



HAL
open science

Performance benchmarking of microbubble-localization algorithms for ultrasound localization microscopy

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

► To cite this version:

Baptiste Heiles, Arthur Chavignon, Vincent Hingot, Pauline Lopez, Elliott Teston, et al.. Performance benchmarking of microbubble-localization algorithms for ultrasound localization microscopy. *Nature Biomedical Engineering*, 2022, 6 (5), pp.605-616. 10.1038/s41551-021-00824-8 . hal-03965571

HAL Id: hal-03965571

<https://hal.science/hal-03965571v1>

Submitted on 31 Jan 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

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}, Elliott Teston², Olivier Couture¹

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

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

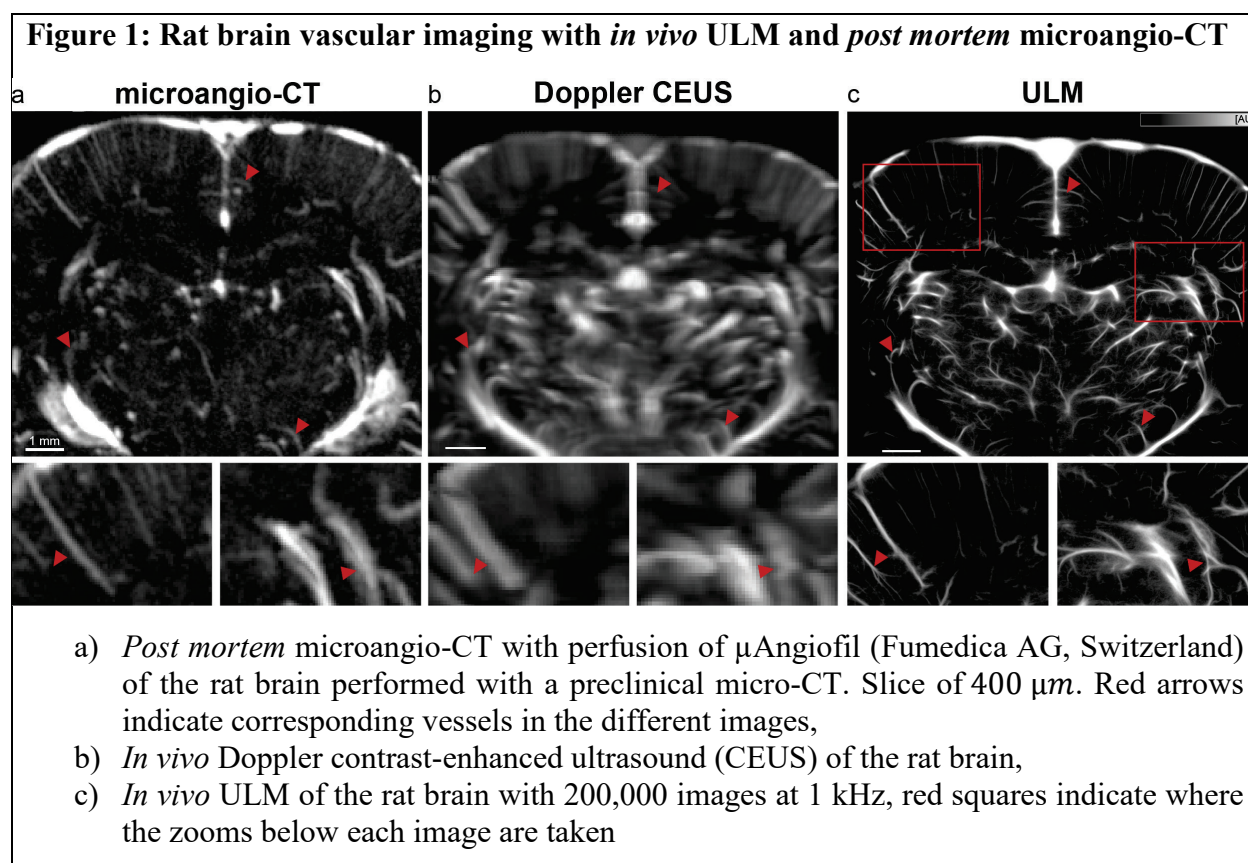
24 **Main**

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

31 Ultrasound imaging is extensively used in medical practice as a non-invasive tool that provides
32 soft-tissue diagnosis, prognosis, or guides interventions. Using the Doppler effect, it can also
33 measure blood flow in real-time. With the advent of plane wave techniques¹, ultrasound has
34 reached frame rates up to 20 kHz making it possible to observe and measure fast occurring
35 changes¹ such as functional changes in the brain², as well as increase sensitivity to blood flow that
36 allows more accurate filtering^{3,4}. To increase blood's contrast, micrometric gas microbubbles can
37 be intravenously injected *in vivo*. Contrast-Enhanced Ultrasound (CEUS) is mostly used for
38 perfusion studies and cardiac imaging⁵.

39 Because conventional ultrasound imaging, Doppler, and contrast-enhanced ultrasound all rely on
40 the propagation of sound waves, ultrasound imaging is largely limited in resolution by diffraction.
41 Recently, Ultrasound Localization Microscopy (ULM) has broken that limit by isolating a small
42 number of microbubbles as subwavelength sources in each image and localizing them with
43 micrometric precision⁶⁻⁸. Similar to PALM (PhotoActivated Localization Microscopy), it uses

44 these agents as individual sensors to map the region of imaging by accumulating thousands of
 45 events and adding them through density-based methods. When implemented at kilohertz frame
 46 rates, ULM can also retrieve the trajectories of microbubbles carried in the bloodstream and
 47 reconstruct several centimeters deep, hemodynamic maps of the vasculature both in 2D and 3D⁹⁻
 48 ¹¹, a feat rising to the challenge of multiscale imaging tools for both the vasculature and
 49 microvasculature. For example, typical imaging of a rat brain with craniectomy *in vivo* with
 50 Doppler contrast-enhanced ultrasound and ULM is shown in **Figure 1** next to a *post mortem*
 51 microangio-computed tomography (microangio-CT) for reference. Microvessels appear on each
 52 modality but sharper and brighter on the ULM rendering (red triangles on **Figure 1**).



53

54 Several teams have been implementing ULM using various localization techniques over the last
55 few years^{6,10,12-14}. The quality of ULM images is directly dependent on the signal-to-noise ratio of
56 the original images or detection filters but most importantly on the localization and rendering
57 techniques. However, because each team uses different *in vitro* and *in vivo* models as well as
58 different probes, ultrasound machines, and acquisition parameters, comparing these algorithms,
59 along with their general approaches, is difficult. Progress in the field of ultrasound super-resolution
60 is hindered by a lack of widely accepted comparison metrics, even though several attempts have
61 been published on filtering¹⁵, localization¹⁶⁻¹⁸, and tracking¹⁵ errors.

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

72

Localization method	Related articles	Comments
Gaussian fitting	(Ackermann and Schmitz, 2016; Luke et al., 2016; O'Reilly and Hynnen, 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.

