nature portfolio

Peer Review File

Three-Dimensional Diffractive Acoustic Tomography

Corresponding Author: Professor Junjie Yao

Parts of this Peer Review File have been redacted as indicated to remove third-party material.

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Attachments originally included by the reviewers as part of their assessment can be found at the end of this file.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
This manuscript introduces three-dimensional diffractive acoustic tomography (3D-DAT), which employs an off-the-shelf linear-array transducer to perform 3D ultrasound and photoacoustic imaging in deep tissues. 3D-DAT utilizes single-slit acoustic diffraction to achieve isotropic resolutions, and its GPU-powered fast focal line (FFL) image reconstruction dramatically improves imaging speed. This innovative approach provides superior anatomical, functional, and molecular imaging performance, as demonstrated in various animal models, including glass frogs, murine tumor models, and pregnant mice. This reviewer believes that the novelty and experimental validation of this manuscript are well organized. Please address a few comments below before acceptance.

- In line 75, please cite literature that provides specific examples/studies of overfitting or hallucination effects in deep learning-based PA imaging.

- Since 3D-DAT using single-slit diffraction is a core aspect of this paper, I recommend creating a supplementary note that explains the principles in detail to help readers from diverse fields understand better.

- For 3D imaging, does the 3D-DAT using the single-slit method require scanning the probe in the elevational direction? Traditional method seems to do. This is not clearly explained. If the probe is not scanned in the elevational direction, what is the maximum FOV and imaging time?

- In Supplementary Figure 1, how does the acoustic diffraction pattern using DAT and traditional method, in the x-y plane?

- The explanation for why traditional methods result in blurred images seems insufficient. Additionally, the traditional method appears to elongate in the elevational direction. It seems to be able to address by optimizing the beamforming algorithm?

- The manuscript appears very optimistic about 3D-DAT using a single slit. I think it would be a very complex system and difficult to reimplement. Have you compared traditional and 3D-DAT images of human?

- Are there any disadvantages or issues with this method, single-slit? Including this in the discussion could improve the manuscript.

Reviewer #2

(Remarks to the Author)
A three-dimensional diffractive acoustic tomography system has been developed using a linear array transducer and a slit for deep tissue imaging. Fast focal line image reconstruction has been employed with 3D-DAT for faster 3D image rendering. Extensive experiments have been performed to characterize the developed system's performance in terms of resolution, imaging speed, and field-of-view. Several animal models have been explored to demonstrate the imaging system's ability for functional and molecular imaging. The results are supportive to the conclusions claimed. Below are my

comments to further improve the manuscript.

1. In Supple Fig. 1, please explain why the signal from P1 decreases when there is a slit, while P2 signal increases. Is this a trade-off noticed for the sources on the elevational axis? If so, why does it happen. This trade-off needs to mentioned in the main manuscript. Also, please explain the underlying physics behind it.

2. It is a bit confusing what is the sparse matrix? In the results section its written it has size of $n \times m$ where n is the number of voxels in the reconstruction volume and m is the total number of signal samples from all transducer elements, whereas in SI Fig. 3, its written as S is m x n, where m is the total number of reconstruction voxels and n is the number of samples times the number of elements times the size of the synthetic aperture. Please clarify. Also, please explain clearly how do you calculate the sparse matrix?

3. Is sparse matrix and the delay matrix both same? If so, then why the size of the dimensions of the delay matrix are [nx ny nz ne nsa]?

4. How is the FFL different FL algorithm from author's previous paper?

5. At a single position of the transducer, do we obtain reconstructed frame as a 2D or 3D image?

6. Its not clear if 3 wavelengths of the laser illuminate the tissue simultaneously or sequentially? From, SI. Fig 2, is it intended that it is sequential? If so, then how do you perform delay operation b/w each wavelength? If not, then how do you unmix the signals from 3 wavelengths?

