# Performance benchmarking of microbubble-localization algorithms for ultrasound localization microscopy

Baptiste Heiles, Arthur Chavignon, Vincent Hingot, Pauline Lopez, Eliott Teston, Olivier Couture

# Open Platform for Ultrasound Localization Microscopy: performance assessment of localization algorithms

Baptiste Heiles[1,2,†*], Arthur Chavignon[1, †], Vincent Hingot[1,2], Pauline Lopez[2,3], Eliott Teston[2], Olivier Couture[1]

[1] Sorbonne Université, CNRS, INSERM, Laboratoire d'Imagerie Biomédicale, Paris, France

[2] ESPCI, CNRS, INSERM, PhysMedParis, Paris, France

[3] Institut Cochin, INSERM U1016, Paris, France

[†] These authors contributed equally

* Corresponding author

## Abstract

Ultrasound Localization Microscopy (ULM) is an ultrasound imaging technique that relies on the acoustic response of sub-wavelength ultrasound scatterers to map the microcirculation with an order of magnitude increase in resolution as compared to conventional ultrasound imaging. Initially demonstrated *in vitro*, ULM has matured and sees implementation *in vivo* for vascular imaging of organs or tumors in both animal models and humans. The performance of localization algorithms greatly defines the quality of vascular mapping. Here, we compiled and implemented a collection of ultrasound localization algorithms and devised three *in silico* and *in vivo* datasets to compare their performance through 11 metrics. We also present two novel algorithms designed to increase speed and overall performance. By providing a comprehensive open package to perform ULM that includes localization algorithms, the datasets used, and the evaluation metrics, we aim to equip researchers with a tool to identify the optimal localization algorithm for their

application, benchmark their own software and enhance the overall quality of their ULM images while uncovering the algorithms' own limits.

## Main

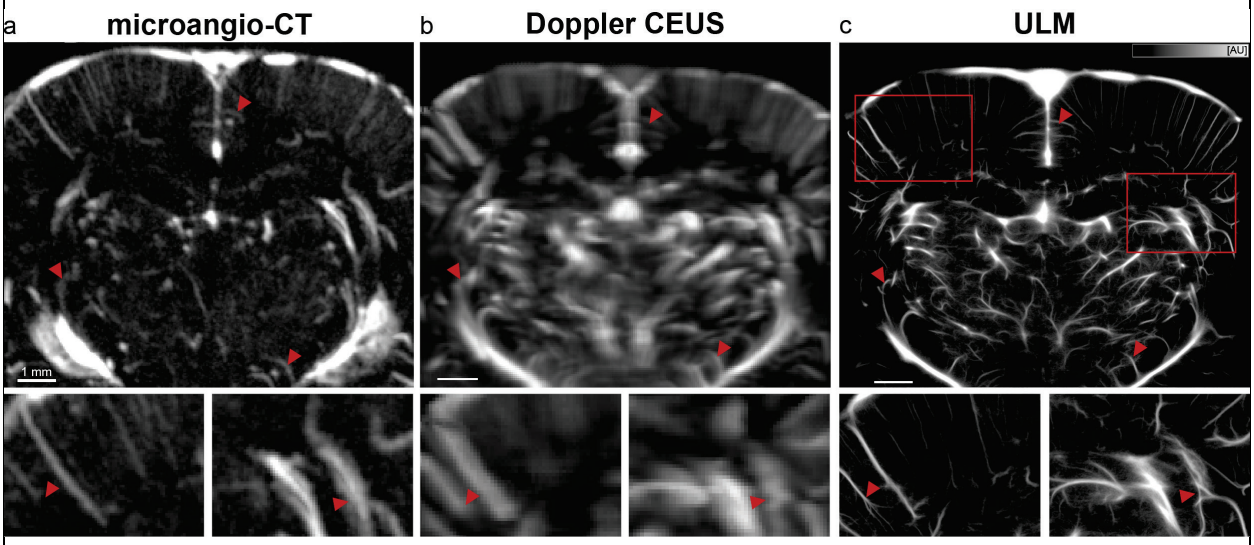
The circulatory system carries the essential nutrients of life to cells in the body. It forms a 100,000 kilometer-long network composed of centimeter-wide arteries down to capillaries that are a few micrometers in diameter at most. The study of the vascular system is essential for both the diagnosis and treatment of cardiovascular diseases, cancer, diabetes, stroke, or organ dysfunction. Due to its diversity of scale, imaging the vasculature is a daunting task and few techniques are capable of measuring micro-hemodynamics deep in the human body.

Ultrasound imaging is extensively used in medical practice as a non-invasive tool that provides soft-tissue diagnosis, prognosis, or guides interventions. Using the Doppler effect, it can also measure blood flow in real-time. With the advent of plane wave techniques[1], ultrasound has reached frame rates up to 20 kHz making it possible to observe and measure fast occurring changes[1] such as functional changes in the brain[2], as well as increase sensitivity to blood flow that allows more accurate filtering[3,4]. To increase blood's contrast, micrometric gas microbubbles can be intravenously injected *in vivo*. Contrast-Enhanced Ultrasound (CEUS) is mostly used for perfusion studies and cardiac imaging[5].

Because conventional ultrasound imaging, Doppler, and contrast-enhanced ultrasound all rely on the propagation of sound waves, ultrasound imaging is largely limited in resolution by diffraction. Recently, Ultrasound Localization Microscopy (ULM) has broken that limit by isolating a small number of microbubbles as subwavelength sources in each image and localizing them with micrometric precision[6–8]. Similar to PALM (PhotoActivated Localization Microscopy), it uses

these agents as individual sensors to map the region of imaging by accumulating thousands of events and adding them through density-based methods. When implemented at kilohertz frame rates, ULM can also retrieve the trajectories of microbubbles carried in the bloodstream and reconstruct several centimeters deep, hemodynamic maps of the vasculature both in 2D and 3D[9–11], a feat rising to the challenge of multiscale imaging tools for both the vasculature and microvasculature. For example, typical imaging of a rat brain with craniectomy *in vivo* with Doppler contrast-enhanced ultrasound and ULM is shown in **Figure 1** next to a *post mortem* microangio-computed tomography (microangio-CT) for reference. Microvessels appear on each modality but sharper and brighter on the ULM rendering (red triangles on **Figure 1**).

**Figure 1: Rat brain vascular imaging with *in vivo* ULM and *post mortem* microangio-CT**




a) *Post mortem* microangio-CT with perfusion of µAngiofil (Fumedica AG, Switzerland) of the rat brain performed with a preclinical micro-CT. Slice of 400 µ*m*. Red arrows indicate corresponding vessels in the different images,
b) *In vivo* Doppler contrast-enhanced ultrasound (CEUS) of the rat brain,
c) *In vivo* ULM of the rat brain with 200,000 images at 1 kHz, red squares indicate where the zooms below each image are taken

Several teams have been implementing ULM using various localization techniques over the last few years[6,10,12–14]. The quality of ULM images is directly dependent on the signal-to-noise ratio of the original images or detection filters but most importantly on the localization and rendering techniques. However, because each team uses different *in vitro* and *in vivo* models as well as different probes, ultrasound machines, and acquisition parameters, comparing these algorithms, along with their general approaches, is difficult. Progress in the field of ultrasound super-resolution is hindered by a lack of widely accepted comparison metrics, even though several attempts have been published on filtering[15], localization[16–18], and tracking[15] errors.

In this paper, we present a comprehensive open data-processing framework for ULM, including six datasets, *in vitro* and *in vivo* – available at http://doi.org/10.5281/zenodo.4343435 - complemented by a series of universal performance metrics to evaluate quantitatively different ULM implementations. We provide five of the most used algorithms for microbubble localization to compare their performances: Lanczos, spline and cubic interpolation (Lz-Interp, Sp-Interp, Cub-Interp), Gaussian fitting of the PSF (Gauss-Fit), and a no-shift method that assumes ideal centering of the microbubble in the pixel. The scripts along with our Localization and tracking Toolbox for Ultrasound Super-resolution (LOTUS), are available on a GitHub repository (https://github.com/AChavignon/PALA). In the table below, we have compiled the localization methods used and their related articles as well as a few comments.

| Localization method | Related articles | Comments |
|---|---|---|
| Gaussian fitting | (Ackermann and Schmitz, 2016; Luke et al., 2016; O'Reilly and Hynynen, 2013; Song et al., 2018b)[19–22] | Usually, these works use a Gaussian convolved PSF rather than a Gaussian fitting with an optimizer. |
| Weighted average based | (Christensen-Jeffries et al., 2015; Hansen et al., 2016; Heiles et al., 2019; Lin et al., 2017; Song et al., 2018a; Soulioti et al., 2018; Viessmann et al., 2013; Zhang et al., 2018; Zhu et al., 2019)[10,13,15,23–28] | Except for Heiles et al 2019, these work on data beamformed with pixels of sizes below the wavelength or data beamformed with commercial scanners which might affect PSF shape and full width at half maximum. |
| Lanczos based interpolation and Gaussian fitting | (Errico et al, 2015)[9] | |
| Spline based interpolation | (Huang et al., 2020; Song et al., 2018b)[22,29] | |
| Linear based interpolation | (Song et al., 2018a)[15] | On top of the linear-based interpolation, this paper convolves with a Gaussian profile. |
| Cubic based interpolation | (Song et al., 2018b) [22] | This paper is a comparison of algorithms |
| Radial symmetry | (Parthasarathy, 2012)[30] (optic super-resolution only) | |
| RF-based | (Brown et al., 2019; Christensen-Jeffries et al., 2017a, 2017b; Desailly et al., 2013, 2015) [16,17,31–33] | These papers are based on radiofrequency data before beamforming. In particular, the papers from the team at Imperial College/Kings College London use the onset of the Hilbert transform of the RF signals. |

In addition, we introduce two algorithms developed by our team called weighted average (WA) and radial symmetry (RS). We provide all algorithms ready for use so that labs lacking the expertise to implement ULM can contribute to the fast development of ultrasound super-resolution imaging (see supplemental data). Importantly, we introduce metrics that address still unanswered questions such as: how accurate and sensitive is the localization algorithm[17], how much is ULM affected by low SNR[13], what affects the precision of the code, how well is it possible to reconstruct difficult structures such as highly tortuous vessels, aneurysms, stenosis, where do the grid artifacts come from[34], and finally how can 3D ULM be implemented without a dramatic increase in

computational power and time? The platform is designed to allow the integration of new metrics and answers to questions the user might come up with on its own.