73

74 In addition, we introduce two algorithms developed by our team called weighted average (WA)

75 and radial symmetry (RS). We provide all algorithms ready for use so that labs lacking the

76 expertise to implement ULM can contribute to the fast development of ultrasound super-resolution

77 imaging (see supplemental data). Importantly, we introduce metrics that address still unanswered

78 questions such as: how accurate and sensitive is the localization algorithm¹⁷, how much is ULM

79 affected by low SNR¹³, what affects the precision of the code, how well is it possible to reconstruct

80 difficult structures such as highly tortuous vessels, aneurysms, stenosis, where do the grid artifacts

81 come from³⁴, and finally how can 3D ULM be implemented without a dramatic increase in

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

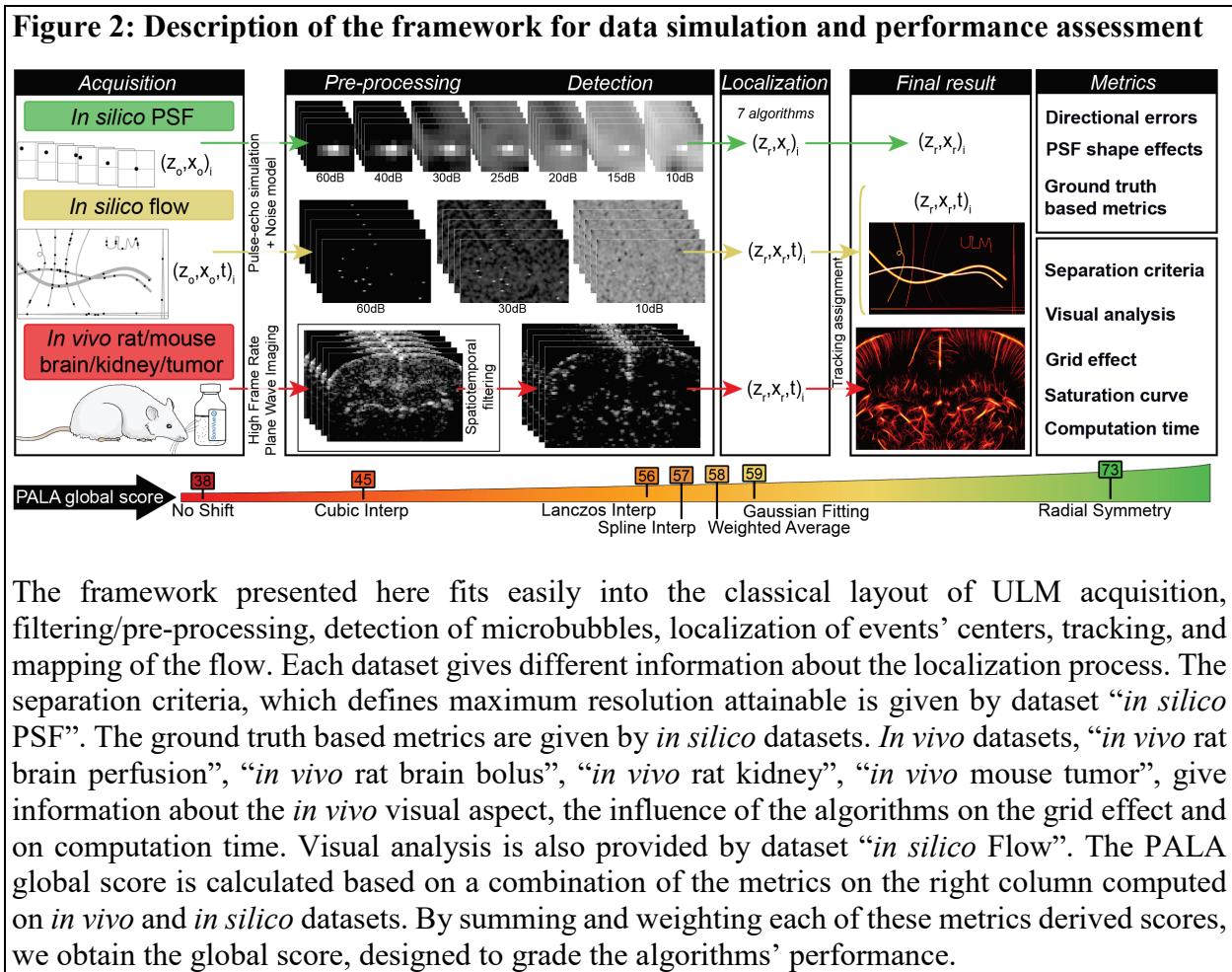
84 **Results**

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

93 *In silico and in vivo datasets*

94 To study the performances of different localization algorithms, the first step was to create and
95 acquire 3 different datasets, each of which we make fully available for reuse in supplementary
96 materials. The first one (“*in silico* PSF”), the point-spread function (PSF) simulation, is designed
97 to study non-uniformity effects on localization and spatial sampling effect on the beamforming
98 process. It consists of a point-like scatterer moved in a $\lambda \times \lambda$ space, the size of one pixel, by $\frac{\lambda}{21}$
99 steps, resulting in 441 positions. The second dataset (“*in silico* Flow”), the microcirculation
100 simulation, allows for testing *in silico* more fundamental aspects of resolution such as the
101 separability of microvessels and the accuracy of microbubble tracking. It is comprised of complex
102 3D tubes in the imaging plane mimicking an actual 2D imaging situation. Inside, point-like
103 scatterers are placed at random in each of the tubes section and then are propagated through 20,000

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



111

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

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

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

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

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

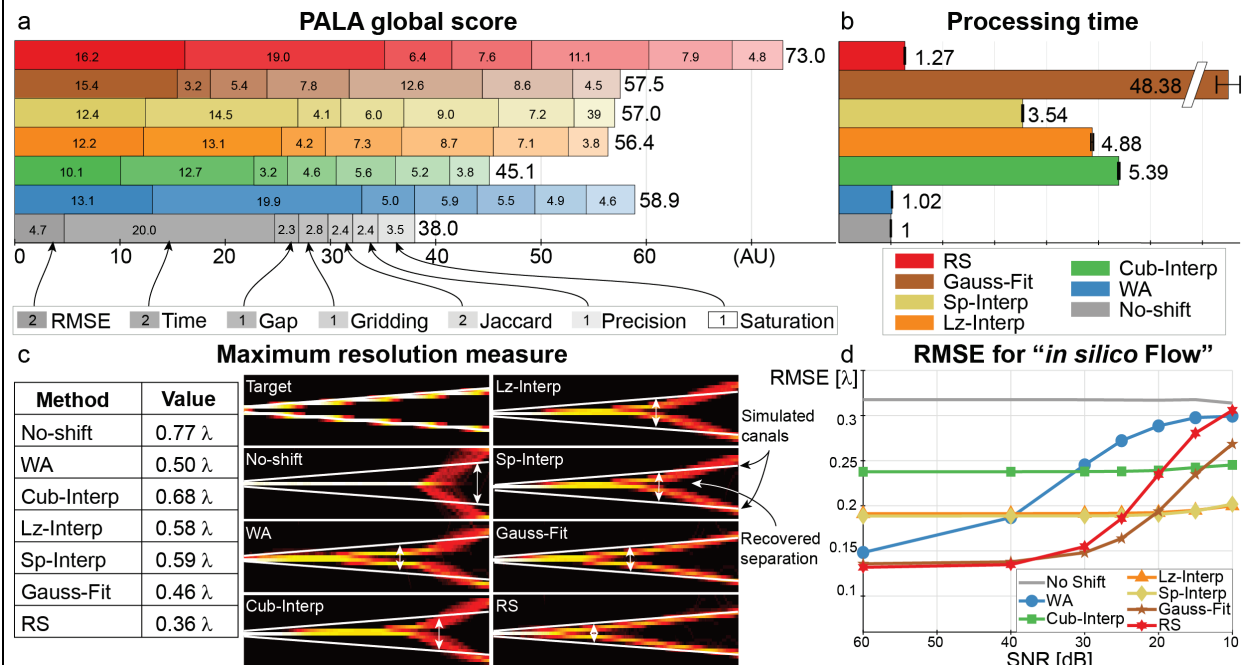
139 Munkres assignment to perform tracking and velocimetry^{36,37}. Final images are reconstructed by
140 density projection or velocity renderings. The whole framework is presented in **Figure 2**.

141 General comparisons of 7 algorithms using 11 metrics

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

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

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



- PALA global score obtained with the different scores from “*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.
- 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).
- Maximum resolution attainable on the dataset “*in silico* Flow” at $SNR = 30\text{ dB}$ 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.
- Root mean square error of the localization schemes for recovery of simulated points in dataset “*in silico* Flow” with different SNR.