Reviewer #3

(Remarks to the Author)
Review of NCOMMS-24-34632-T: Three-Dimensional Diffractive Acoustic Tomography in Deep Tissues

Please see the attached PDF containing the review comments

Reviewer #4

(Remarks to the Author)
I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
The authors well address all of my previous comments. Further, the revised manuscript shows compelling results which have never done before with the conventional PA and US methods. I strongly recommend this manuscript for publication.

I have one suggestion: the authors show a series structural, functional, and molecular imaging of 3D DAT in various animals. Can you perform any human imaging like in human forearm or palm? Then, the potential of clinical translation would be highly feasible.

Reviewer #2

(Remarks to the Author)
All comments are addressed. The manuscript can be accepted for publication.

Reviewer #3

(Remarks to the Author)
The authors have addressed all reviewers' comments well and the paper can be accepted.

Reviewer #4

(Remarks to the Author)
I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

Three-Dimensional Diffractive Acoustic Tomography

Luca Menozzi [a*], Tri Vu [a*], Aidan J Canning [a*], Harshal Rawtani [b], Carlos Taboada [a,c], Marie Elise Abi Antoun [d], Chenshuo Ma [a], Jesse Delia [e], Van Tu Nguyen [a], Soon-Woo Cho [a], Jianing Chen [a], Theresa Charity [f,g], Yirui Xu [a], Phuong Tran [a,h], Jun Xia [i], Gregory M. Palmer[f,g], Tuan Vo-Dinh [a,j], Liping Feng [b], Junjie Yao [a,k]

[a] *Department of Biomedical Engineering, Duke University, Durham, North Carolina, USA*

[b] *Duke University School of Medicine, Durham, North Carolina, USA*

[c] *Department of Biological Sciences, Vanderbilt University, Nashville, Tennessee, USA*

[d] *Tufts Medical Center, Boston, Massachusetts, USA*

[e] *American Museum of Natural History, New York City, New York, USA*

[f] *Department of Radiation Oncology, Duke University School of Medicine, Durham, NC 27710, USA*

[g] *Department of Surgery, Duke University School of Medicine, Durham, NC 27710, USA*

[h] *Department of Biomedical Engineering, North Carolina State University, Raleigh, North Carolina, USA*

[i] *Department of Biomedical Engineering, University at Buffalo, Buffalo, New York, USA*

[j] *Department of Chemistry, Duke University, Durham, NC 27708, USA*

[k] *Department of Neurology, Duke University of School of Medicine, Durham, NC 27710, USA*

The authors greatly appreciate the time and efforts of the editor and reviewers for providing helpful comments and suggestions that have improved the manuscript. Please find the point-by-point responses below. All changes have been colored with red font in the revised manuscript.

Response to Reviewer 1:

Reviewer #1 (Remarks to the Author):

This manuscript introduces three-dimensional diffractive acoustic tomography (3D-DAT), which employs an off-the-shelf linear-array transducer to perform 3D ultrasound and photoacoustic imaging in deep tissues. 3D-DAT utilizes single-slit acoustic diffraction to achieve isotropic resolutions, and its GPU-powered fast focal line (FFL) image reconstruction dramatically improves imaging speed. This innovative approach provides superior anatomical, functional, and molecular imaging performance, as demonstrated in various animal models, including glass frogs, murine tumor models, and pregnant mice. This reviewer believes that the novelty and experimental validation of this manuscript are well organized. Please address a few comments below before acceptance.

Response: We sincerely appreciate the reviewer's very positive comments and support.

1. In line 75, please cite literature that provides specific examples/studies of overfitting or hallucination effects in deep learning-based PA imaging.

Response: We thank the reviewer for the great suggestion. We have reworded the statement to be more specific to the challenges of deep learning for enhancing 3D PA/US image quality. Additionally, we have included four new citations that are more specific to deep-learning PA on the overfitting/hallucination effects, which are also provided below.