## Results

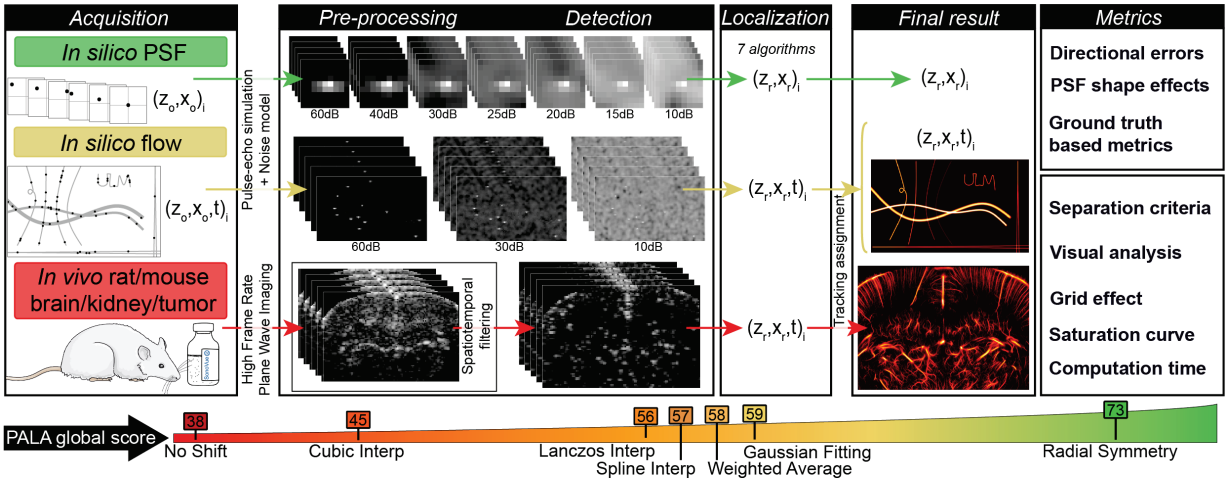
Ultrasound localization microscopy is a multi-step process comprising the acquisition of a large number of frames, pre-processing, detection, localization, and image reconstruction as described in **Figure 2** and ***Supplementary Figure 1-1***. Although this article is mostly focused on the comparison of localization algorithms, readers will find, in supplementary material, the scripts and appropriate open software (LOTUS) to perform each of these steps in the most straightforward way. These scripts can be easily adapted to different acquisition types or various computer setups and parallelized. In LOTUS, users can load their own acquired data, filter it, localize microbubbles, track them and generate ultrasound super-resolution renderings.

### *In silico and in vivo datasets*

To study the performances of different localization algorithms, the first step was to create and acquire 3 different datasets, each of which we make fully available for reuse in supplementary materials. The first one ("*in silico* PSF"), the point-spread function (PSF) simulation, is designed to study non-uniformity effects on localization and spatial sampling effect on the beamforming process. It consists of a point-like scatterer moved in a $\lambda \times \lambda$ space, the size of one pixel, by $\frac{\lambda}{21}$ steps, resulting in 441 positions. The second dataset ("*in silico* Flow"), the microcirculation simulation, allows for testing *in silico* more fundamental aspects of resolution such as the separability of microvessels and the accuracy of microbubble tracking. It is comprised of complex 3D tubes in the imaging plane mimicking an actual 2D imaging situation. Inside, point-like scatterers are placed at random in each of the tubes section and then are propagated through 20,000

frames according to a Poiseuille flow model assuming continuity on streamlines. Concentrations, speeds, and noise level can all be modified by the user and are presented in the online methods section. The maximum velocities in each of these tubes range from $0.05\lambda/frame$ to $3\lambda/frame$. We studied the localization errors and the effect of different signal-to-noise ratios (SNR) using this dataset. For these datasets, the point-like scatterer's ultrasound response is simulated using the Verasonics Research Ultrasound Simulator (Verasonics Inc., Kirkland Washington, USA), using a 128 elements linear probe, with elements' pitch at a wavelength ($\lambda = 100\ \mu m$).

**Figure 2: Description of the framework for data simulation and performance assessment**




The framework presented here fits easily into the classical layout of ULM acquisition, filtering/pre-processing, detection of microbubbles, localization of events' centers, tracking, and mapping of the flow. Each dataset gives different information about the localization process. The separation criteria, which defines maximum resolution attainable is given by dataset "*in silico* PSF". The ground truth based metrics are given by *in silico* datasets. *In vivo* datasets, "*in vivo* rat brain perfusion", "*in vivo* rat brain bolus", "*in vivo* rat kidney", "*in vivo* mouse tumor", give information about the *in vivo* visual aspect, the influence of the algorithms on the grid effect and on computation time. Visual analysis is also provided by dataset "*in silico* Flow". The PALA global score is calculated based on a combination of the metrics on the right column computed on *in vivo* and *in silico* datasets. By summing and weighting each of these metrics derived scores, we obtain the global score, designed to grade the algorithms' performance.

While the simulation-based datasets give access to the ground truth and are useful to compare the error of localization-based metrics, the best validation is still *in vivo* experiments as we are

confronted with multi-factorial signal noise, breathing motion, speed of sound changes coming from propagation through fatty tissue, pulsating blood and other phenomena arduous to reproduce in simulation. Imaging different organs allow us to compare different physiologies, hemodynamics, and anatomical geometries. The *in vivo* datasets comprise of:

- *"in vivo* rat brain perfusion", "*in vivo* rat brain bolus": two Sprague Dawley rat brains at different coronal sections. One was acquired while continuously injecting microbubbles through the jugular vein (192,000 ultrasound images at a SNR of 29 dB), another one by injecting a bolus of microbubbles in the caudal vein (170,000 frames at a SNR of 28 dB).
- *"in vivo* rat kidney": a Sprague Dawley rat kidney (188,020 frames at a SNR of 26 dB)
- *"in vivo* mouse tumor": A subcutaneous tumor implanted in an FvB strain mouse (30,000 frames at a SNR of 25 dB)

The *in vivo* angiography of the rat brain with continuous injection is named "*in vivo* rat brain", which is the most complete dataset and one that has already been used as reference [8,35]. The skull of the rat was previously removed following craniotomy surgery protocol (detailed in the Online methods section). The other rat brain we imaged went through the same protocol with a different method for injection. The rat kidney dataset was acquired after isolation of the kidney outside of the rat following lumbotomy protocol (detailed in Online methods). The tumor dataset was obtained on a subcutaneous tumor in mice after implantation of primary cancerous cells developed in mouse mammary tumor virus-polyoma middle tumor-antigen (MMTV-PyMT) donor mice (detailed in Online methods). The kidney and mouse tumor datasets are more subject to breathing motion than the rat brain since the animal is not fixed. These datasets will thus allow us to compare the performance of the algorithms in low SNR, moderate motion situations. The mouse tumor which is expected to have a disorganized vasculature is of particular interest.

On each of these datasets, and based on the state of the art we selected and applied 7 different localization algorithms to retrieve subwavelength positions. They are then processed by a Kuhn-

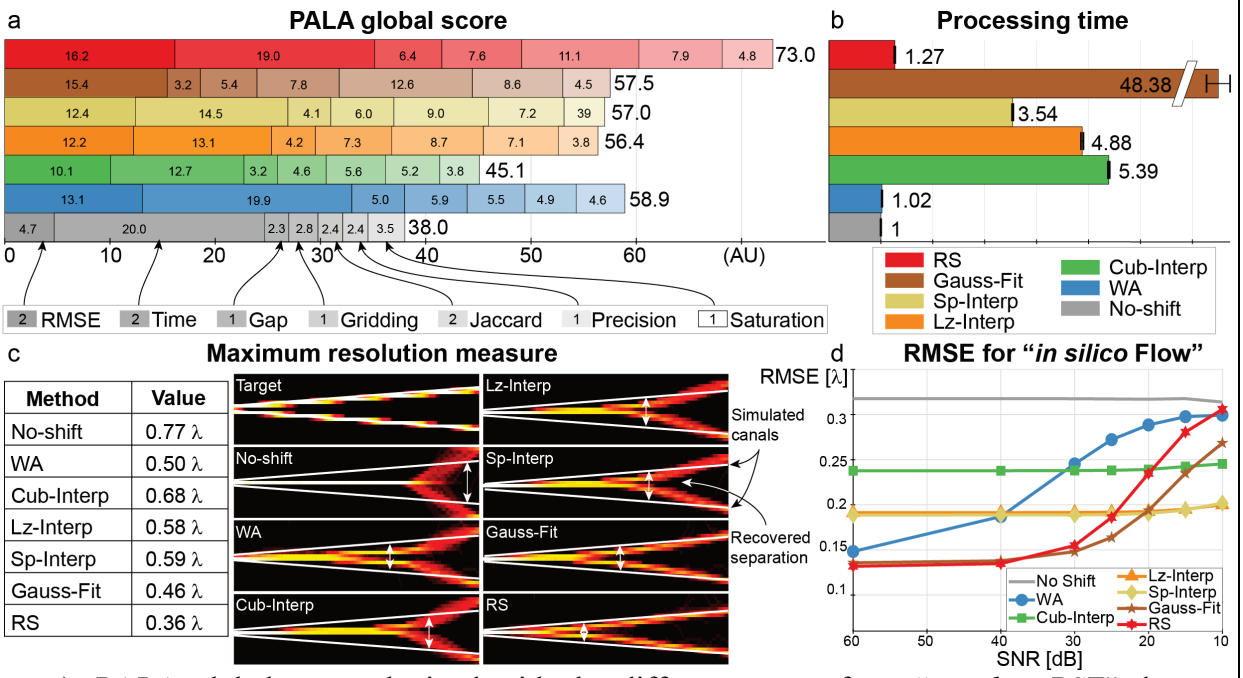
Munkres assignment to perform tracking and velocimetry[36,37]. Final images are reconstructed by density projection or velocity renderings. The whole framework is presented in **Figure 2**.

*General comparisons of 7 algorithms using 11 metrics*

To alleviate the current difficulty in comparing the attainable resolution in different organs and by different groups, we report here 11 different metrics to compare various ULM algorithms. In each simulated dataset, we look at the directional errors, statistical measures using a binary classification test, and separation criteria to define maximum attainable resolution. For *in vivo* results, we report an in-depth visual analysis as well as grid effect characterization, and computation time. In the end, the localization algorithms are rated with an overall scoring called PALA global score that takes into account errors, the ability to detect microbubbles accurately, and computation time, computed on *in silico* and *in vivo* datasets (see **Online methods** and ***Supplementary Figure 5-1 to 5-4***). The results are presented in **Figure 3**.

Two algorithms, Gaussian fitting, and radial symmetry come out on top on almost every quantitative index that we have devised. However, when looking at the time it takes to perform ULM with tracking and localization, the ranking is modified. The weighted average and the radial symmetry-based algorithms are the fastest localization implementations, while the Gaussian fitting is almost 50 times slower. The values presented in **Figure 3** b were normalized according to the fastest implementation which takes for the *in vivo* dataset around 3 minutes, faster than the actual acquisition of the images.

**Figure 3: Results of the study combined in the PALA global score for each algorithm**




a) PALA global score obtained with the different scores form "*in silico* PSF" dataset: RMSE, "*in silico* flow" dataset: minimum separation measured between the two canals (i.e. maximum resolution), Jaccard index, precision value, and "*in vivo* rat brain perfusion": the time factor, gridding index and saturation index based on saturation curves (see **Figure 7**). The bigger the score, the higher is the performance of the algorithm. To reflect the importance of the different indices, some of them were multiplied by coefficients (value in legend). These arbitrary coefficients allow us to double the impact of RMSE, time factor, and Jaccard index, the three indices we consider most critical. The weights used as well as the exact formulae for the conversion of the metrics to arbitrary unit scores on a scale of 0 to 100 are available in the online methods and a supplementary dynamic table "PALA_GlobalScores_DynamicTable.xlsx" Different weights have also been chosen to reflect implementations where time or SNR influence might be less crucial in the supplementary information.

b) Computation time for each algorithm compared to the fastest implementation, the no-shift implementation taking 3 minutes on our high-end computer (Intel Core i9 @ 2.9 GHz 12cores, NVidia RTX 2080Ti, 128GB RAM @ 2133 MHz).

c) Maximum resolution attainable on the dataset "*in silico* Flow" at $SNR = 30\ dB$ with different localization schemes. For each rendering, the point where the intensities of the two canals reconstructed are distinct by more than 3% is calculated. The lateral position at which it is located is taken to calculate the real distance between the two simulated canals. That is taken as the minimum separation that can be accurately recovered by the algorithm. It is the separation index.

d) Root mean square error of the localization schemes for recovery of simulated points in dataset "*in silico* Flow" with different SNR.