159 Overall the radial symmetry (RS) based and weighted average (WA) algorithms showed the
160 highest performance with the weights chosen. Looking closer, the maximum resolution attainable
161 measured as the separation criteria is the highest for the RS and Gauss-Fit based algorithms at
162 0.36λ and 0.46λ . The time factor, however, is the lowest for the WA with under 3 minutes of
163 calculation which is lower than the acquisition time with our computer. The large increase and
164 very low root mean square error (RMSE) minimum of the weighted average algorithm (0.13λ)
165 confirm that it is a very good candidate for localization at a high signal-to-noise ratio (SNR) but
166 that one should switch to either Gauss-Fit or RS when the SNR worsens below 30 dB . At 15 dB
167 and downwards, one should theoretically consider using interpolation-based algorithms (see
168 ***Supplementary Figures 3-3***). The gridding index, which is a quantitative measure of the
169 reprojection of localized microbubbles on the original beamformed grid, impairing accurate vessel
170 delineation, is maximized for both RS and Gauss-Fit approaches. This gridding index increases
171 when specific spatial frequencies are overexpressed.

172 *Accuracy metrics in simulated datasets*

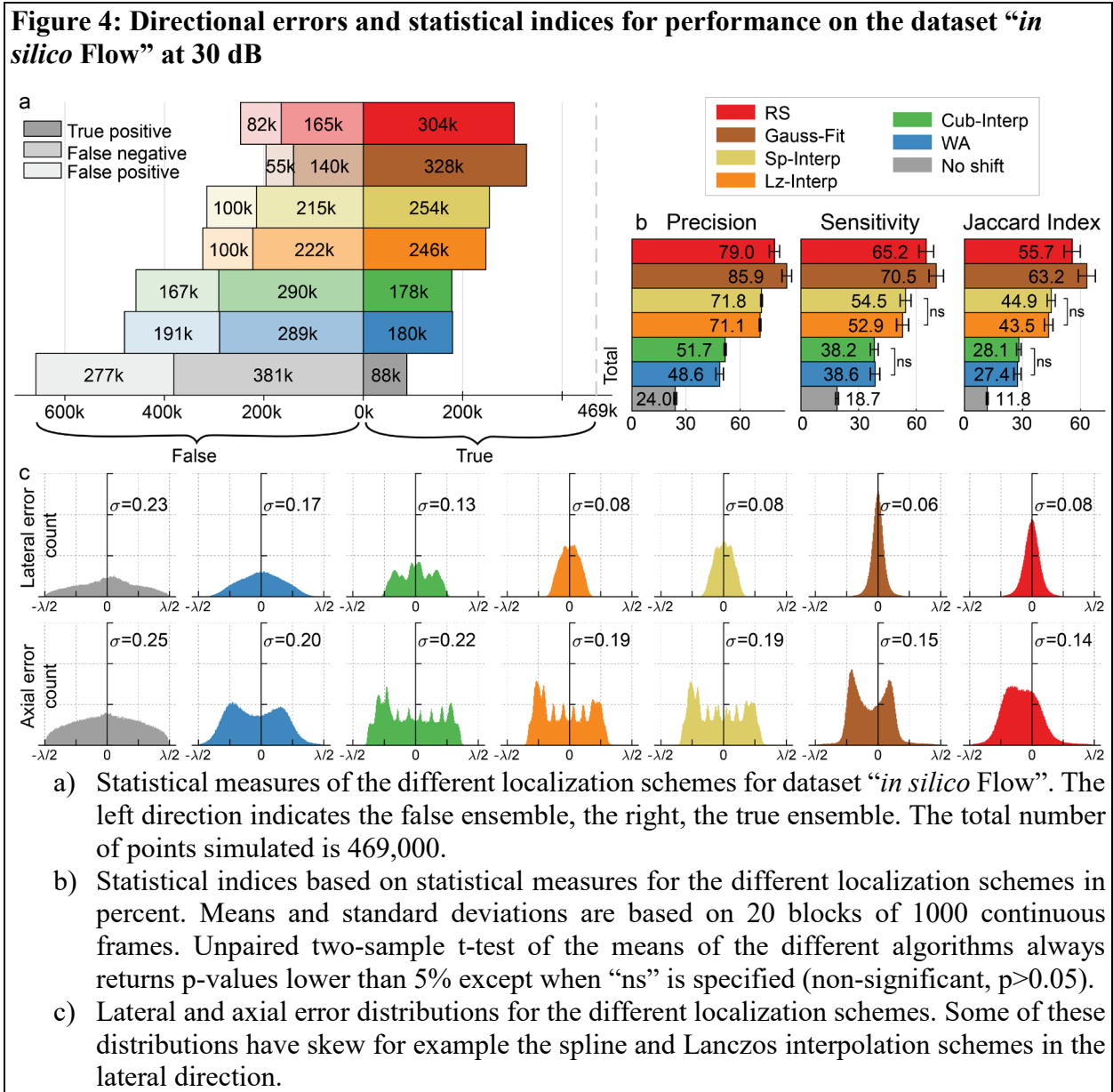
173 We simulated both *in silico* datasets with 7 different signal-to-noise ratios from 10 dB to 60 dB
174 (white Gaussian noise model - 0.2 ohm impedance noise, $\pm 10 \text{ dB}$ amplitude - for clutter filtered
175 with 0.7 pixel wide 2D Gaussian kernel, see ***Supplementary figure 2-3***). The different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

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

189 The sensitivities of WA and the best interpolation-based algorithms are quite low at around 40 %,
190 while the precision is higher for the latter at 71.8 % (see **Figure 4 b**). The no-shift algorithm has
191 the lowest values albeit not as low as one would expect given that no localization is involved. The
192 Jaccard index which represents a detection rate is at most 63.2 % and can be as low as 11.8 % for
193 the no-shift scheme. Gauss-Fit and RS are more precise than others. This is due to a low
194 localization error, and as such a high cardinality of true positives. As the total number of
195 $\{TP + FN\}$ is finite, this also means that the sensitivity index is high. For the weighted average
196 scheme, we measure a high precision and high sensitivity although its localization error is the
197 highest among the schemes. It gives us a glimpse into why this happens: the TP and FN values are
198 almost identical to the best interpolation schemes.

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

205 false negatives and a decrease in the number of true positives which can be verified in **Figure 4 a**,
 206 $TP_{WA} = 180k, TP_{Cub-interp} = 178k$ and $FN_{WA} = 289k, FN_{Cub-Interp} = 290k$