1. Zheng, Wenhan, et al. "Deep learning enhanced volumetric photoacoustic imaging of vasculature in human." *Advanced Science* 10.29 (2023): 2301277.
2. Hauptmann, Andreas, and Ben Cox. "Deep learning in photoacoustic tomography: current approaches and future directions." *Journal of Biomedical Optics* 25.11 (2020): 112903-112903.
3. N. Wang and J. Yao, "Sound Out the Deep Clarity: Super-resolution Photoacoustic Imaging at Depths," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control,* 2024
4. H. Deng, H. Qiao, Q. Dai, and C. Ma, "Deep learning in photoacoustic imaging: a review," *Journal of Biomedical Optics,* vol. 26, no. 4, pp. 040901-040901, 2021

2. Since 3D-DAT using single-slit diffraction is a core aspect of this paper, I recommend creating a supplementary note that explains the principles in detail to help readers from diverse fields understand better.

Response: We thank the reviewer for the helpful comment. We have added a supplementary note (**Supplementary Note 1**) on single-slit diffraction, and how the acoustic diffraction helps to improve resolution of 3D-DAT.

3. For 3D imaging, does the 3D-DAT using the single-slit method require scanning the probe in the elevational direction? Traditional method seems to do. This is not clearly explained. If the probe is not scanned in the elevational direction, what is the maximum FOV and imaging time?

Response: We thank the reviewer for the important question. Yes, 3D-DAT requires elevational scanning. The scanning process for 3D-DAT is identical to that of traditional imaging. In traditional imaging, only a transducer is scanned along the elevational axis. In 3D-DAT, both the transducer and single-slit are scanned together along the elevational axis (the relative position between the transducer and slit always remains the same). This was described in lines 110-111 and **Figure 1** lines 640-642 of the original submission.

4. In Supplementary Figure 1, how does the acoustic diffraction pattern using DAT and traditional method, in the x-y plane?

Response: We thank the reviewer for the great question. Because the single-slit provides a narrow acoustic opening only along the elevational (y) axis, the diffraction only occurs along the y-axis. There is essentially no diffraction along the lateral (x) axis, as the slit aperture is completely open along the x-axis. We have included discussion of this in the newly-added **Supplementary Note 1** and corresponding **Supplementary Figure 2**.

5. The explanation for why traditional methods result in blurred images seems insufficient. Additionally, the traditional method appears to elongate in the elevational direction. It seems to be able to address by optimizing the beamforming algorithm?

Response: The traditional imaging method results in blurred images due to the limited detection angle of the linear-array transducer with elevational focusing. This limited detection angle causes the elongation and blurring of structures along the elevational direction. The new materials in **Supplementary Note 1** and **Supplementary Figure 2** address the source of this problem in more detail. The beamforming method is fundamentally limited by the detection angle along the elevational axis. As such, a 3D beamforming algorithm applied to a traditional linear-array setup (without slit) would provide little if any improvement in image quality. However, with 3D-DAT, the detection angle along the elevational axis is physically enlarged, which directly results in the improved elevational resolution.

6. The manuscript appears very optimistic about 3D-DAT using a single slit. I think it would be a very complex system and difficult to reimplement. Have you compared traditional and 3D-DAT images of human?

Response: We thank the reviewer for the important comment. We have included new data showing the implementation of 3D-DAT to a different commercially available linear-array transducer (GE 9L-D) to show that the method can be applied to other traditional PA/US imaging systems (**Supplementary Figure 14**). We agree that human data for comparison would benefit this study and show the clinical potential of 3D-DAT. However, we are still in the process of applying for our institutional review board (IRB) approval for human imaging, and as such are unable to include human data in this study. While we do believe that this method is clinically translatable, we have removed statements attesting to the clinical translatability of the method throughout the manuscript, as we do not yet have data to support this claim.

7. Are there any disadvantages or issues with this method, single-slit? Including this in the discussion could improve the manuscript.

Response: We thank the reviewer for the great suggestion. There are three main technical tradeoffs with 3D-DAT.