Overall the radial symmetry (RS) based and weighted average (WA) algorithms showed the highest performance with the weights chosen. Looking closer, the maximum resolution attainable measured as the separation criteria is the highest for the RS and Gauss-Fit based algorithms at $0.36\ \lambda$ and $0.46\ \lambda$. The time factor, however, is the lowest for the WA with under 3 minutes of calculation which is lower than the acquisition time with our computer. The large increase and very low root mean square error (RMSE) minimum of the weighted average algorithm ($0.13\ \lambda$) confirm that it is a very good candidate for localization at a high signal-to-noise ratio (SNR) but that one should switch to either Gauss-Fit or RS when the SNR worsens below $30\ dB$. At $15\ dB$ and downwards, one should theoretically consider using interpolation-based algorithms (see ***Supplementary Figures 3-3)***. The gridding index, which is a quantitative measure of the reprojection of localized microbubbles on the original beamformed grid, impairing accurate vessel delineation, is maximized for both RS and Gauss-Fit approaches. This gridding index increases when specific spatial frequencies are overexpressed.

*Accuracy metrics in simulated datasets*

We simulated both *in silico* datasets with 7 different signal-to-noise ratios from $10\ dB$ to $60\ dB$ (white Gaussian noise model - 0.2 ohm impedance noise, ± 10 dB amplitude - for clutter filtered with 0.7 pixel wide 2D Gaussian kernel, see ***Supplementary figure 2-3***). The different localizations in the 30 dB set were classified depending on their distance to the scatterer simulated position. When this distance is lower than $\frac{\lambda}{4}$, the localization is deemed successful, otherwise, it is either classified as a false negative (FN) if an existing scatterer was not detected or as a false positive if no real scatterer was present (FP) (see ***Figure 4*** **a**, and ***Supplementary Figure 3-6***). The highest number of localizations 386,000, calculated as the total of false and true positives is obtained in the radial symmetry case, with 12% more than the worst-rated algorithm, the Cub-

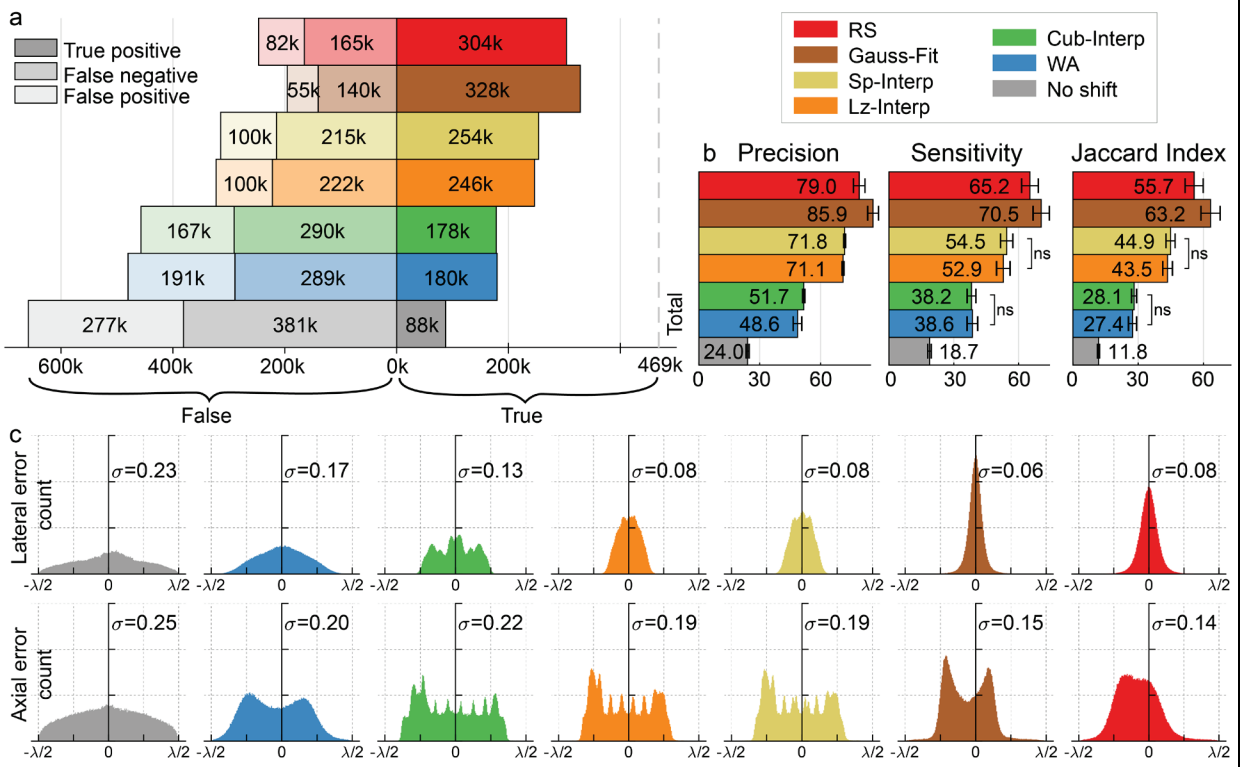
Interp. The number of true positives is the highest in the Gauss-Fit/RS-based algorithms with a value of 3.7 and 3.5 times higher respectively than the no-shift algorithm. The interpolation-based algorithms have similar TP values while the cubic interpolation fares the worst. The same trends in FP and FN numbers can be observed. To interpret these more easily, the precision, Jaccard, and sensitivity indices were also plotted (**Figure 4** b). For the three indices, the Gauss-Fit and RS come out on top with as much as 85.9% simulated scatterers being picked up by the best localization algorithm.

The sensitivities of WA and the best interpolation-based algorithms are quite low at around 40 %, while the precision is higher for the latter at 71.8 % (see **Figure 4** b). The no-shift algorithm has the lowest values albeit not as low as one would expect given that no localization is involved. The Jaccard index which represents a detection rate is at most 63.2 % and can be as low as 11.8 % for the no-shift scheme. Gauss-Fit and RS are more precise than others. This is due to a low localization error, and as such a high cardinality of true positives. As the total number of $\{TP + FN\}$ is finite, this also means that the sensitivity index is high. For the weighted average scheme, we measure a high precision and high sensitivity although its localization error is the highest among the schemes. It gives us a glimpse into why this happens: the TP and FN values are almost identical to the best interpolation schemes.

The directional errors are calculated for every simulated point and every recovered localization (**Figure 4** c). At first look, WA and Cub-Interp localization algorithms have similar error values and standard deviations. However, their error distributions are very different and we can uncover further insight about localization behavior by closely studying them. The distribution axial error of the latter has $\frac{\lambda}{10}$ peaks compared to the others' smooth distribution. That points towards jumps in the localizations. We hypothesize that this will cause a significant increase in the number of

false negatives and a decrease in the number of true positives which can be verified in **Figure 4** a, $TP_{WA} = 180k, TP_{Cub-interp} = 178k$ and $FN_{WA} = 289k, FN_{Cub-Interp} = 290k$

**Figure 4: Directional errors and statistical indices for performance on the dataset "*in silico* Flow" at 30 dB**




a) Statistical measures of the different localization schemes for dataset "*in silico* Flow". The left direction indicates the false ensemble, the right, the true ensemble. The total number of points simulated is 469,000.

b) Statistical indices based on statistical measures for the different localization schemes in percent. Means and standard deviations are based on 20 blocks of 1000 continuous frames. Unpaired two-sample t-test of the means of the different algorithms always returns p-values lower than 5% except when "ns" is specified (non-significant, p>0.05).

c) Lateral and axial error distributions for the different localization schemes. Some of these distributions have skew for example the spline and Lanczos interpolation schemes in the lateral direction.

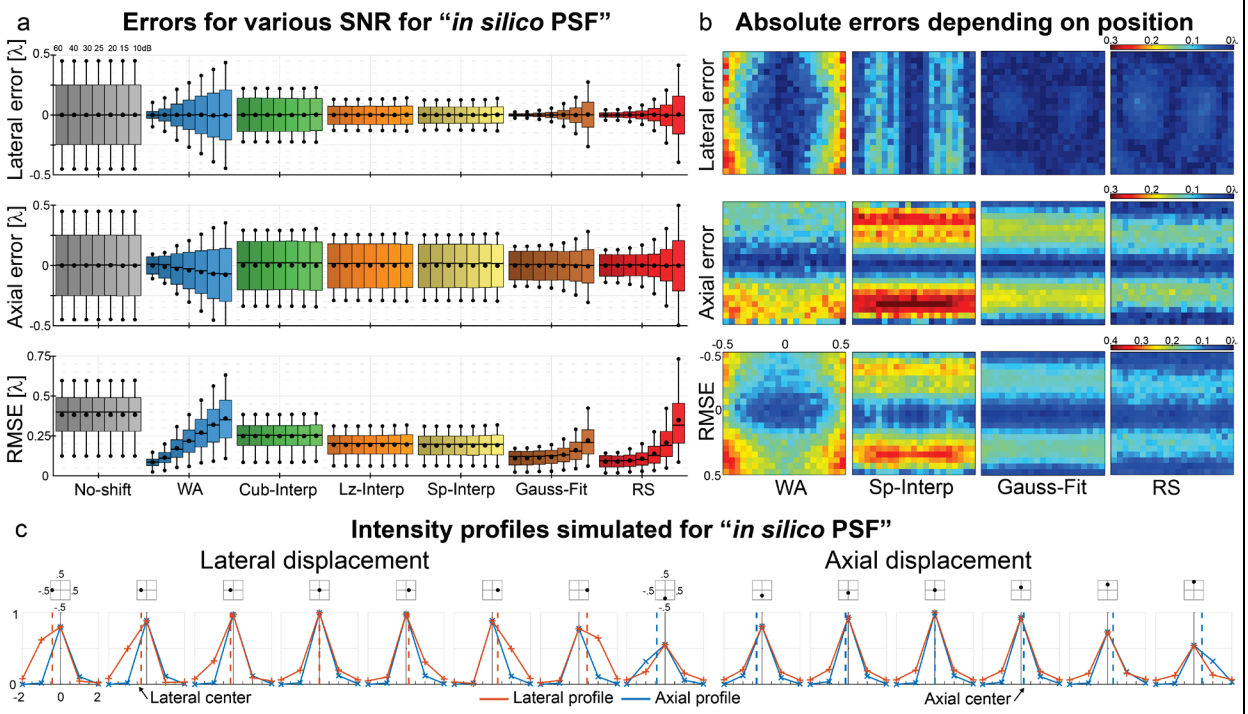
The largest standard deviation of the lateral error was found for the no-shift method as expected. For the axial error, the algorithms have standard deviations between 0.06 and 0.23, with the lowest being the GF and the highest being the no shift. For the Gauss-Fit and RS-based algorithm, their

lateral standard deviation error is around half lower than their axial error. The Lanczos and Spline interpolations schemes have similar ratios between the lateral and axial error standard deviation. The histogram distributions in the lateral direction for these algorithms seem to follow a Power-law with a high exponent whereas the weighted average, Lanczos, and spline kernel-based interpolation follow a linear law. The axial error distributions all present sharp peaks in different patterns. The most notable pattern is the fork pattern at $\frac{\lambda}{10}$ on the interpolation-based methods reproducing the interpolation grid used. Another pattern present in the radial symmetry and Gaussian fitting-based methods takes the shape of two sharp peaks on either side of the center value showing a bimodal behavior where the microbubble is localized in two preferred locations outside of the actual center. The RMSE is highly affected by this difference in localization in both directions and while the RMSE of the radial symmetry and Gaussian fitting-based schemes are better than the others, the improvement in localization brought by these algorithms in the lateral direction is diminished in the final value of the RMSE. These errors were calculated for seven different SNR and are presented in ***Supplementary Figures 3-3***.