207

208 The largest standard deviation of the lateral error was found for the no-shift method as expected.

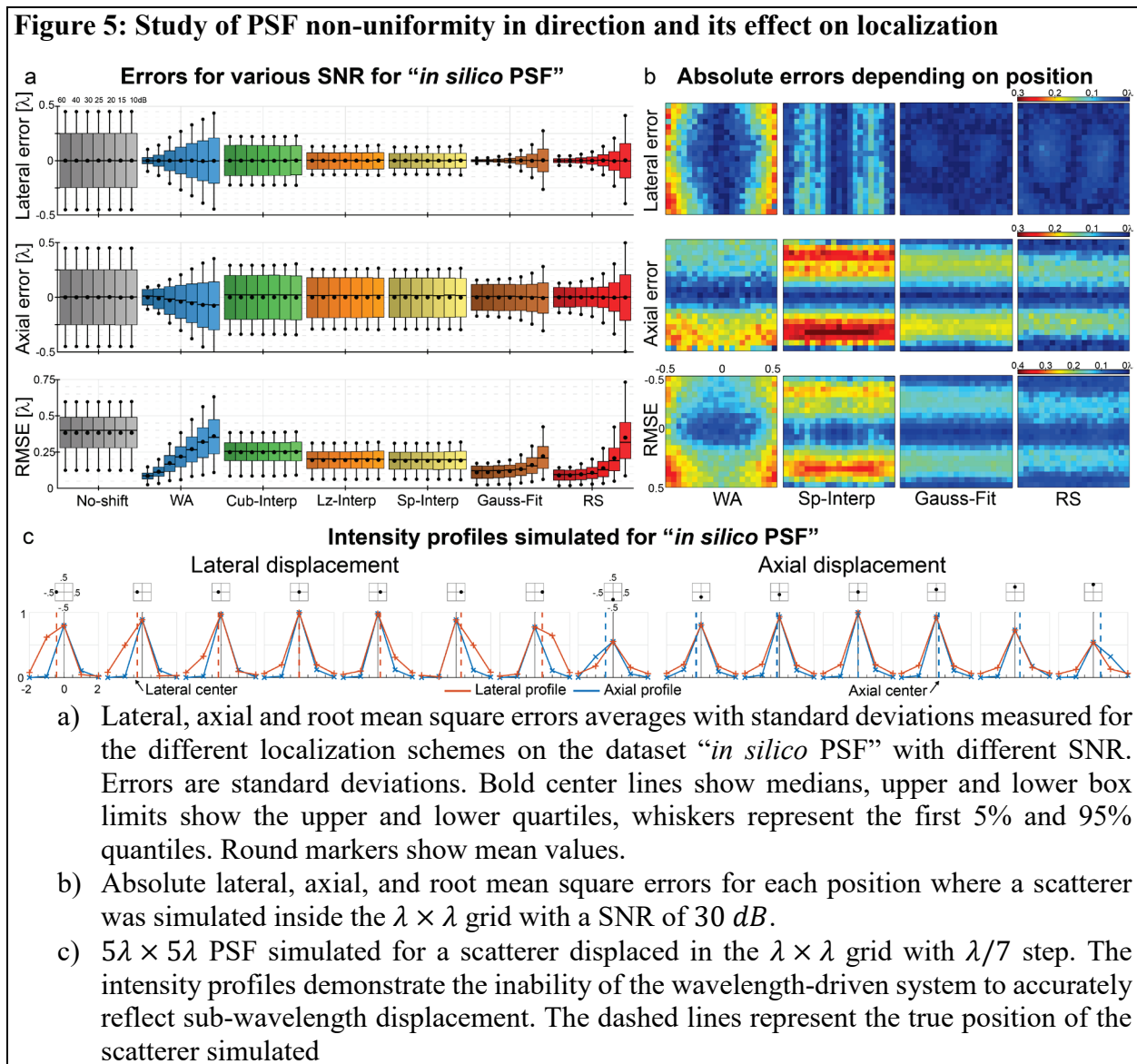
209 For the axial error, the algorithms have standard deviations between 0.06 and 0.23, with the lowest

210 being the GF and the highest being the no shift. For the Gauss-Fit and RS-based algorithm, their

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

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

234 distributed in two bands in each direction up until a certain SNR value. At 15 dB the weighted
 235 average exhibits this behavior very clearly for the lateral error while the GF and RS have an almost
 236 uniform distribution. All distributions and maps are presented in *Supplementary Figures 2-5 to*
 237 *2-6 and 3-3*. If we look more closely at the intensity profiles generated by our framework (**Figure**
 238 **5 c**), we can see that displacement in any direction induces a displacement of the maximum
 239 amplitude.



240

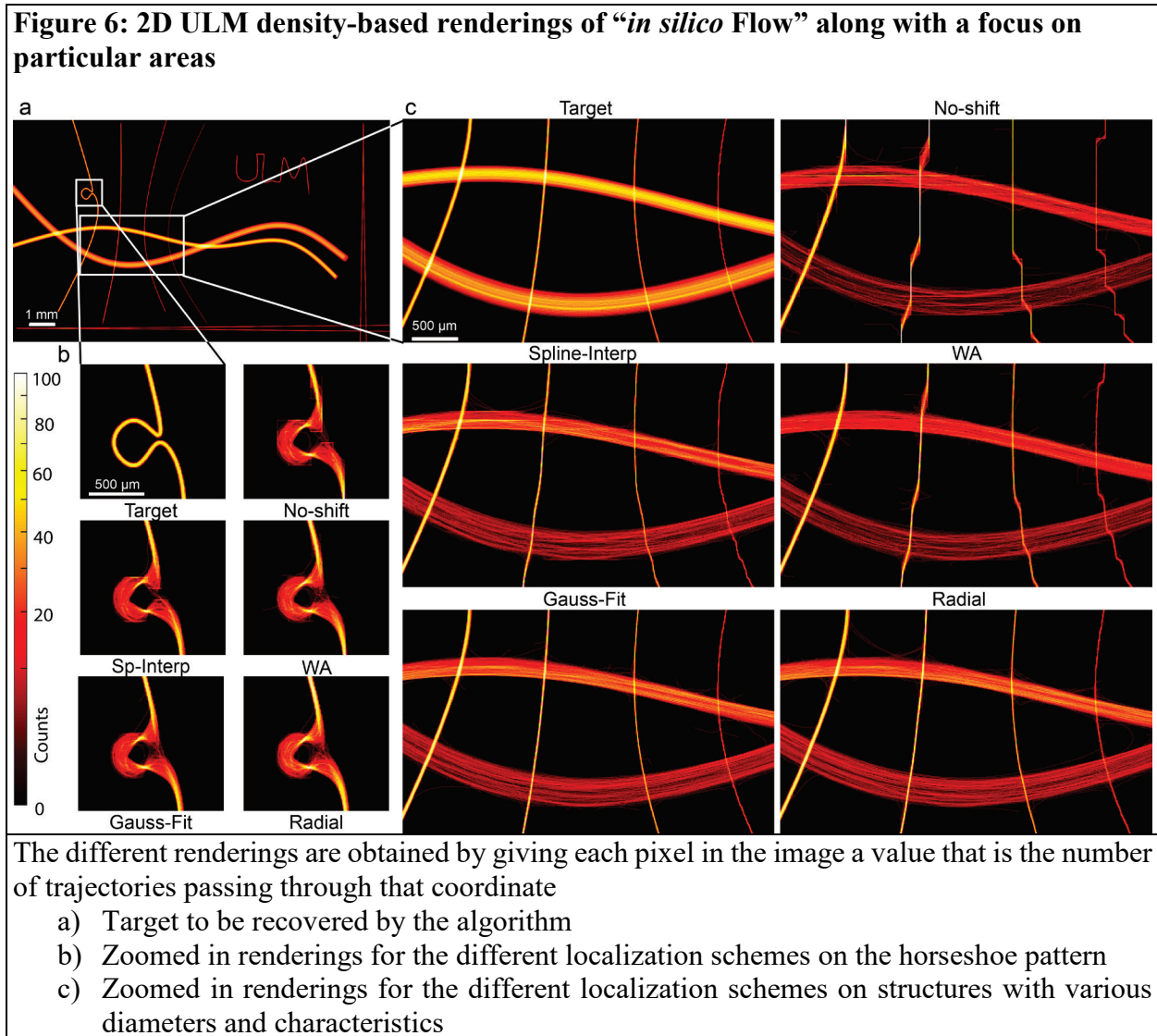
241 However, even when moving close to the neighboring pixel situated $\left(\frac{\lambda}{5}\right)$ further from the center,
242 the maximum amplitude remains at the center of the pixel $((x, z) = 0)$. The maximum drop in
243 intensity for each displacement is 12 dB below the peak amplitude on the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

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

254 Image quality metrics in silico and in vivo

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

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



266

267 The no-shift is valuable to study a hypothesis made in a previous paper about 3D ULM³⁷ stating
 268 that the grid effect appears because of erroneous localization in areas of low SNR. Because the no-
 269 shift does not involve sub-pixel localization, it should suffer the most from the grid effect and one
 270 can see that it does, appearing in all of the structures smaller than the wavelength of the ultrasound