Firstly, when performing two-dimensional (2D) imaging, the inclusion of the slit will make the image quality much worse than the traditional imaging without the slit. This is due to the slit effectively functioning as a negative acoustic lens by broadening the acoustic beam (both transmission and receiving). This broadening is beneficial for 3D imaging, although disadvantageous for 2D imaging. To address this tradeoff, we implemented the dynamic slit for 3D-DAT which is able to be precisely controlled via a fast motorized stage. Doing so, we can retain both 3D imaging and 2D imaging, by opening the slit for 2D imaging and closing the slit for 3D imaging.

Secondly, as we have discussed extensively in the manuscript, 3D-DAT partially blocks the on-axis acoustic energy that could have arrived at the transducer at each scanning position, compared to the in-focus signals in traditional imaging without a slit. To address this tradeoff, we perform 3D image reconstruction, taking advantage of the larger detection angle with a slit present, which ultimately leads to comparable SNR and CNR.

Thirdly, to accommodate the slit's large detection angle with 3D-DAT, we need to develop a more computationally-expensive 3D reconstruction. To address this tradeoff, we have developed the FFL reconstruction, which optimized the implementation of the focal line reconstruction concept and accelerated the image reconstruction speed to several seconds for an FOV of 40 x 14 x 15 mm^3.

Per the reviewer's suggestion, we have included further discussion on these points in the manuscript on page 16.

Response to Reviewer 2:

Reviewer #2 (Remarks to the Author):

A three-dimensional diffractive acoustic tomography system has been developed using a linear array transducer and a slit for deep tissue imaging. Fast focal line image reconstruction has been employed with 3D-DAT for faster 3D image rendering. Extensive experiments have been performed to characterize the developed system's performance in terms of resolution, imaging speed, and field-of-view. Several animal models have been explored to demonstrate the imaging system's ability for functional and molecular imaging. The results are supportive to the conclusions claimed. Below are my comments to further improve the manuscript.

Response: We sincerely appreciate the reviewer's positive comments and support.

1. In Supple Fig. 1, please explain why the signal from P1 decreases when there is a slit, while P2 signal increases. Is this a trade-off noticed for the sources on the elevational axis? If so, why does it happen. This trade-off needs to mentioned in the main manuscript. Also, please explain the underlying physics behind it.

Response: The decrease in the on-axis P1 signal is due to the slit, which partially blocks the passage of the acoustic energy. However, due to the increased detection angle with the slit's diffraction effect, the off-axis P2 signal becomes stronger. This improved detection angle is the fundamental advantage of 3D-DAT. The improvement in resolution due to single-slit diffraction is explained in the newly-added **Supplementary Note 1** and **Supplementary Figure 2**. Additionally, **Supplementary Figure 8** includes a new simulation quantifying the decrease in the in-focus acoustic pressure and the increase in the detection angle. The main results of this new simulation are also included in **Figure 2**, and presented in **Figure R1** below.