The various algorithms spatially shift the expected localization in different ways. To study this effect more precisely, it is useful to look at the results of the PSF simulation. The results are presented in **Figure 5**. The two main results are visible at once: first, the interpolation-based algorithms are scarcely affected by noise as they only take into account the closest neighboring sample to build the final function, and second, they exhibit a quite high standard deviation of their errors in the axial direction ($sd\left(err_{axspline}\right) = 0.19\ \lambda$, $sd(err_{axcubic})) = 0.22\ \lambda$). Putting it into perspective when plotting the absolute error at each of the $\frac{\lambda}{21}$ grid points helps to understand the dynamics of the noise impact (**Figure 5** b). When the SNR diminishes, the errors of the best-ranked algorithms tend to show the bimodal behavior of the worst-ranked: axial and lateral errors are

distributed in two bands in each direction up until a certain SNR value. At 15 $dB$ the weighted average exhibits this behavior very clearly for the lateral error while the GF and RS have an almost uniform distribution. All distributions and maps are presented in ***Supplementary Figures 2-5 to 2-6 and 3-3.*** If we look more closely at the intensity profiles generated by our framework (**Figure 5** c), we can see that displacement in any direction induces a displacement of the maximum amplitude.

**Figure 5: Study of PSF non-uniformity in direction and its effect on localization**




a) Lateral, axial and root mean square errors averages with standard deviations measured for the different localization schemes on the dataset "*in silico* PSF" with different SNR. Errors are standard deviations. Bold center lines show medians, upper and lower box limits show the upper and lower quartiles, whiskers represent the first 5% and 95% quantiles. Round markers show mean values.

b) Absolute lateral, axial, and root mean square errors for each position where a scatterer was simulated inside the $\lambda \times \lambda$ grid with a SNR of 30 $dB$.

c) $5\lambda \times 5\lambda$ PSF simulated for a scatterer displaced in the $\lambda \times \lambda$ grid with $\lambda/7$ step. The intensity profiles demonstrate the inability of the wavelength-driven system to accurately reflect sub-wavelength displacement. The dashed lines represent the true position of the scatterer simulated

However, even when moving close to the neighboring pixel situated $\left(\frac{\lambda}{5}\right)$ further from the center, the maximum amplitude remains at the center of the pixel $\left((x, z) = 0\right)$. The maximum drop in intensity for each displacement is $12\ dB$ below the peak amplitude on the connected components of the central pixel. This means that the bimodal behavior observed above in the axial error is due to low SNR that prevents the algorithms to use the information present in the neighboring pixels to perform accurate localization. This confirms that the interpolation schemes don't vary with SNR as much as the other schemes as they only use 3 points to perform the interpolation. At SNR lower than $15\ dB$, localization might fail in the WA and RS schemes at giving accurate enough results and providing enhanced resolution in the axial direction.

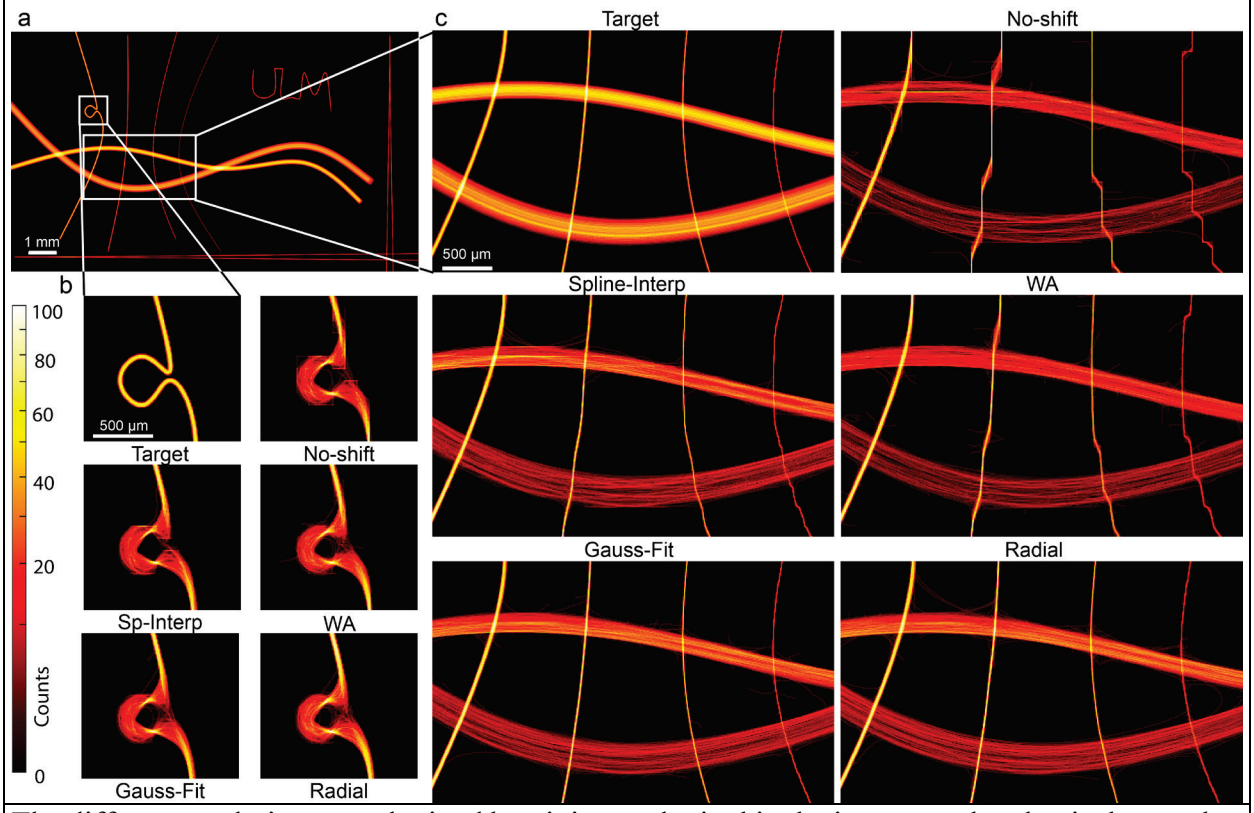
The spatial frequency peak to baseline analysis detailed in the methods section is presented in **Figure 7**. The cubic interpolation has a large score that is in line with what is described in the next section. The Sp-Interp and WA seem to have the same level of gridding. The radial and Gaussian fitting schemes come out on top with values almost 3 times less than the worst scheme.

*Image quality metrics in silico and in vivo*

With the provided datasets, both simulated and real microcirculation can be evaluated with the various algorithms. For instance, **Figure 6** shows the whole view of the ULM rendering of the simulated microcirculation. There are notable differences in how well algorithms manage to recover the initial shapes. In some cases, the overall contrast of the structures compared to the target is lower, with some vessels or part of vessels disappearing for example in the no-shift or cubic interpolation schemes. The horseshoe pattern and ULM watermark (**Figure 6** b) are the hardest structures to recover, few algorithms seem to perform well. Gauss-Fit/RS/Sp-Interp seem to yield the rendering closest to what was simulated, especially the curvature before and after the

horseshoe. This shape is interesting as it is similar to saccular aneurysms. The smallest tubes are correctly represented in most cases but the separation index presented in **Figure 3** gives us an idea of the maximum resolution attainable which is the gold standard for resolution determination.

**Figure 6: 2D ULM density-based renderings of "*in silico* Flow" along with a focus on particular areas**




The different renderings are obtained by giving each pixel in the image a value that is the number of trajectories passing through that coordinate

a) Target to be recovered by the algorithm
b) Zoomed in renderings for the different localization schemes on the horseshoe pattern
c) Zoomed in renderings for the different localization schemes on structures with various diameters and characteristics

The no-shift is valuable to study a hypothesis made in a previous paper about 3D ULM[37] stating that the grid effect appears because of erroneous localization in areas of low SNR. Because the no-shift does not involve sub-pixel localization, it should suffer the most from the grid effect and one can see that it does, appearing in all of the structures smaller than the wavelength of the ultrasound

sent (100 $\mu m$). Some behaviors are only visible when zooming in (**Figure 6** c). The WA and the Cub-Interp scheme seem to suffer similar grid effects on the same tubes. The Sp-Interp only suffers gridding on the two smaller curved tubes (10 and 5 $\mu m$). The Gauss-Fit and RS, on the other hand, adapt completely and do not suffer from any gridding on any curved tubes. More renderings are produced in ***Supplementary Figures 3-1, 3-2, 4-2, and 4-3.***

We devised an index to measure the grid-effect by a spatial frequency peak to baseline analysis in both directions. The value of that index grows with the severity of gridding. The values are reported in **Figure 7** next to the renderings of the brain vasculature. The no-shift-based ULM performs very poorly *in vivo* as predicted by the simulation and has the highest gridding index. The vessels in the cortex rendered by that scheme suffer from considerable gridding making their structure disorganized. The Cub-Interp also suffers from gridding as predicted, however, the WA localization-based ULM is in appearance not gridded.

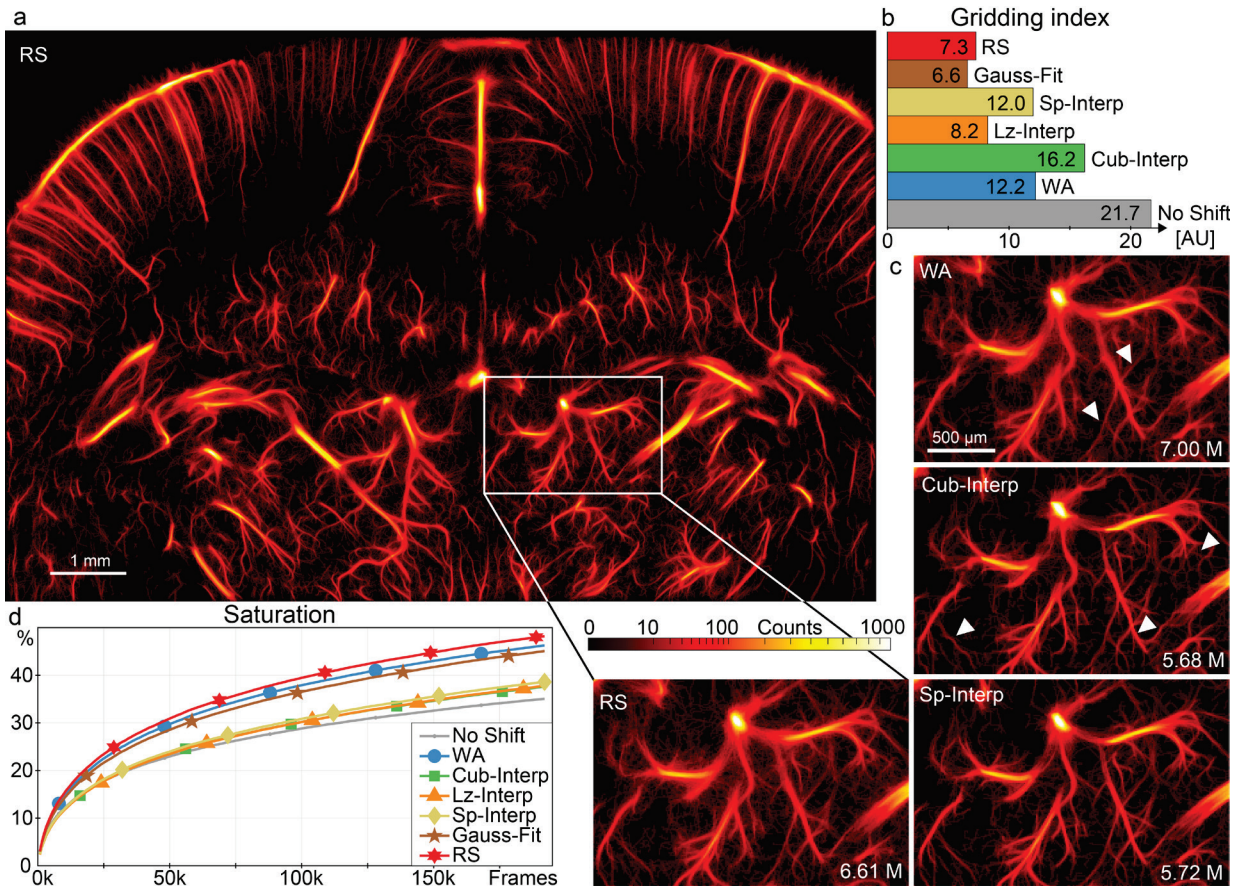
To further investigate the effects of localization, we zoomed in on a specific subcortical region where the grid effect is maximum, and where there is a large variety of vessel sizes (**Figure 7** b). Our previous conclusion holds partially. The Cub-Interp is still massively subjected to gridding and we can see some gridding appearing in specific regions in WA (indicated by white triangles). For the four best-ranked algorithms, almost no gridding is present in the images *in vivo*. One considerable difference can be noted though between the RS-based ULM and the interpolation schemes: the vessels seem more detailed and sharper in the interpolation schemes. In RS, there seems to be a lot of small crossing trajectories in between the brightest and largest vessels.