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

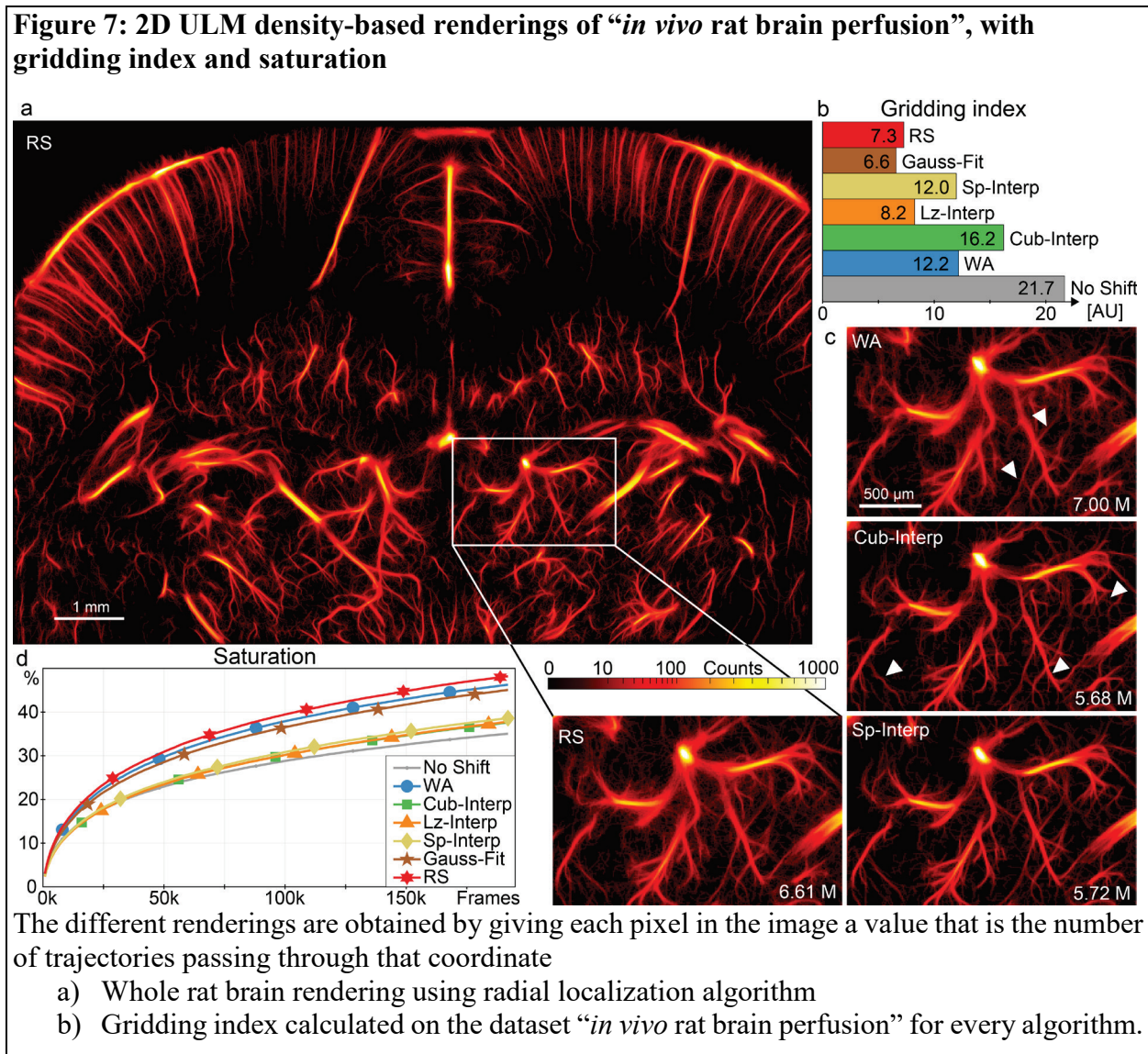
276 We devised an index to measure the grid-effect by a spatial frequency peak to baseline analysis in
277 both directions. The value of that index grows with the severity of gridding. The values are reported
278 in **Figure 7** next to the renderings of the brain vasculature. The no-shift-based ULM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

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

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

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

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



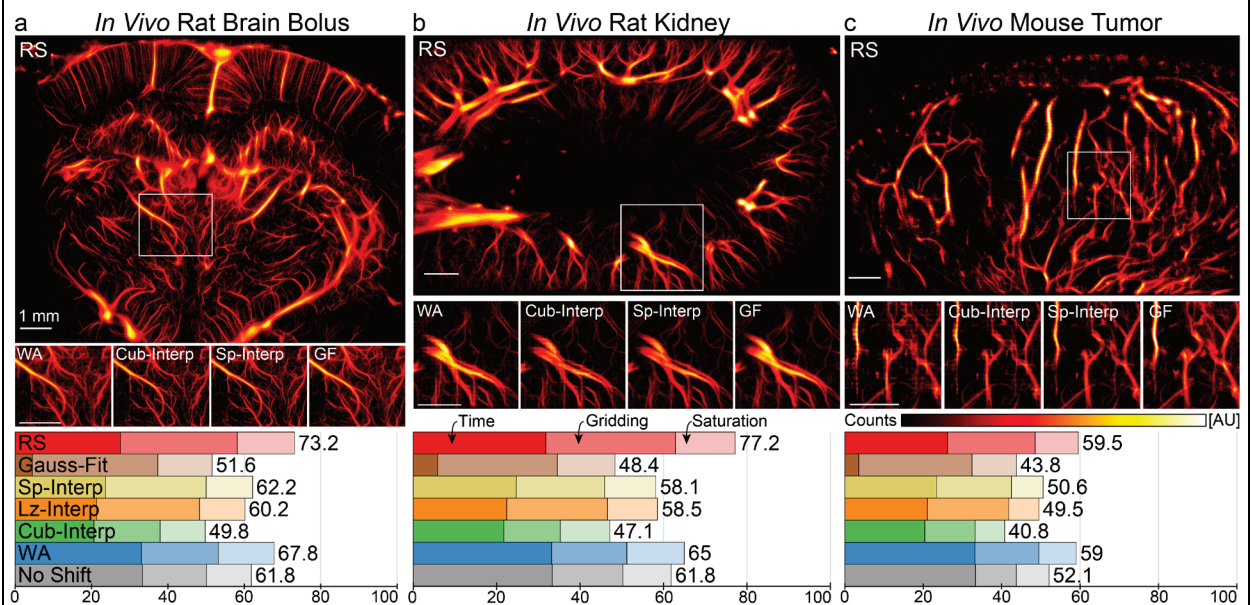
- c) Zoomed in portions of rat brain ULM using cubic interpolation, weighted average, spline interpolation, and radial symmetry-based localizations.
- d) Saturation coefficient calculated for each algorithm

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

301

302 The results obtained on the three other *in vivo* datasets are presented in **Figure 8** (see also
303 **Supplementary Figures 4-5 to 4-7**). The patterns described in the previous dataset still hold. In
304 the *in vivo* rat brain bolus, the superiority of the GF and RS algorithms is confirmed, the overall
305 score obtained by the RS is even higher than its score on the rat brain perfusion dataset. This
306 superiority and the GF algorithms' high-quality imaging is best illustrated with the tumor where
307 few vessels are present and are characterized by a disorganized structure. The GF and RS algorithm
308 depict evidently more vessels than the other schemes. The gridding effect is clearly seen in the
309 WA and Cub-Interp localization schemes. As for the Sp-Interp, its imaging quality is quite high
310 for example in the kidney, it is almost identical to the GF algorithm, but in the tumor dataset, we
311 can see clear evidence that it localizes fewer vessels than GF and RS. The partial score (obtained
312 from time, gridding, and saturation) difference between rat brain and tumor for the RS is 23%
313 while for the Weighted Average it is considerably less at 15%. This difference for the interpolation-
314 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.

- 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.
- 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
- 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.

315

316 Discussion

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

322 contrast agent types^{39,40}; ultrasound acquisition parameters²² as well as tracking algorithms³⁸ but
323 it is hard to draw general conclusions for the techniques as these papers do not use the same
324 datasets. We hope that with the datasets provided in this paper, comparison and accurate
325 characterization of all the steps will be feasible. Devising metrics to compare and rate specific
326 aspects of localization is a crucial step in providing a better understanding of super-resolution
327 imaging. Visual scoring by radiologists, neurologists, or physiologists will be required to establish
328 quality grading for each application. However, ULM will first need to gain better uniformity
329 through quantitative and objective metrics assessments.