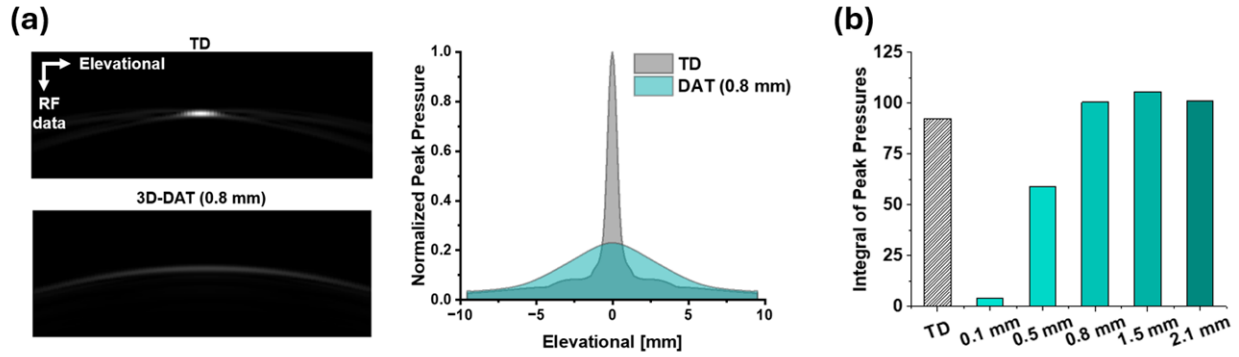

Figure R1. The impact of the single-slit on the PA signals. *(**a**) While the peak pressure for an in-focus target is higher in the traditional imaging compared to 3D-DAT, the off-axis pressure is higher with 3D-DAT. (**b**) In the reconstruction (FFL), the pressures are integrated along the elevational axis. Thus, taking the integral of the peak pressures along the elevational axis, we can see that the total received pressure for 3D-DAT at 0.8 mm slit width is nearly the same as that with the traditional method without a slit, which is why the SNR remains similar between the two methods.*

2. It is a bit confusing what is the sparse matrix? In the results section its written it has size of $n \times m$ where n is the number of voxels in the reconstruction volume and m is the total number of signal samples from all transducer elements, whereas in SI Fig. 3, its written as S is m x n, where m is the total number of reconstruction voxels and n is the number of samples times the number of elements times the size of the synthetic aperture. Please clarify. Also, please explain clearly how do you calculate the sparse matrix?

Response: We thank the reviewer for pointing out this contradiction and we apologize for the confusion in our previous writing. We have updated the results section to clarify the structure of the sparse matrix, and it now matches the correct description in **Supplementary Figure 5**. The calculation of the sparse matrix is based on the focal line concept [1], and geometrically shown in **Figure 1e**. The propagation distance from each voxel to each transducer element considers the propagation path through the slit aperture (shown by the black arrows of **Figure 1e**). **Supplementary Figure 5** shows how the

sparse matrix is arranged such that a single matrix multiplication generates the fully reconstructed data. More information on sparse matrix multiplication based beamforming can be found in the newly added citation [2].

[1] Xia, Jun, et al. "Three-dimensional photoacoustic tomography based on the focal-line concept." *Journal of biomedical optics* 16.9 (2011): 090505-090505.

[2] Perrot, Vincent, et al. "So you think you can DAS? A viewpoint on delay-and-sum beamforming." Ultrasonics 111 (2021): 106309.

3. Is sparse matrix and the delay matrix both same? If so, then why the size of the dimensions of the delay matrix are [nx ny nz ne nsa]?

Response: We thank the reviewer for this question. The sparse matrix and delay matrix are both used in the beamforming operation, however they are not the same. For the delay matrix, each value in the matrix corresponds to a certain index in the RF data that represents the travel time of the PA pressure wave. Thus, the values could range from 1 to Ns, where Ns is the number of time samples, and the reconstruction is performed by indexing the RF data using the delay matrix. Conversely, for the sparse matrix, the location of the nonzero values in the matrix corresponds to an index in the RF data. The values in the sparse matrix are either 0 or 1 (the matrix is highly sparse, >99% of the values are 0) and the reconstruction is performed by multiplying the RF data with the sparse matrix. This operation, especially when performed on a GPU, can be far more parallelized and efficient than the indexing operation, and thus allows for a fast reconstruction time. Further clarification on the difference between the sparse matrix and delay matrix has been added to **Supplementary Note 2**.

4. How is the FFL different FL algorithm from author's previous paper?

Response: We thank the reviewer for the good question. In principle, FFL is the same as our previously described FL reconstruction [1], however FFL is a much more efficient implementation. The reconstruction is formulated as a sparse beamforming matrix [2] that is directly multiplied with the RF data to generate the beamformed image, additionally accelerated by GPU processing. To improve parallelization, the nonzero entries of the columns represent each of the elements of the transducer array (n_e) at each of the positions that the transducer is scanned (n_{sa}). Additionally, our investigation of the synthetic aperture width (**Supplementary Figure 6**) further improves the efficiency of the reconstruction, because it reveals that only a synthetic aperture width of a few millimeters is necessary to achieve the optimal SNR and resolution enhancement. These factors combined to allow FFL to achieve nearly two orders of magnitude (10 seconds vs 720 seconds) reduction in computation time compared to the FL reconstruction (shown in the previously published table below), for the same number of voxels [3]. This dramatically improves the applicability of the FFL method.