Further insight can be drawn from **Figure 8** where the other *in vivo* datasets are presented. In the kidney, where the vasculature is organized differently than in the brain, these small crossing trajectories are not observed for the RS scheme. In the tumor, the vasculature is disorganized but

no crossing trajectories are observed anyway. The difference between the best of the interpolation schemes (Sp-Interp) and the GF or RS algorithms is notable in the tumor where fewer vessels are being picked up by the former than the latter.

The saturation curves, counting the number of illuminated pixels throughout the reconstruction, show us the capacity of algorithms to detect new structures. When interpolated based localizations saturate, WA/Gauss-Fit/RS are still recovering new pixels in the image. Finally, the RS obtain the highest final saturation value at 48.2%, 1.9% above the Weighted-Average.

**Figure 7: 2D ULM density-based renderings of "*in vivo* rat brain perfusion", with gridding index and saturation**




The different renderings are obtained by giving each pixel in the image a value that is the number of trajectories passing through that coordinate

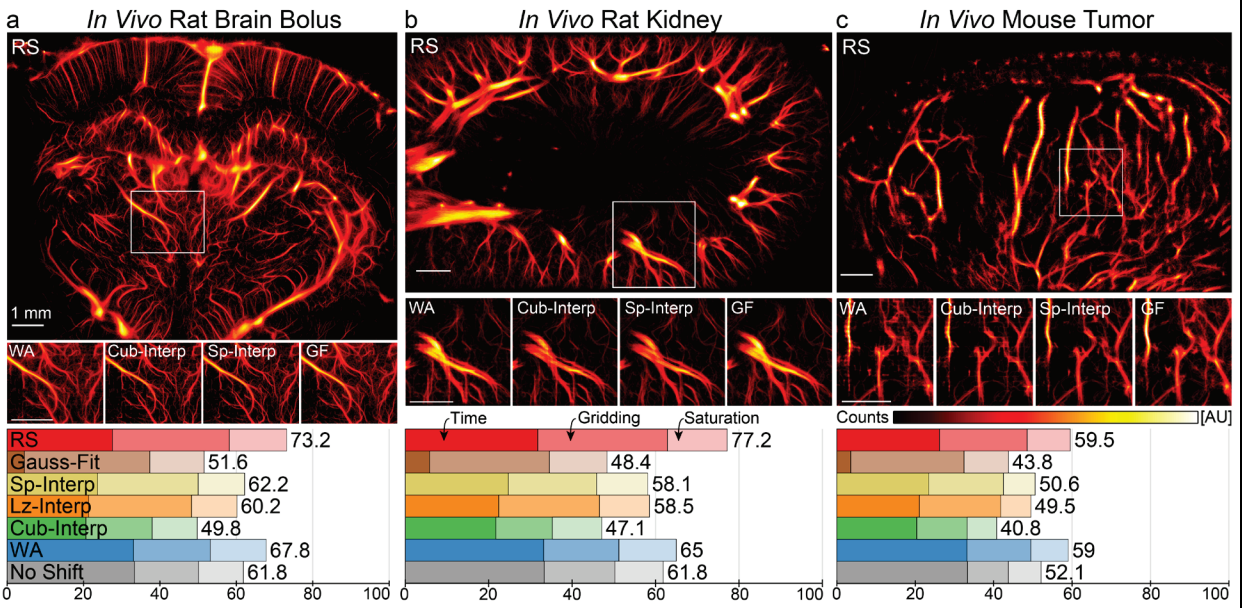
a) Whole rat brain rendering using radial localization algorithm
b) Gridding index calculated on the dataset "*in vivo* rat brain perfusion" for every algorithm.

c) Zoomed in portions of rat brain ULM using cubic interpolation, weighted average, spline interpolation, and radial symmetry-based localizations.
d) Saturation coefficient calculated for each algorithm

For more detailed renderings, see ***Supplementary Figures 4-2 to 4-3.***

The results obtained on the three other *in vivo* datasets are presented in **Figure 8** (see also ***Supplementary Figures 4-5 to 4-7***). The patterns described in the previous dataset still hold. In the *in vivo* rat brain bolus, the superiority of the GF and RS algorithms is confirmed, the overall score obtained by the RS is even higher than its score on the rat brain perfusion dataset. This superiority and the GF algorithms' high-quality imaging is best illustrated with the tumor where few vessels are present and are characterized by a disorganized structure. The GF and RS algorithm depict evidently more vessels than the other schemes. The gridding effect is clearly seen in the WA and Cub-Interp localization schemes. As for the Sp-Interp, its imaging quality is quite high for example in the kidney, it is almost identical to the GF algorithm, but in the tumor dataset, we can see clear evidence that it localizes fewer vessels than GF and RS. The partial score (obtained from time, gridding, and saturation) difference between rat brain and tumor for the RS is 23% while for the Weighted Average it is considerably less at 15%. This difference for the interpolation-based schemes is interestingly constant at 22%. For the no-shift, the ratio is 19%.

**Figure 8: 2D ULM density-based renderings of 3 other *in vivo* datasets, with partial scores**




Density renderings obtained with radial symmetry are presented for 3 other *in vivo* datasets from the rat brain, rat kidney, and mouse tumor. Below each image is presented a cropped area (in white) of the same respective datasets obtained with Weighted Average, Cub-Interp, Sp-Interp, and Gaussian fitting renderings. Three dataset-specific metrics are presented on each of these datasets for all localization algorithms, computation time, gridding index, and saturation index.

a) *In vivo* rat brain imaging obtained by injecting microbubbles with bolus in the tail vein. The imaging plane acquired is different from the previous dataset presented. Noise is present on the far left corner of this dataset due to the remaining bone presence.
b) *In vivo* rat kidney imaging obtained after isolating the kidney outside of the abdomen. Motion has been corrected for. The vascular structure is clearly recognizable with the two main arteries and veins on the left of the image
c) *In vivo* mouse tumor imaging obtained on a subcutaneous back tumor. The vascular structure here is clearly disorganized and not as sharp as in the kidney and brain images. Motion has been corrected for.

## Discussion

The ULM workflow evaluated here includes every step after data acquisition/generation to final rendering, along with the calculations of the various proposed metrics for ULM quality. The localization step alone was benchmarked as it is the most variable aspect of ULM algorithms used by different teams. However, filtering, ultrasound acquisition parameters, contrast agent types, and tracking algorithms should also be studied. Partial studies have been made about filtering[37,38],

contrast agent types[39,40]; ultrasound acquisition parameters[22] as well as tracking algorithms[38] but it is hard to draw general conclusions for the techniques as these papers do not use the same datasets. We hope that with the datasets provided in this paper, comparison and accurate characterization of all the steps will be feasible. Devising metrics to compare and rate specific aspects of localization is a crucial step in providing a better understanding of super-resolution imaging. Visual scoring by radiologists, neurologists, or physiologists will be required to establish quality grading for each application. However, ULM will first need to gain better uniformity through quantitative and objective metrics assessments.

Thanks to the different datasets generated, we can quantify the influence of PSF inhomogeneity, size, complexity, the position of the shape to recover, and behavior *in vivo* of the localization schemes. The *in vivo* datasets are particularly important as they validate the different approaches in a pre-clinical setting. Specifically, they focus on brain, kidney, subcutaneous tumors which have very different vascular organizations and exhibit different experimental characteristics. For example, the tumor and kidney images will be subjected to more significant breathing motion, the tumor will also be fed by disorganized and low perfused vessels. The metrics were chosen to be as universal as possible and should work in any workflow, especially the ground truth-based metrics. The availability of the *in vivo* datasets should be seen as an incentive to both enhancing existing localization algorithms and comparing the imaging quality of new *in vivo* applications. These *in vivo* datasets can also be used to try new approaches without going through the many stumbling points strewn along with preclinical experimentation.

We think the metrics presented here could be used whenever a new processing step or technology for ULM is implemented. Firstly, the improvement in localization following the integration of this innovation should be characterized by using the *in silico* PSF and *in silico* flow datasets. Rather

than relying on the full width at half maximum of vessels, these metrics are based on the ground truth and thus are more accurate in quantifying localization improvements. Secondly, the increase in the number of localizations should be quantified by using the statistical indices (True Positive, False Negative, and Positive), and the impact of that increase be measured with the Separation Index thanks to the codes available on the platform. Finally, we recommend that the impact of any innovation in ultrasound super-resolution is benchmarked on *in vivo* data as this technique's main application is pre-clinical and clinical imaging of the microvasculature. In order to include this, the *in vivo* datasets given here should be used as databases and the metrics Number of detections, Saturation, Gridding Index, and Processing Times should be computed and discussed in the report. The combination of these scores in a global score (identical to the PALA global score or with different weights to reflect other scenarii) is made easy by the dynamic table on the platform and will allow the community to evaluate the innovation at a quick glance.

An interesting aspect brought by the distribution errors point of view is the peaks present in the cubic, Lanczos, and spline interpolation schemes for the axial error. They are periodic with a frequency corresponding to the $\frac{\lambda}{10}$ interpolation grid. This would lead to a clear pattern on the final images.

On image quality metrics, we notice that the grid effect is especially present for interpolation-based localization algorithms. It is very similar to quantization in its form and as the no-shift method confirms, it is due to errors in the localization. This means that in interpolation schemes, gridding comes from a reprojection of the localized center towards the center of the original pixel grid. It can be linked to low SNR in certain regions following this line of thought: if SNR is low, the full width at half maximum of the intensity increases so much that covered by the clutter noise, not enough information is present in the profiles to perform an accurate interpolation or

localization. When confronted with low SNR images, the algorithms tend to calculate the super-resolved center very close to the original center because it lacks sufficient signal, and thus all positions align along one line, tracing that gridding pattern. This is confirmed by looking at where gridding happens in Lz-Interp and Sp-Interp schemes *in silico* or *in vivo*. In the large vessels, the number of events is sufficient to compensate for that effect, but for smaller vessels, the density of microbubbles is too low. This can be seen in the tumor images.