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

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

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

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

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

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

374 As such, attention should be paid to the reconstruction grid used for rendering. Decreasing the
375 pixel size too much would increase the grid effect and the interpolation kernels, making the
376 calculations more time-consuming. *In vivo*, the interpolation factor depends on the maximum
377 theoretical resolution attainable, and as such, we advise that the super-resolved image pixel size
378 and the maximum theoretical resolution be matched. One can note that even though localization is
379 performed poorly in the no-shift scheme, the rendering still appears believable and that should be
380 kept in mind when evaluating the quality of ULM. Additionally, as was seen in the no-shift
381 localization scheme, if the reconstruction grid is too fine but the tracking algorithm contains an
382 interpolation of tracks or smoothing, the grid effect may disappear and some artifacts may appear.
383 These artifacts will lead to erroneous velocity estimations which can be measured and
384 characterized by looking at the Poiseuille distributed profiles for these small vessels. Another
385 difference to be noted is the lower gridding index calculated for WA compared to Sp-Interp
386 although in **Figure 7 b** the gridding effect is clearly higher. With 7.00M events localized in the *in*
387 *vivo* rat brain perfusion dataset, WA is the algorithm that localizes the most events (1.3M more
388 than Sp-Interp). The lower precision and higher RMSE compared to the Sp-Interp means that the
389 final rendering contains a lot more erroneous localizations thus decreasing the principal peaks'
390 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 microbubbles⁴¹. 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 microbubbles¹⁷. 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

437 enhancing the PSF and CNR, as well as the different tracking algorithms available for ULM.
438 However, thanks to the versatility of our framework, these can be easily included and benchmarked
439 *ad hoc*. This versatility can be put to use to develop new metrics. One very valuable metric would
440 be an index to benchmark the quality of ULM algorithms in between organs and in between
441 datasets coming from different groups. Comparing organs that have a very different vascular
442 organization, such as the kidney or the brain, remains a challenge. In 2D, the possibility to have
443 an atlas of the vasculature of the rat's brain could be explored but will probably run into out-of-
444 plane related issues (motion, size of the plane at focus, interrupted tracks, reprojection of vessels).
445 The noise simulated in the *in silico* dataset is a simple model aimed at imitating clutter noise.
446 Added to the beamformed images, it does not take into account jitter, false peak errors in the
447 radiofrequency data, readout noise from the electronics, nor does it represent potential motion. As
448 the *in vivo* data we have presented have been filtered with a spatiotemporal clutter filter, our aim,
449 by choosing this simplistic approach, was to reproduce SNR values and aspects seen in the *in vivo*
450 cases. However, thanks to the versatility of our platform, any noise model could be added to the
451 *in silico* dataset easily.

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

457 **Outlook**

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

472 **References**

- 473 1. Tanter, M. & Fink, M. Ultrafast imaging in biomedical ultrasound. *IEEE Trans. Ultrason.*
474 *Ferroelectr. Freq. Control* **61**, 102–119 (2014).
- 475 2. Macé, E. *et al.* Functional ultrasound imaging of the brain. *Nat. Methods* **8**, 662–664 (2011).
- 476 3. Macé, E. *et al.* High sensitivity brain angiography using ultrafast Doppler. in (2010).
477 doi:10.1109/ultsym.2010.5935810.
- 478 4. Demene, C. *et al.* Spatiotemporal Clutter Filtering of Ultrafast Ultrasound Data Highly
479 Increases Doppler and fUltrasound Sensitivity. *IEEE Trans. Med. Imaging* **34**, 2271–2285
480 (2015).
- 481 5. Frinking, P., Segers, T., Luan, Y. & Tranquart, F. Three Decades of Ultrasound Contrast
482 Agents: A Review of the Past, Present and Future Improvements. *Ultrasound Med. Biol.* **46**,
483 892–908 (2020).
- 484 6. Couture, O., Besson, B., Montaldo, G., Fink, M. & Tanter, M. Microbubble ultrasound super-
485 localization imaging (MUSLI). in 1285–1287 (IEEE, 2011).
486 doi:10.1109/ULTSYM.2011.6293576.
- 487 7. Olivier Couture, Mickael Tanter & Mathias Fink. Patent 889 Cooperation Treaty
488 (PCT)/FR2011/052810.
- 489 8. Couture, O., Hingot, V., Heiles, B., Muleki-Seya, P. & Tanter, M. Ultrasound Localization
490 Microscopy and Super-Resolution: A State of the Art. *IEEE Trans. Ultrason. Ferroelectr.*
491 *Freq. Control* **65**, 1304–1320 (2018).
- 492 9. Errico, C. *et al.* Ultrafast ultrasound localization microscopy for deep super-resolution
493 vascular imaging. *Nature* **527**, 499–502 (2015).

- 494 10. Christensen-Jeffries, K., Browning, R. J., Tang, M.-X., Dunsby, C. & Eckersley, R. J. In Vivo
495 Acoustic Super-Resolution and Super-Resolved Velocity Mapping Using Microbubbles. *IEEE*
496 *Trans. Med. Imaging* **34**, 433–440 (2015).
- 497 11. Heiles, B. Ultrafast Volumetric Ultrasound Localization Microscopy in vivo. in (2018).
- 498 12. Viessmann, O. M., Eckersley, R. J., Christensen-Jeffries, K., Tang, M. X. & Dunsby, C.
499 Acoustic super-resolution with ultrasound and microbubbles. *Phys. Med. Biol.* **58**, 6447–6458
500 (2013).
- 501 13. Lin, F. *et al.* 3-D Ultrasound Localization Microscopy for Identifying Microvascular
502 Morphology Features of Tumor Angiogenesis at a Resolution Beyond the Diffraction Limit of
503 Conventional Ultrasound. *Theranostics* **7**, 196–204 (2017).
- 504 14. O'Reilly, M. A. & Hynynen, K. A super-resolution ultrasound method for brain vascular
505 mapping: Super-resolution ultrasound method for brain vascular mapping. *Med. Phys.* **40**,
506 110701 (2013).
- 507 15. Song, P. *et al.* Improved Super-Resolution Ultrasound Microvessel Imaging With
508 Spatiotemporal Nonlocal Means Filtering and Bipartite Graph-Based Microbubble Tracking.
509 *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **65**, 149–167 (2018).
- 510 16. Brown, J. *et al.* Investigation of Microbubble Detection Methods for Super-Resolution
511 Imaging of Microvasculature. *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **66**, 676–691
512 (2019).
- 513 17. Christensen-Jeffries, K. *et al.* Microbubble Axial Localization Errors in Ultrasound Super-
514 Resolution Imaging. *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **64**, 1644–1654 (2017).
- 515 18. Kanoulas, E. *et al.* Super-Resolution Contrast-Enhanced Ultrasound Methodology for the
516 Identification of In Vivo Vascular Dynamics in 2D: *Invest. Radiol.* **54**, 500–516 (2019).

- 517 19. Ackermann, D. & Schmitz, G. Detection and Tracking of Multiple Microbubbles in
518 Ultrasound B-Mode Images. *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **63**, 72–82
519 (2016).
- 520 20. Luke, G. P., Hannah, A. S. & Emelianov, S. Y. Super-Resolution Ultrasound Imaging in Vivo
521 with Transient Laser-Activated Nanodroplets. *Nano Lett.* **16**, 2556–2559 (2016).
- 522 21. O'Reilly, M. A. & Hynynen, K. A super-resolution ultrasound method for brain vascular
523 mapping: Super-resolution ultrasound method for brain vascular mapping. *Med. Phys.* **40**,
524 110701 (2013).
- 525 22. Song, P., Manduca, A., Trzasko, J. D., Daigle, R. E. & Chen, S. On the Effects of Spatial
526 Sampling Quantization in Super-Resolution Ultrasound Microvessel Imaging. *IEEE Trans.*
527 *Ultrason. Ferroelectr. Freq. Control* **65**, 2264–2276 (2018).
- 528 23. Hansen, K. B. *et al.* Robust microbubble tracking for super resolution imaging in ultrasound.
529 in *2016 IEEE International Ultrasonics Symposium (IUS)* 1–4 (IEEE, 2016).
530 doi:10.1109/ULTSYM.2016.7728793.
- 531 24. Heiles, B. *et al.* Ultrafast 3D Ultrasound Localization Microscopy using a 32x32 Matrix Array.
532 *IEEE Trans. Med. Imaging* **13** (2019).
- 533 25. Soulioti, D. E., Espíndola, D., Dayton, P. A. & Pinton, G. Super resolution imaging through
534 the human skull. *ArXiv181110653 Cond-Mat Physicsphysics* (2018).
- 535 26. Viessmann, O. M., Eckersley, R. J., Christensen-Jeffries, K., Tang, M. X. & Dunsby, C.
536 Acoustic super-resolution with ultrasound and microbubbles. *Phys. Med. Biol.* **58**, 6447–6458
537 (2013).