[panel a redacted]

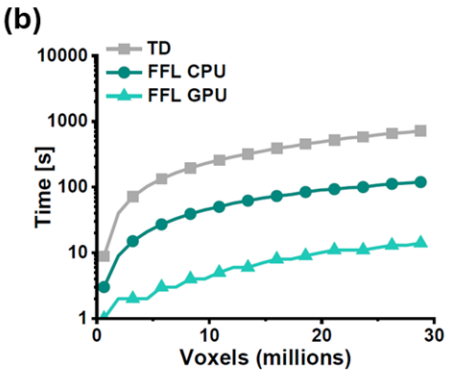

Figure R2. Comparison of previously published 3DFL reconstruction with FFL. *(**a**) Table showing recently published [1] 3DFL reconstruction times with two different sets of computing hardware. The computation times are shown for a reconstruction of 26,875,000 voxels. (**b**) Computation comparisons of delay matrix beamforming, FFL using a CPU, and FFL using a GPU. For the same number of voxels as shown in (**a**), GPU-accelerated FFL requires only ~10 seconds.*

[1] Xia, Jun, et al. "Three-dimensional photoacoustic tomography based on the focal-line concept." *Journal of biomedical optics* 16.9 (2011): 090505-090505.

[2] Perrot, Vincent, et al. "So you think you can DAS? A viewpoint on delay-and-sum beamforming." Ultrasonics 111 (2021): 106309.

[3] Zhang, Huijuan, et al. "Deep-E: A fully-dense neural network for improving the elevation resolution in linear-array-based photoacoustic tomography." *IEEE transactions on medical imaging* 41.5 (2021): 1279-1288.

5. At a single position of the transducer, do we obtain reconstructed frame as a 2D or 3D image?

Response: The transducer and slit are scanned across the entire sample, and once all the data is collected, we can perform the 3D FFL reconstruction. At each scanning position, we receive signals from a 3D volume, but the signals at a single position are not sufficient for performing image reconstruction, neither 2D nor 3D. 3D-DAT is effectively a synthetic aperture imaging method. We have clarified this point in the manuscript on page 5.

6. Its not clear if 3 wavelengths of the laser illuminate the tissue simultaneously or sequentially? From, SI. Fig 2, is it intended that it is sequential? If so, then how do you

perform delay operation b/w each wavelength? If not, then how do you unmix the signals from 3 wavelengths?

Response: We thank the reviewer for the important question. The laser excitation is sequential, separated by 10 ms per laser pulse. This is done by triggering each of the lasers separately. The RF data from each of the laser pulses are then reconstructed separately. This process is shown in **Figure 1a, d**. To avoid confusion, we have modified the caption of **Supplementary Figure 4** to clarify that the signal acquisition of the three wavelengths is sequential.

Response to Reviewer 3:

Summary of the paper

The authors present the use of a diffraction slit, created by two opposing glass slides that move along the centre of a 2D US detection array. A photoacoustically (PA) or ultrasonically (US) induced sound wave is re-focussed through the slit, travelling directly towards the detection elements afterwards. In the US configuration, the sound waves pass the diffraction slit twice, leading to an initial broadening of the sound waves. It is an effective idea which successfully increases the lateral resolution of acquired PA and US images.

The authors present a whole suite of experiments to demonstrate the capabilities of their device design, especially focusing on the improved elevational resolution of their system. For this, they include phantom experiments, but also pre-clinical small animal experiments that cover a range of purposes: characterising the resolution of the device and showing that it can characterise time-resolved phenomena as well as functional parameters; potentially with higher quantitative accuracy due to the higher resolution and therefore less pronounced partial volume/blurring effects.