As such, attention should be paid to the reconstruction grid used for rendering. Decreasing the pixel size too much would increase the grid effect and the interpolation kernels, making the calculations more time-consuming. *In vivo,* the interpolation factor depends on the maximum theoretical resolution attainable, and as such, we advise that the super-resolved image pixel size and the maximum theoretical resolution be matched. One can note that even though localization is performed poorly in the no-shift scheme, the rendering still appears believable and that should be kept in mind when evaluating the quality of ULM. Additionally, as was seen in the no-shift localization scheme, if the reconstruction grid is too fine but the tracking algorithm contains an interpolation of tracks or smoothing, the grid effect may disappear and some artifacts may appear. These artifacts will lead to erroneous velocity estimations which can be measured and characterized by looking at the Poiseuille distributed profiles for these small vessels. Another difference to be noted is the lower gridding index calculated for WA compared to Sp-Interp although in **Figure 7** b the gridding effect is clearly higher. With 7.00M events localized in the *in vivo* rat brain perfusion dataset, WA is the algorithm that localizes the most events (1.3M more than Sp-Interp). The lower precision and higher RMSE compared to the Sp-Interp means that the final rendering contains a lot more erroneous localizations thus decreasing the principal peaks' amplitude in the frequency analysis devised for the gridding index.

391 *In vivo,* interpolation-based localizations seem to have a very sharp delineation of the vessels
392 whereas RS/Gauss-Fit have many small trajectories going in and out of the most intense vessels.
393 This makes the image obtained by the spline interpolation localization look better than that
394 obtained with the radial symmetry-based localization. The latter looks like it has more noise than
395 its interpolation-based counterpart. However, the *in silico* metrics tell us that the RS localization
396 is more sensitive and precise, meaning that it localizes microbubbles more correctly. It is
397 confirmed by looking at the total number of particles localized (see **Figure 4**), where the increase
398 in the number of microbubbles localized in the radial symmetry-based algorithm is 21%, and by
399 the saturation, 4% higher than the second algorithm. We hypothesize that the small trajectories in
400 between vessels are also small vessels that can represent the capillary bed. When looking at how
401 the kidney and tumor images are depicted by the different algorithms, we note the absence of cross
402 trajectories. The duration of ultrasound imaging in these organs does not allow to represent the
403 capillary bed clearly as the recovery of small vessels needs longer acquisition times. If the cross
404 trajectories were due to noise, however, we would see them regardless of recording duration. A
405 possible way to verify this *in vivo* would be to have access to the ground truth for example by
406 observing fluorescently labeled microbubbles with ULM and optical imaging. A recently
407 published study has indeed used such microbubbles[41]. Although coming in second after the
408 Gaussian fitting algorithm in precision, and Jaccard index, the radial symmetry remains a good
409 candidate for analysis of the capillary bed. Globally if time and low RMSE are more critical than
410 an accurate classification of microbubbles, then the radial symmetry should be favored.

411 The time factor, measured for the fastest algorithm, is critical to the fast testing of many different
412 parameters. If the algorithm of localization runs for hours before being able to deliver a full image,
413 changing parameters and waiting for the result will take days. This becomes critical in 3D as the

414 third dimension makes everything considerably more complex. In our team, we have chosen to
415 give priority to speedy calculations. That meant that in 2D, it only takes us 3 minutes to perform
416 full ULM on 192,000 images which is less than the total acquisition time (4 minutes). Our
417 computer is largely above average in terms of performance (Intel Core i9 @ 2.9 GHz 12 cores,
418 NVidia RTX 2080Ti, 128GB RAM @ 2133 MHz) but the ratio of speed improvement would be
419 similar for any machine. This implementation appears much faster than many other algorithms. It
420 improves the applicability of ULM, by allowing a simultaneous acquisition and reconstruction. In
421 a clinical setting, that also means that the practitioner does not have to wait for its result and can
422 either make another image in a matter of minutes or deliver his diagnosis.

423 The PALA global score we have devised and presented here aims at compiling all of these different
424 metrics to aid algorithm selection. Our own defined global score was aimed at favoring fast and
425 low error algorithms. However, one might want to favor calculation time above all other metrics
426 for real-time applications or 3D. To do this we provide a dynamic table in supplementary
427 (PALA_GlobalScores_DynamicTable.xlsx) where 6 different scenarios were designed to reflect
428 experimental requirements: Non-weighted, 3D ULM, 2D scanning, Real-time, Low SNR, and our
429 own defined global score. An additional line called New Entry allows the reader to set its own set
430 of weights.

431 Several limitations remain in this study. For instance, the localization is performed on
432 microbubbles imaged at a high-frequency (15 MHz), limiting the effects linked to the resonant
433 oscillations of microbubbles[17]. While that allows all of the results on the errors to be readily
434 transposable to other frequencies, this is only applicable if linear imaging is used. Also as we have
435 no access to ground truth *in vivo* at these depths in living animals, the structure of the capillary can
436 only be hypothesized. We have not included the impact of beamforming techniques aimed at

enhancing the PSF and CNR, as well as the different tracking algorithms available for ULM. However, thanks to the versatility of our framework, these can be easily included and benchmarked *ad hoc*. This versatility can be put to use to develop new metrics. One very valuable metric would be an index to benchmark the quality of ULM algorithms in between organs and in between datasets coming from different groups. Comparing organs that have a very different vascular organization, such as the kidney or the brain, remains a challenge. In 2D, the possibility to have an atlas of the vasculature of the rat's brain could be explored but will probably run into out-of-plane related issues (motion, size of the plane at focus, interrupted tracks, reprojection of vessels).

The noise simulated in the *in silico* dataset is a simple model aimed at imitating clutter noise. Added to the beamformed images, it does not take into account jitter, false peak errors in the radiofrequency data, readout noise from the electronics, nor does it represent potential motion. As the *in vivo* data we have presented have been filtered with a spatiotemporal clutter filter, our aim, by choosing this simplistic approach, was to reproduce SNR values and aspects seen in the *in vivo* cases. However, thanks to the versatility of our platform, any noise model could be added to the *in silico* dataset easily.

We hope that these algorithms, datasets, and metrics will improve the comparisons and discussions among the growing ULM community and allow more users with less ultrasonic expertise to exploit these new methods. Besides potentially becoming a valuable clinical imaging modality, ULM opens a new window to observe the microcirculation in-depth and complements other laboratory imaging methods that lack either resolution or penetration.

**Outlook**

We have developed a platform to analyze and measure the performance of the most used localization schemes in Ultrasound Localization Microscopy. This led us to integrate these in an open-source algorithm, which is available in the supplementary information, to facilitate the dissemination of the technique for research groups around the world. Furthermore, we provide the data and scripts needed to benchmark future techniques or other aspects of the ULM process. Also, these algorithms can be readily used on novel contrast agents such as acoustically activated nanodroplets, biogenic gas vesicles, or novel imaging applications such as 4D ultrasound or non-linear imaging. This open framework could benefit from further development such as non-linear behavior simulation or tracking improvement. Finally, we have presented two novel algorithms, one of which obtains the highest score according to the criteria proposed here. With the advent of new artificial intelligence-based super-resolution methods, the metrics devised will help to compare and characterize such methods. We see great potential in the collective use of these resources to improve our knowledge of ULM, increase its quality, and provide better and faster imaging for 3D developments.

## References


1. Tanter, M. & Fink, M. Ultrafast imaging in biomedical ultrasound. *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **61**, 102–119 (2014).
2. Macé, E. *et al.* Functional ultrasound imaging of the brain. *Nat. Methods* **8**, 662–664 (2011).
3. Macé, E. *et al.* High sensitivity brain angiography using ultrafast Doppler. in (2010). doi:10.1109/ultsym.2010.5935810.
4. Demene, C. *et al.* Spatiotemporal Clutter Filtering of Ultrafast Ultrasound Data Highly Increases Doppler and fUltrasound Sensitivity. *IEEE Trans. Med. Imaging* **34**, 2271–2285 (2015).
5. Frinking, P., Segers, T., Luan, Y. & Tranquart, F. Three Decades of Ultrasound Contrast Agents: A Review of the Past, Present and Future Improvements. *Ultrasound Med. Biol.* **46**, 892–908 (2020).
6. Couture, O., Besson, B., Montaldo, G., Fink, M. & Tanter, M. Microbubble ultrasound super-localization imaging (MUSLI). in 1285–1287 (IEEE, 2011). doi:10.1109/ULTSYM.2011.6293576.
7. Olivier Couture, Mickael Tanter & Mathias Fink. Patent 889 Cooperation Treaty (PCT)/FR2011/052810.
8. Couture, O., Hingot, V., Heiles, B., Muleki-Seya, P. & Tanter, M. Ultrasound Localization Microscopy and Super-Resolution: A State of the Art. *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **65**, 1304–1320 (2018).
9. Errico, C. *et al.* Ultrafast ultrasound localization microscopy for deep super-resolution vascular imaging. *Nature* **527**, 499–502 (2015).

10. Christensen-Jeffries, K., Browning, R. J., Tang, M.-X., Dunsby, C. & Eckersley, R. J. In Vivo Acoustic Super-Resolution and Super-Resolved Velocity Mapping Using Microbubbles. *IEEE Trans. Med. Imaging* **34**, 433–440 (2015).
11. Heiles, B. Ultrafast Volumetric Ultrasound Localization Microscopy in vivo. in (2018).
12. Viessmann, O. M., Eckersley, R. J., Christensen-Jeffries, K., Tang, M. X. & Dunsby, C. Acoustic super-resolution with ultrasound and microbubbles. *Phys. Med. Biol.* **58**, 6447–6458 (2013).
13. Lin, F. *et al.* 3-D Ultrasound Localization Microscopy for Identifying Microvascular Morphology Features of Tumor Angiogenesis at a Resolution Beyond the Diffraction Limit of Conventional Ultrasound. *Theranostics* **7**, 196–204 (2017).
14. O'Reilly, M. A. & Hynynen, K. A super-resolution ultrasound method for brain vascular mapping: Super-resolution ultrasound method for brain vascular mapping. *Med. Phys.* **40**, 110701 (2013).
15. Song, P. *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* **65**, 149–167 (2018).
16. Brown, J. *et al.* Investigation of Microbubble Detection Methods for Super-Resolution Imaging of Microvasculature. *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **66**, 676–691 (2019).
17. Christensen-Jeffries, K. *et al.* Microbubble Axial Localization Errors in Ultrasound Super-Resolution Imaging. *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **64**, 1644–1654 (2017).
18. Kanoulas, E. *et al.* Super-Resolution Contrast-Enhanced Ultrasound Methodology for the Identification of In Vivo Vascular Dynamics in 2D: *Invest. Radiol.* **54**, 500–516 (2019).