- 538 27. Zhang, G. *et al.* Acoustic wave sparsely activated localization microscopy (AWSALM):
539 Super-resolution ultrasound imaging using acoustic activation and deactivation of
540 nanodroplets. *Appl. Phys. Lett.* **113**, 014101 (2018).
- 541 28. Zhu, J. *et al.* 3D Super-Resolution US Imaging of Rabbit Lymph Node Vasculature in Vivo
542 by Using Microbubbles. *Radiology* **291**, 642–650 (2019).
- 543 29. Huang, C. *et al.* Short Acquisition Time Super-Resolution Ultrasound Microvessel Imaging
544 via Microbubble Separation. *Sci. Rep.* **10**, 6007 (2020).
- 545 30. Parthasarathy, R. Rapid, accurate particle tracking by calculation of radial symmetry centers.
546 *Nat. Methods* **9**, 724–726 (2012).
- 547 31. Christensen-Jeffries, K. *et al.* 3-D In Vitro Acoustic Super-Resolution and Super-Resolved
548 Velocity Mapping Using Microbubbles. *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **64**,
549 1478–1486 (2017).
- 550 32. Desailly, Y., Couture, O., Fink, M. & Tanter, M. Sono-activated ultrasound localization
551 microscopy. *Appl. Phys. Lett.* **103**, 174107 (2013).
- 552 33. Desailly, Y., Pierre, J., Couture, O. & Tanter, M. Resolution limits of ultrafast ultrasound
553 localization microscopy. *Phys. Med. Biol.* **60**, 8723–8740 (2015).
- 554 34. Song, P., Manduca, A., Trzasko, J. D., Daigle, R. E. & Chen, S. On the Effects of Spatial
555 Sampling Quantization in Super-Resolution Ultrasound Microvessel Imaging. *IEEE Trans.*
556 *Ultrason. Ferroelectr. Freq. Control* **65**, 2264–2276 (2018).
- 557 35. Hingot, V. *et al.* Microvascular flow dictates the compromise between spatial resolution and
558 acquisition time in Ultrasound Localization Microscopy. *Sci. Rep.* **9**, (2019).
- 559 36. Kuhn, H. W. The Hungarian method for the assignment problem. *Nav. Res. Logist. Q.* **2**, 83–
560 97 (1955).

561 37. Heiles, B. *et al.* Ultrafast 3D Ultrasound Localization Microscopy using a 32x32 Matrix Array.
562 *Accept. IEEE Trans. Med. Imaging* 13 (2019).

563 38. Song, P. *et al.* Improved Super-Resolution Ultrasound Microvessel Imaging With
564 Spatiotemporal Nonlocal Means Filtering and Bipartite Graph-Based Microbubble Tracking.
565 *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **65**, 149–167 (2018).

566 39. Luke, G. P., Hannah, A. S. & Emelianov, S. Y. Super-Resolution Ultrasound Imaging in Vivo
567 with Transient Laser-Activated Nanodroplets. *Nano Lett.* **16**, 2556–2559 (2016).

568 40. Zhang, G. *et al.* Acoustic wave sparsely activated localization microscopy (AWSALM):
569 Super-resolution ultrasound imaging using acoustic activation and deactivation of
570 nanodroplets. *Appl. Phys. Lett.* **113**, 014101 (2018).

571 41. Lowerison, M. R., Huang, C., Lucien, F., Chen, S. & Song, P. Ultrasound localization
572 microscopy of renal tumor xenografts in chicken embryo is correlated to hypoxia. *Sci. Rep.*
573 **10**, 2478 (2020).

574 42. Forsberg, F., Leeman, S. & Jensen, J. A. Assessment of hybrid speckle reduction algorithms.
575 *Phys. Med. Biol.* **36**, 1539–1549 (1991).

576 43. Ledoux, L. A. F., Brands, P. J. & Hoeks, A. P. G. Reduction of the Clutter Component in
577 Doppler Ultrasound Signals Based on Singular Value Decomposition: A Simulation Study.
578 *Ultrason. Imaging* **19**, 1–18 (1997).

579 44. Desailly, Y. *et al.* Contrast enhanced ultrasound by real-time spatiotemporal filtering of
580 ultrafast images. *Phys. Med. Biol.* **62**, 31–42 (2017).

581

582

583 **Acknowledgments**

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

597 **Author information**

598 *Affiliations*

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

600 Baptiste Heiles, Arthur Chavignon, Vincent Hingot, Olivier Couture

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

602 Baptiste Heiles, Vincent Hingot, Eliott Teston, Pauline Lopez

603 *Institut Cochin, INSERM U1016, Paris, France*

604 Pauline Lopez

605 Contributions

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

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

613 Corresponding author

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

615 **Ethics declarations**

616 Competing interests

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

619 Animal experimentation

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

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

628 **Online methods**

629 Point Spread Function dataset (“in silico PSF”)

630 The Point-Spread Function (PSF) simulation is designed to study non-uniformity effects on
 631 localization and spatial sampling effect on the beamforming process. A point-like scatterer is
 632 simulated using the Verasonics Research Ultrasound Simulator (Verasonics Inc., Kirkland
 633 Washington, USA). The position of the scatterer is moved by $\frac{\lambda}{7}$ increments in both directions in a
 634 $\lambda \times \lambda$ space. For each of these positions, the radiofrequency response, as well as its beamformed
 635 image, was produced using 3 tilted plane waves emitted from a 15MHz linear probe, with 128
 636 elements spaced by one wavelength. The speed of sound was set to 1540 m/s.

637 In silico angiography dataset (“in silico Flow”)

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

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

640

641 To simulate moving microbubbles, we chose initial random positions in the tubes and recorded
642 them as point-like scatterers. For each of these positions, a trajectory calculated by Poiseuille's
643 law was assigned based on their position in the cross-section of the tube. We considered that
644 microbubbles could not jump from one trajectory to another and so had to follow the same
645 trajectory for as long as it is simulated. Using the Verasonics Research Ultrasound Simulator, we
646 simulated radiofrequency and beamformed images of a moving scatterer in a $7 \times 14.9 \text{ mm}^2$ area
647 insonified with 3 tilted plane waves with a 15MHz linear probe, with 128 elements spaced by one
648 wavelength. The simulation sequence was based on the Vantage example script (Vantage release
649 4.0.0) *FlashAngles* and adapted to save either radiofrequency data and beamformed images with
650 $\lambda \times \lambda$ pixel size. The speed of sound was set to 1540 *m/s*.

