.9958562](https://doi.org/10.1109/IUS54386.2022.9958562).
- [42] G. Revach, N. Shlezinger, X. Ni, A. L  pez Escoriza, R. J. G. van Sloun, and Y. C. Eldar, "KalmanNet: Neural network aided Kalman filtering for partially known dynamics," *IEEE Trans. Signal Process.*, vol. 70, pp. 1532–1547, 2022, doi: [10.1109/TSP.2022.3158588](https://doi.org/10.1109/TSP.2022.3158588).
- [43] H. W. Kuhn, "The Hungarian method for the assignment problem," *Nav. Res. Logistics (NRL)*, vol. 52, no. 1, pp. 7–21, Feb. 2005, doi: [10.1002/nav.20053](https://doi.org/10.1002/nav.20053).
- [44] M. Piepenbrock, S. Dencks, and G. Schmitz, "Microbubble tracking with a nonlinear motion model," in *Proc. IEEE Int. Ultrason. Symp. (IUS)*, Las Vegas, NV, USA, Sep. 2020, pp. 1–4, doi: [10.1109/IUS46767.2020.9251581](https://doi.org/10.1109/IUS46767.2020.9251581).
- [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, doi: [10.1109/TUFFC.2017.2778941](https://doi.org/10.1109/TUFFC.2017.2778941).
- [46] S. Tang, "Kalman filter-based microbubble tracking for robust super-resolution ultrasound microvessel imaging," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 67, no. 9, pp. 1738–1751, Sep. 2020, doi: [10.1109/TUFFC.2020.2984384](https://doi.org/10.1109/TUFFC.2020.2984384).
- [47] S. Schoen, Z. Zhao, A. Alva, C. Huang, S. Chen, and C. Arvanitis, "Morphological reconstruction improves microvessel mapping in super-resolution ultrasound," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 68, no. 6, pp. 2141–2149, Jun. 2021, doi: [10.1109/TUFFC.2021.3057540](https://doi.org/10.1109/TUFFC.2021.3057540).
- [48] S. Schoen, A. E. Samir, and V. Kumar, "MR for ULTRA-SR: Improved localization with morphological image processing," in *Proc. IEEE Int. Ultrason. Symp. (IUS)*, Oct. 2022, pp. 1–4, doi: [10.1109/IUS54386.2022.9957276](https://doi.org/10.1109/IUS54386.2022.9957276).
- [49] C. Deme  , "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, doi: [10.1109/TMI.2015.2428634](https://doi.org/10.1109/TMI.2015.2428634).
- [50] L. Vincent, "Morphological grayscale reconstruction in image analysis: Applications and efficient algorithms," *IEEE Trans. Image Process.*, vol. 2, no. 2, pp. 176–201, Apr. 1993, doi: [10.1109/83.217222](https://doi.org/10.1109/83.217222).
- [51] J.-Y. Tinevez. (Nov. 17, 2022). *Simpletracker—A Simple Particle Tracking Algorithm for MATLAB That Can Deal With Gaps*. Accessed: Apr. 4, 2023. [Online]. Available: <https://github.com/tinevez/simpletracker>
- [52] F. Bourgeois and J.-C. Lassalle, "An extension of the Munkres algorithm for the assignment problem to rectangular matrices," *Commun. ACM*, vol. 14, no. 12, pp. 802–804, Dec. 1971, doi: [10.1145/362919.362945](https://doi.org/10.1145/362919.362945).
- [53] M. Ashikuzzaman, B. Helfield, and H. Rivaz, "Analytic optimization-based microbubble tracking in ultrasound super-resolution microscopy," in *Proc. IEEE Int. Ultrason. Symp. (IUS)*, Oct. 2022, pp. 1–4, doi: [10.1109/IUS54386.2022.9957900](https://doi.org/10.1109/IUS54386.2022.9957900).
- [54] V. Hingot, A. Chavignon, B. Heiles, and O. Couture, "Measuring image resolution in ultrasound localization microscopy," *IEEE Trans. Med. Imag.*, vol. 40, no. 12, pp. 3812–3819, Dec. 2021, doi: [10.1109/TMI.2021.3097150](https://doi.org/10.1109/TMI.2021.3097150).
- [55] 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, doi: [10.1109/TUFFC.2018.2832600](https://doi.org/10.1109/TUFFC.2018.2832600).
- [56] Y. Shin et al., "Context-aware deep learning enables high-efficacy localization of high concentration microbubbles for super-resolution ultrasound localization microscopy," *Nature Commun.*, vol. 15, no. 1, p. 2023, Apr. 23, 2024, doi: [10.1038/s41467-024-47154-2](https://doi.org/10.1038/s41467-024-47154-2).
- [57] J. Sobolewski, S. Dencks, and G. Schmitz, "Influence of image discretization and patch size on ULM localization precision," in *Proc. IEEE Int. Ultrason. Symp. (IUS)*, Sep. 2023, pp. 1–3, doi: [10.1109/IUS51837.2023.10306429](https://doi.org/10.1109/IUS51837.2023.10306429).
- [58] B. Huang, J. Yan, M. Morris, V. Sinnett, N. Somaiah, and M.-X. Tang, "Acceleration-based Kalman tracking for super-resolution ultrasound imaging in vivo," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 70, no. 12, pp. 1739–1748, Dec. 2023, doi: [10.1109/TUFFC.2023.3326863](https://doi.org/10.1109/TUFFC.2023.3326863).