19. Ackermann, D. & Schmitz, G. Detection and Tracking of Multiple Microbubbles in Ultrasound B-Mode Images. *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **63**, 72–82 (2016).
20. Luke, G. P., Hannah, A. S. & Emelianov, S. Y. Super-Resolution Ultrasound Imaging in Vivo with Transient Laser-Activated Nanodroplets. *Nano Lett.* **16**, 2556–2559 (2016).
21. O'Reilly, M. A. & Hynynen, K. A super-resolution ultrasound method for brain vascular mapping: Super-resolution ultrasound method for brain vascular mapping. *Med. Phys.* **40**, 110701 (2013).
22. Song, P., Manduca, A., Trzasko, J. D., Daigle, R. E. & Chen, S. On the Effects of Spatial Sampling Quantization in Super-Resolution Ultrasound Microvessel Imaging. *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **65**, 2264–2276 (2018).
23. Hansen, K. B. *et al.* Robust microbubble tracking for super resolution imaging in ultrasound. in *2016 IEEE International Ultrasonics Symposium (IUS)* 1–4 (IEEE, 2016). doi:10.1109/ULTSYM.2016.7728793.
24. Heiles, B. *et al.* Ultrafast 3D Ultrasound Localization Microscopy using a 32x32 Matrix Array. *IEEE Trans. Med. Imaging* 13 (2019).
25. Soulioti, D. E., Espíndola, D., Dayton, P. A. & Pinton, G. Super resolution imaging through the human skull. *ArXiv181110653 Cond-Mat Physicsphysics* (2018).
26. Viessmann, O. M., Eckersley, R. J., Christensen-Jeffries, K., Tang, M. X. & Dunsby, C. Acoustic super-resolution with ultrasound and microbubbles. *Phys. Med. Biol.* **58**, 6447–6458 (2013).

27. Zhang, G. *et al.* Acoustic wave sparsely activated localization microscopy (AWSALM): Super-resolution ultrasound imaging using acoustic activation and deactivation of nanodroplets. *Appl. Phys. Lett.* **113**, 014101 (2018).
28. Zhu, J. *et al.* 3D Super-Resolution US Imaging of Rabbit Lymph Node Vasculature in Vivo by Using Microbubbles. *Radiology* **291**, 642–650 (2019).
29. Huang, C. *et al.* Short Acquisition Time Super-Resolution Ultrasound Microvessel Imaging via Microbubble Separation. *Sci. Rep.* **10**, 6007 (2020).
30. Parthasarathy, R. Rapid, accurate particle tracking by calculation of radial symmetry centers. *Nat. Methods* **9**, 724–726 (2012).
31. Christensen-Jeffries, K. *et al.* 3-D In Vitro Acoustic Super-Resolution and Super-Resolved Velocity Mapping Using Microbubbles. *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **64**, 1478–1486 (2017).
32. Desailly, Y., Couture, O., Fink, M. & Tanter, M. Sono-activated ultrasound localization microscopy. *Appl. Phys. Lett.* **103**, 174107 (2013).
33. Desailly, Y., Pierre, J., Couture, O. & Tanter, M. Resolution limits of ultrafast ultrasound localization microscopy. *Phys. Med. Biol.* **60**, 8723–8740 (2015).
34. Song, P., Manduca, A., Trzasko, J. D., Daigle, R. E. & Chen, S. On the Effects of Spatial Sampling Quantization in Super-Resolution Ultrasound Microvessel Imaging. *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **65**, 2264–2276 (2018).
35. Hingot, V. *et al.* Microvascular flow dictates the compromise between spatial resolution and acquisition time in Ultrasound Localization Microscopy. *Sci. Rep.* **9**, (2019).
36. Kuhn, H. W. The Hungarian method for the assignment problem. *Nav. Res. Logist. Q.* **2**, 83–97 (1955).

37. Heiles, B. *et al.* Ultrafast 3D Ultrasound Localization Microscopy using a 32x32 Matrix Array. *Accept. IEEE Trans. Med. Imaging* 13 (2019).
38. Song, P. *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* **65**, 149–167 (2018).
39. Luke, G. P., Hannah, A. S. & Emelianov, S. Y. Super-Resolution Ultrasound Imaging in Vivo with Transient Laser-Activated Nanodroplets. *Nano Lett.* **16**, 2556–2559 (2016).
40. Zhang, G. *et al.* Acoustic wave sparsely activated localization microscopy (AWSALM): Super-resolution ultrasound imaging using acoustic activation and deactivation of nanodroplets. *Appl. Phys. Lett.* **113**, 014101 (2018).
41. Lowerison, M. R., Huang, C., Lucien, F., Chen, S. & Song, P. Ultrasound localization microscopy of renal tumor xenografts in chicken embryo is correlated to hypoxia. *Sci. Rep.* **10**, 2478 (2020).
42. Forsberg, F., Leeman, S. & Jensen, J. A. Assessment of hybrid speckle reduction algorithms. *Phys. Med. Biol.* **36**, 1539–1549 (1991).
43. Ledoux, L. A. F., Brands, P. J. & Hoeks, A. P. G. Reduction of the Clutter Component in Doppler Ultrasound Signals Based on Singular Value Decomposition: A Simulation Study. *Ultrason. Imaging* **19**, 1–18 (1997).
44. Desailly, Y. *et al.* Contrast enhanced ultrasound by real-time spatiotemporal filtering of ultrafast images. *Phys. Med. Biol.* **62**, 31–42 (2017).

**Acknowledgments**

The project has been funded by the European Research Council under the European Union Horizon H2020 program/ERC Consolidator grant agreement No 772786-ResolveStroke. The project was also supported by Agence Nationale de la Recherche (ANR), within the project ANR Predic and the Plan Cancer UICT. We thank the laboratories Institut Langevin and PhysMed Paris for technical support and particularly Line Rahal for helping with data acquisition. We thank Cyrille Orset (INSERM UMR-S U1237, Physiopathology and Imaging of Neurological Disorders, GIP Cyceron, Institut Blood and Brain @ Caen-Normandie (BB@C), Caen, France) for animals' preparation and perfusion of contrast agent and the biomedical imaging platform CYCERON (UMS 3408 Unicaen/CNRS, Caen, France), particularly, Hanadi Skeif and Vincent Beaudouin for the micro-CT of brains. We would like to thank Ruslan Hlushchuk for all his valuable advices for µAngiofil perfusion, David Maresca for useful discussion and our collaborators Franck Lager, Gilles Renault and Alba Nicolas-Boluda for helping with the mouse tumor model and the *in vivo* experiments.

**Author information**

*Affiliations*

*Sorbonne Université, CNRS, INSERM, Laboratoire d'Imagerie Biomédicale, Paris, France*

Baptiste Heiles, Arthur Chavignon, Vincent Hingot, Olivier Couture

*ESPCI, CNRS, INSERM, PhysMedParis, Paris, France*

Baptiste Heiles, Vincent Hingot, Eliott Teston, Pauline Lopez

*Institut Cochin, INSERM U1016, Paris, France*

Pauline Lopez

*Contributions*

All authors conceived the project. ET and PL are the main collaborators on the kidney and tumor experiments. BH and VH laid out the general framework and algorithm for ULM with contributions from all authors.

BH wrote the original codes for localization algorithms with significant improvements brought by AC. AC wrote the simulation framework and implemented the metrics. VH, AC, ET, and PL acquired *in vivo* data. BH wrote the manuscript with input from all authors. OC has directed the work of BH, AC, and VH as thesis/postdoctoral supervisor.

*Corresponding author*

Correspondence to Baptiste Heiles: baptiste.heiles@gmail.com

**Ethics declarations**

*Competing interests*

BH, AC, PL, ET, and VH declare no competing financial interests. OC is a co-inventor of an ultrasound super-resolution patent (PCT)/FR2011/052810.

*Animal experimentation*

All animals received humane care in compliance European Community Council Directive of 22nd September 2010 (010/63/UE) and approved by the institutional committee C2EA-59: "Comité d'éthique en matière d'expérimentation animale Paris Centre et Sud" under the protocol APAFIS#13363-2018020321116321 (rat brain), APAFIS # 16874-2017122914243628 (rat kidney), the institutional committee C2EA-034: "Comité d'éthique en matière d'expérimentation

animale Paris Descartes", APAFIS #25169-202008071746473 (mouse tumor), and the institutional committee C2EA-54 "Comité d'éthique Normandie en matière d'expérimentation animale" under the protocol APAFIS #22544-2019093017523797 (rat brain with micro-CT)

## Online methods

### *Point Spread Function dataset* ("*in silico* PSF")

The Point-Spread Function (PSF) simulation is designed to study non-uniformity effects on localization and spatial sampling effect on the beamforming process. A point-like scatterer is simulated using the Verasonics Research Ultrasound Simulator (Verasonics Inc., Kirkland Washington, USA). The position of the scatterer is moved by $\frac{\lambda}{7}$ increments in both directions in a $\lambda \times \lambda$ space. For each of these positions, the radiofrequency response, as well as its beamformed image, was produced using 3 tilted plane waves emitted from a 15MHz linear probe, with 128 elements spaced by one wavelength. The speed of sound was set to $1540\ \mathrm{m/s}$.

### *In silico angiography dataset* ("*in silico* Flow")

The media simulated comprised of 11 tubes with various geometries with various complexity (see **Figure *6*** and ***Supplementary Figure 2-1***):

| Structure | Diameters | Maximal velocities |
|---|---|---|
| A pseudo double helix | $3\ \lambda$<br>$2\ \lambda$ | $v_{max} = 3\ \lambda/frame$<br>$v_{max} = 2\ \lambda/frame$ |
| A curved tube with a constant diameter and a horseshoe pattern | $0.5\ \lambda$ | $v_{max} = 0.5\ \lambda/frame$ |
| 3 curved tubes with 3 different diameters | $0.2\ \lambda$<br>$0.1\ \lambda$<br>$0.05\ \lambda$ | $v_{max} = 0.2\ \lambda/frame$<br>$v_{max} = 0.1\ \lambda/frame$<br>$v_{max} = 0.05\ \lambda/frame$ |
| 4 spreading tubes with a constant diameter | $0.1\ \lambda$ | $v_{max} = 0.9\ \lambda/frame$ |
| A watermark comprised of the word ULM that does not serve other purposes but identification | $0.1\ \lambda$ | $v_{max} = 0.4\ \lambda/frame$ |

To simulate moving microbubbles, we chose initial random positions in the tubes and recorded them as point-like scatterers. For each of these positions, a trajectory calculated by Poiseuille's law was assigned based on their position in the cross-section of the tube. We considered that microbubbles could not jump from one trajectory to another and so had to follow the same trajectory for as long as it is simulated. Using the Verasonics Research Ultrasound Simulator, we simulated radiofrequency and beamformed images of a moving scatterer in a $7 \times 14.9\ mm^2$ area insonified with 3 tilted plane waves with a 15MHz linear probe, with 128 elements spaced by one wavelength. The simulation sequence was based on the Vantage example script (Vantage release 4.0.0) *FlashAngles* and adapted to save either radiofrequency data and beamformed images with $\lambda \times \lambda$ pixel size. The speed of sound was set to $1540\ m/s$.

For each image in the two *in silico* datasets, we added clutter noise modeled by a Gaussian filtered white Gaussian noise (see ***Supplementary Figure 4-2***). The parameters for the white Gaussian Noise were matched with a typical $30dB$ noise present in our *in vivo* images: we used *wgn* Matlab (The MathWorks, Inc., Natick Massachusetts, USA) function to generate a 0.2 ohm impedance noise, with a $\pm 10dB$ amplitude. The resulting noise was then filtered using a 2D Gaussian kernel of size 0.7 pixel. Finally, the noise was added to the beamformed image by choosing an amplitude in dB between the maximum intensity value of the beamformed image and the average intensity of the generated noise. This process results in noised images, with a user-selected SNR, and a pseudo clutter with $20dB$ of amplitude.