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

660 *In vivo* dataset (“*in vivo* rat brain perfusion” and “*in vivo* rat brain bolus”)

661 All animals received humane care in compliance European Community Council Directive
662 of 22nd September 2010 (010/63/UE) and approved by the institutional committee C2EA-59:
663 “Comité d’Ethique en matière d’Expérimentation Animale Paris Centre et Sud” under the protocol
664 2015–23. 8-10 weeks Sprague-Dawley rats were obtained from Janvier Labs (Le Genest-Saint-
665 Isle, France). The animals were kept for at least a week before surgery. Water and a commercial
666 pelleted diet SAFE A04-10 (Augy, France) were available *ad libitum*. SAFE FLAKE sawdust was
667 used as bedding, enrichment such as pieces of cardboard, paper tunnels were provided. After 5
668 minutes in the induction cage filled with a mix of air and 5% of Isoflurane, the animal was placed
669 on the back on a heated plate with a respiratory mask. The mix was replaced with O₂ mixed with
670 4% of Isoflurane. The depth of the anesthesia was tested by the absence of withdrawal reflex when
671 pinching the toes of the hind limb. A jugular vein catheter was then placed and craniotomy surgery
672 was used to remove the skull on a window in between the Lambda and Bregma planes and 14 mm
673 wide. 400 μ L of Sonovue microbubbles (Bracco Imaging®, Milan, Italy) were injected at a steady
674 rate of 80 μ L/min to keep a stable concentration in the bloodstream or with boli in the second
675 dataset. A continuous ultrasound acquisition comprised of 240 blocs (or 213 blocs for bolus
676 dataset) of 800 frames taken at a compounded frame rate of 1000 Hz with 3 tilted plane waves
677 $[-3; 0; 3]^\circ$ or $[-5; 0; 5]^\circ$ for bolus dataset) using a 15 MHz centered frequency probe
678 manufactured by Vermon®, Tours, France with a 0.1 mm pitch. Acquisitions lasted 4 min for the
679 perfusion dataset, and 3 min 30 sec for the bolus dataset. Elevation focusing is provided through a
680 plastic lens to reach 500 μ m at 8 mm. The radiofrequency data were beamformed on the
681 ultrasound machine’s GPU (SuperSonic Imagine, Aix-en-Provence, France) and transferred to an
682 external hard drive for later processing. The microbubbles’ SNR was measured on a filtered image
683 as the intensity ratio between the 10 brightest microbubbles and the average image intensity.

684 *In vivo dataset (“in vivo rat kidney”)*

685 All animals received humane care in compliance European Community Council Directive
686 of 22nd September 2010 (010/63/UE) and approved by the institutional committee C2EA-59:
687 “Comité d’Ethique en matière d’Expérimentation Animale Paris Centre et Sud” under the protocol
688 APAFIS # 16874-2017122914243628. The same housing protocol and strain of rat were used as
689 the animals used for the brain experiments. The kidney was exposed through a lumbotomy incision
690 and placed on an acoustically absorbing material to facilitate imaging. A heat lamp was used to
691 prevent hypothermia of the animal. Ultrasound gel was applied on top of the imaging area to
692 provide coupling, Sonovue microbubbles (Bracco Imaging ®, Milan, Italy) were injected in boli
693 via a jugular vein catheter. A continuous ultrasound acquisition comprised of 240 blocs of 800
694 frames taken at a compounded frame rate of 1000 Hz with 5 tilted plane waves
695 $[-10; -5; 0; 5; 10]^{\circ}$ using a 15 MHz centered frequency probe manufactured by Vermon®, Tours,
696 France with a 0.1 mm pitch. The acquisition lasted 4 minutes. Elevation focusing is provided
697 through a plastic lens to reach 500 μm at 8 mm.

698 *In vivo dataset (“in vivo mouse tumor”)*

699 All animals received humane care in compliance European Community Council Directive of 22nd
700 September 2010 (010/63/UE) and approved by the institutional committee C2EA-34: “Comité
701 d’Ethique en matière d’Expérimentation Animale Paris Descartes” under the protocol
702 APAFIS #25169-202008071746473. For this study, MMTV-PyMT transgenic mice were
703 backcrossed with FvB/NCrl mice (Charles River Laboratories, Wilmington Massachusetts, USA).
704 A transplanted model of the MMTV-PyMT was used for this study. Tumor implantation was
705 performed using primary tumors developed in donor mice. Donor mice with primary tumors of
706 ~10–15 mm were killed and tumors were pooled, dissociated, and then re-injected orthotopically

707 in receiver mice. Tumors were dissociated mechanically and incubated for 30 minutes at 37°C
708 with DNase I (100 µg/ml, Roche Diagnostics GmbH, Mannheim, Germany), collagenase (1mg/ml,
709 Roche Diagnostics GmbH, Mannheim, Germany), and hyaluronidase (1 µg/ml, Sigma-Aldrich, St.
710 Louis Missouri, USA) in RPMI with 2% serum. After red blood cell lysis and filtration on a 70
711 µm sieve, the cell suspension was rinsed 3 times in PBS and used for tumor transplantation. Thus,
712 10^6 cells isolated from tumors of MMTV-PyMT mice were injected in 100 µl of PBS in the
713 mammary fat pad of 8 weeks-old FvB mice. About 10 d after inoculation, transplanted PyMT
714 tumors reached a diameter of ~ 6mm. A continuous ultrasound acquisition comprised of 150 blocs
715 of 200 frames taken at a compounded frame rate of 500Hz with 5 tilted plane waves
716 $[-10; -5; 0; 5; 10]^\circ$ using a 15 MHz centered frequency probe manufactured by Vermon®, Tours,
717 France with a 0.1 mm pitch. The acquisition lasted 2 min 30 sec. Elevation focusing is provided
718 through a plastic lens to reach 500 µm at 8 mm.

719 ULM workflow and detection

720 ULM is a multi-step method. The general layout is presented in **Figure 2** and **Supplementary**
721 **Figure 4-8**. Firstly, one needs to acquire a large number of frames in the region of imaging. Then,
722 a detection process takes place to enhance the signal of microbubbles. Close microbubbles are
723 isolated. *In silico*, such a step is dispensed with as the signal-to-noise ratio is set high enough in
724 the simulation. For the *in vivo* dataset, we use the Singular Value Decomposition approach^{4,42,43}.
725 By rearranging a continuous set of beamformed images as a Casorati matrix ($N_z \times N_x, N_t$), we
726 will use the new basis given by the SVD to filter the tissue from the microbubbles according to its
727 spatiotemporal decorrelation⁴⁴. We remove the four highest energy vectors out of a basis of 800.

728 Tracking algorithm

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

736 Reconstruction

737 We build the final image by accumulating all the super-resolved positions of the microbubbles
 738 through time in a user defined pixel grid in a Maximum Intensity Projection fashion. The pixel
 739 intensity represents the number of tracks passing through this pixel (see **Supplementary Figure 4-**
 740 **9**).

741 In silico metrics

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

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

747
$$\delta_z(i) = z_i^0 - z_i; \delta_x(i) = x_i^0 - x_i$$

748
$$RMSE = \frac{1}{N} \sum_i^N \sqrt{(x_i - x_i^0)^2 + (z_i - z_i^0)^2}$$

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

759 *In vivo metrics*

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

765
$$F_z = FFT \left(\sum_x I(z, x) \right); F_x = FFT \left(\sum_z I(z, x) \right)$$

766 The peak to the baseline value of the two first peaks of the power spectral density are summed in
767 a logarithmic scale to quantify the amount of aliasing present in the image. For more details see

768 ***Supplementary Figure 3-16.***

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

771 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

772

773 Conversion metrics to scores

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

Metric	Unit	Conversion	Score Range
RMSE	Wavelengths	$Score(x_i) = 100 * (1 - \frac{x_i}{0.5\lambda})$	0 λ : 100/100 0.5 λ : 0/100
Processing Time	Seconds	$Score(x_i) = 100 * (1 - 0.5 * \log_{10}(\frac{x_i}{min(x)}))$	$min(x)$: 100/100 100 * $min(x)$: 0/100
Separation Index	Wavelengths	$Score(x_i) = 100 * (1 - \frac{x_i}{1\lambda})$	0 λ : 100/100 1 λ : 0/100
Gridding Index	dB	$Score(x_i) = 100 * (1 - \frac{x_i}{30})$	0 dB : 100/100 30 dB : 0/100
Jaccard	[%]	$Score(x_i) = x_i$	100 % : 100/100 0 % : 0/100

Precision	[%]	$Score(x_i) = x_i$	100 % : 100/100 0 % : 0/100
Saturation	[%]	$Score(x_i) = x_i$	100 % : 100/100 0 % : 0/100

777

778

Micro angiography Computed Tomography (microangio-CT) with μ Angiofil

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 μ Angiofil (Fumedica

781

AG, Switzerland) was perfused until the brain, tongue, and eyes became blue. After polymerization

782

of the μ Angiofil (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 μ m.