*In vivo dataset* ("*in vivo* rat brain perfusion" and "*in vivo* rat brain bolus")

All animals received humane care in compliance European Community Council Directive of 22nd September 2010 (010/63/UE) and approved by the institutional committee C2EA-59: "Comité d'Ethique en matière d'Expérimentation Animale Paris Centre et Sud" under the protocol 2015–23. 8-10 weeks Sprague-Dawley rats were obtained from Janvier Labs (Le Genest-Saint-Isle, France). The animals were kept for at least a week before surgery. Water and a commercial pelleted diet SAFE A04-10 (Augy, France) were available *ad libitum*. SAFE FLAKE sawdust was used as bedding, enrichment such as pieces of cardboard, paper tunnels were provided. After 5 minutes in the induction cage filled with a mix of air and 5% of Isoflurane, the animal was placed on the back on a heated plate with a respiratory mask. The mix was replaced with $O_2$ mixed with 4% of Isoflurane. The depth of the anesthesia was tested by the absence of withdrawal reflex when pinching the toes of the hind limb. A jugular vein catheter was then placed and craniotomy surgery was used to remove the skull on a window in between the Lambda and Bregma planes and 14 $mm$ wide. 400 $\mu L$ of Sonovue microbubbles (Bracco Imaging ®, Milan, Italy) were injected at a steady rate of 80 $\mu L/min$ to keep a stable concentration in the bloodstream or with boli in the second dataset. A continuous ultrasound acquisition comprised of 240 blocs (or 213 blocs for bolus dataset) of 800 frames taken at a compounded frame rate of 1000 Hz with 3 tilted plane waves $[-3; 0; 3]°$ or ($[-5; 0; 5]°$ for bolus dataset) using a 15 MHz centered frequency probe manufactured by Vermon®, Tours, France with a 0.1 $mm$ pitch. Acquisitions lasted 4 min for the perfusion dataset, and 3 min 30 sec for the bolus dataset. Elevation focusing is provided through a plastic lens to reach 500 $\mu m$ at 8 $mm$. The radiofrequency data were beamformed on the ultrasound machine's GPU (SuperSonic Imagine, Aix-en-Provence, France) and transferred to an external hard drive for later processing. The microbubbles' SNR was measured on a filtered image as the intensity ratio between the 10 brightest microbubbles and the average image intensity.

*In vivo dataset* ("*in vivo* rat kidney")

All animals received humane care in compliance European Community Council Directive of 22nd September 2010 (010/63/UE) and approved by the institutional committee C2EA-59: "Comité d'Ethique en matière d'Expérimentation Animale Paris Centre et Sud" under the protocol APAFIS # 16874-2017122914243628. The same housing protocol and strain of rat were used as the animals used for the brain experiments. The kidney was exposed through a lumbotomy incision and placed on an acoustically absorbing material to facilitate imaging. A heat lamp was used to prevent hypothermia of the animal. Ultrasound gel was applied on top of the imaging area to provide coupling, Sonovue microbubbles (Bracco Imaging ®, Milan, Italy) were injected in boli via a jugular vein catheter. A continuous ultrasound acquisition comprised of 240 blocs of 800 frames taken at a compounded frame rate of 1000 Hz with 5 tilted plane waves $[-10; -5; 0; 5; 10]°$ using a 15 MHz centered frequency probe manufactured by Vermon®, Tours, France with a $0.1\ mm$ pitch. The acquisition lasted 4 minutes. Elevation focusing is provided through a plastic lens to reach $500\ \mu m$ at $8\ mm$.

*In vivo dataset* ("*in vivo* mouse tumor")

All animals received humane care in compliance European Community Council Directive of 22nd September 2010 (010/63/UE) and approved by the institutional committee C2EA-34: "Comité d'Ethique en matière d'Expérimentation Animale Paris Descartes" under the protocol APAFIS #25169-202008071746473. For this study, MMTV-PyMT transgenic mice were backcrossed with FvB/NCrl mice (Charles River Laboratories, Wilmington Massachusetts, USA). A transplanted model of the MMTV-PyMT was used for this study. Tumor implantation was performed using primary tumors developed in donor mice. Donor mice with primary tumors of ~10–15 mm were killed and tumors were pooled, dissociated, and then re-injected orthotopically

in receiver mice. Tumors were dissociated mechanically and incubated for 30 minutes at 37°C with DNase I (100 µg/ml, Roche Diagnostics GmbH, Mannheim, Germany), collagenase (1mg/ml, Roche Diagnostics GmbH, Mannheim, Germany), and hyaluronidase (1 µg/ml, Sigma-Aldrich, St. Louis Missouri, USA) in RPMI with 2% serum. After red blood cell lysis and filtration on a 70 µm sieve, the cell suspension was rinsed 3 times in PBS and used for tumor transplantation. Thus, $10^6$ cells isolated from tumors of MMTV-PyMT mice were injected in 100 µl of PBS in the mammary fat pad of 8 weeks-old FvB mice. About 10 d after inoculation, transplanted PyMT tumors reached a diameter of ~ 6mm. A continuous ultrasound acquisition comprised of 150 blocs of 200 frames taken at a compounded frame rate of 500Hz with 5 tilted plane waves $[-10; -5; 0; 5; 10]°$ using a 15 MHz centered frequency probe manufactured by Vermon®, Tours, France with a $0.1\ mm$ pitch. The acquisition lasted 2 min 30 sec. Elevation focusing is provided through a plastic lens to reach $500\ \mu m$ at $8\ mm$.

*ULM workflow and detection*

ULM is a multi-step method. The general layout is presented in **Figure *2*** and ***Supplementary Figure 4-8***. Firstly, one needs to acquire a large number of frames in the region of imaging. Then, a detection process takes place to enhance the signal of microbubbles. Close microbubbles are isolated. *In silico,* such a step is dispensed with as the signal-to-noise ratio is set high enough in the simulation. For the *in vivo* dataset, we use the Singular Value Decomposition approach[4,42,43]. By rearranging a continuous set of beamformed images as a Casorati matrix $(N_z \times N_x, N_t)$, we will use the new basis given by the SVD to filter the tissue from the microbubbles according to its spatiotemporal decorrelation[44]. We remove the four highest energy vectors out of a basis of 800.

*Tracking algorithm*

After microbubbles' positions are measured and stored, a tracking algorithm was implemented based on the Hungarian method or Kuhn-Munkres algorithm for assignment. For each particle, the squared distance to all of the particles in the subsequent frame is computed. We obtain a vector containing $N_{particles}(t + dt)$, the number of particles in the frame $(t + dt)$. By applying the Kuhn-Munkres algorithm, we are able to find the optimal pairing for the subsequent frames by minimizing the total squared distance. We can calculate the velocities of each microbubble by differentiating each position in a trajectory according to the time vector.

*Reconstruction*

We build the final image by accumulating all the super-resolved positions of the microbubbles through time in a user defined pixel grid in a Maximum Intensity Projection fashion. The pixel intensity represents the number of tracks passing through this pixel (see ***Supplementary Figure 4-9)***.

*In silico metrics*

Thanks to the simulation tool, we have access to the ground truth and can devise numerical and several statistical metrics based on classification:

- the error of localization in two directions $Z, X$ and calculate the average root mean square error (RMSE). $(z^0; x^0)$ represent the absolute position of the scatterer, and $(z; x)$ the position measured by algorithms.

$$\delta_z(i) = {z_i}^0 - z_i \; ; \; \delta_x(i) = {x_i}^0 - x_i$$

$$RMSE = \frac{1}{N}\sum_{i}^{N}\sqrt{(x_i - {x_i}^0)^2 + (z_i - {z_i}^0)^2}$$

- True-positives (TP): a microbubble was simulated and a microbubble was localized in a $\frac{\lambda}{4}$ excluding neighborhood
- False-positive (FP): a microbubble was localized but no microbubble was simulated in the neighborhood
- False-negative (FN): no microbubble was localized within a distance $\frac{\lambda}{4}$ of the simulated microbubbles
- the Jaccard index as: $JAC = \frac{TP}{FN+TP+FP}$
- positive predictive value (precision) $p$ and recall (sensitivity): $p = \frac{TP}{FP+TP}, r = \frac{TP}{FN+TP}$
- separation index: minimal distance between inside borders of simulated canals, maximum resolution criteria

*In vivo metrics*

A visual analysis of the *in vivo* ULM rendering is done for each of the algorithms following several criteria: the number of localizations, likeliness of the trajectories, histogram of trajectory lengths, and visual aspects such as aliasing and branching vessels. The saturation curve represents the fraction of non-zero pixels in the image. The power spectral density of the spectral content in the $(Z, X)$ directions of the intensity was calculated as:

$$F_z = FFT\left(\sum_x I(z,x)\right) \; ; \; F_x = FFT\left(\sum_z I(z,x)\right)$$

The peak to the baseline value of the two first peaks of the power spectral density are summed in a logarithmic scale to quantify the amount of aliasing present in the image. For more details see ***Supplementary Figure 3-16.***

Finally, the computation time was evaluated for each algorithm in the *in vivo* case and compared to the fastest algorithm.

All 11 metrics can be summarized in a table:

| Number | Measurement | Dataset | | |
|---|---|---|---|---|
| | | **In silico PSF** | **In silico flow** | **In vivo** |
| 1 | Lateral error | X | X | |
| 2 | Axial error | X | X | |
| 3 | RMSE | X | X | |
| 4 | True Positive | | X | |
| 5 | False Negative | | X | |
| 6 | False Positive | | X | |
| 7 | Gap | | X | |
| 8 | Number of detections | | | X |
| 9 | Saturation | | | X |
| 10 | Gridding Index | | | X |
| 11 | Processing time | | | X |

*Conversion metrics to scores*

Before building the global score, metrics are normalized into a percentage value. The following table explains the conversion from metrics results to scores. The global score is a weighted average of these individual scores.

| Metric | Unit | Conversion | Score Range |
|---|---|---|---|
| RMSE | Wavelengths | $Score(x_i) = 100 * (1 - \frac{x_i}{0.5\lambda})$ | $0\ \lambda : 100/100$<br>$0.5\ \lambda : 0/100$ |
| Processing Time | Seconds | $Score(x_i) = 100 * (1 - 0.5 * log10\left(\frac{x_i}{min(x)}\right))$ | $min(x)$: $100/100$<br>$100 * min(x) : 0/100$ |
| Separation Index | Wavelengths | $Score(x_i) = 100 * (1 - \frac{x_i}{1\lambda})$ | $0\ \lambda : 100/100$<br>$1\ \lambda : 0/100$ |
| Gridding Index | dB | $Score(x_i) = 100 * (1 - \frac{x_i}{30})$ | $0\ dB : 100/100$<br>$30\ dB : 0/100$ |
| Jaccard | [%] | $Score(x_i) = x_i$ | $100\ \% : 100/100$<br>$0\ \% :\ 0/100$ |

| | | | |
|---|---|---|---|
| Precision | [%] | $Score(x_i) = x_i$ | 100 % : 100/100<br>0 % : 0/100 |
| Saturation | [%] | $Score(x_i) = x_i$ | 100 % : 100/100<br>0 % : 0/100 |

777

778 *Micro angiography Computed Tomography (microangio-CT) with µAngiofil*

779 The animal was heparinized and carotid arteries were cannulated. The brain was perfused with
780 saline (NaCl) at the animal's temperature to wash out the blood. Then the µAngiofil (Fumedica
781 AG, Switzerland) was perfused until the brain, tongue, and eyes became blue. After polymerization
782 of the µAngiofil (at least 30 minutes), the head was removed and fixed with paraformaldehyde
783 (PFA). Later, the brain was carefully extracted from the skull. The brain was scanned with a pre-
784 clinical micro-CT Siemens Inveon PET-CT (Siemens, Germany) with a voxel size of 37 $\mu m$.