Summary of the main strengths and weaknesses of the paper

The main strength of the paper is the comprehensive suite of testing experiments. Not only did they reproduce experiments from the literature, but also included a completely novel study on the investigation of PFAS and their influence on embryonal development. The manuscript is well written and the introduction strikes a great balance of catering to both a general and specialised audience by introducing the main concepts in an appropriate level of detail. The authors also (generally) go into great detail to describe their methods, leading to a high chance that many aspects of this work will be reproducible in the future.

Response: We sincerely appreciate the reviewer's positive comments and support.

The main drawback of the work is, that the claims and conclusions are at times not clearly supported. Examples:

- The idea of using a diffraction slit is not novel but has already been proposed by [Wang, Y et al. 2017], though the authors claim that they *"develop a novel technique for 3D PA/US imaging using an off-the-shelf linear-array transducer—3D diffractive acoustic tomography (3D-DAT)"*.

Response: The reviewer is correct. The previous works using a diffraction slit were cited in the original submission. We would like to highlight that our novelty includes the diffractive ultrasound imaging (not published before), the thorough characterization and validation of the technique through simulation, phantom, and in-vivo imaging (both for PA and US), the enhanced efficiency of the reconstruction method (making the technique practical), and the experimental applications in functional and molecular imaging.

- The authors compare the computation time of the sparse array reconstruction method with a conventional delay-and-sum approach, which is inherently not as optimised.

Response: Per the reviewer's suggestions, the section on the performance of FFL has been moved to the supplementary materials. In the manuscript, the comparisons in computation time are made directly to the most recently published data on reconstruction speed of the 3DFL algorithm [1].

[1] Zhang, Huijuan, et al. "Deep-E: A fully-dense neural network for improving the elevation resolution in linear-array-based photoacoustic tomography." *IEEE transactions on medical imaging* 41.5 (2021): 1279-1288.

- Traditional method images look confounded by artefacts and of poor quality, not just in the lateral resolution, but also in the imaging plane, where many discontinuities can be visualised (*e.g.* in Figures 3 b and f) which were not seen in past work published by the authors.

Response: The discontinuities in the imaging plane are mostly due to breathing motion artifacts in our *in vivo* experiments. These discontinuities are often prominent when performing traditional 3D imaging with a linear-array transducer.

- The authors use imprecise language in the text, *e.g.* "optimal", "greatly", "superior", etc., without providing quantifications and details how and why this was achieved.

Response: We thank the reviewer for pointing this out. We have added further quantifications to support statements made throughout the manuscript when applicable, as well as removed imprecise language throughout.

- In the introduction and conclusion, the authors claim the potential for clinical translation of the method, but none of the experiments directly support this claim.

Response: While we do see the clinical translatability of 3D-DAT, we do not yet have human data to directly support this claim. As such, we have removed claims relating to clinical translation from the manuscript.

While the impressive number of experiments makes the testing of the methodology quite comprehensive, it simultaneously makes the paper unfocussed and the purpose of each individual experiment does not always come across clearly and does not always align with the general storyline of the paper. For the reader, the experiments can thus feel quite disconnected from the extensive methods section. The image analysis, performance quantification and statistical analysis should also be extensively improved.

Response: We appreciate the comment. To improve the manuscript flow, we have moved some experiment data to the supplementary materials and enhanced the image analysis and quantification as suggested.

Evaluation of the paper

While the authors present an extensive validation effort for the *Diffractive Acoustic Tomography* method, we believe that less is more in the case of this paper. The authors should consider reducing the sheer number of experiments and instead focus on improving the quality but primarily the significance of each, by focussing on more quantitative image analysis strategies. The authors should also carefully revise the claims they are making and the conclusions they are drawing from the experiments and only state claims that can be directly supported.

As such, the paper is not fit for publication at this stage and requires major revisions before being considered.

Major high-level comments

1. *Concerning the novelty of the work:* The authors have to acknowledge that the diffraction slit has already been presented in the literature and cite the relevant paper [Wang, Y et al. 2017]. The authors should state if and how this work differentiates from the previously reported state-of-the-art methods. The authors should clearly outline the novelty of their work: the extensive experimental validation of the 3D DPAT/DUST method.

Response: We thank the reviewer for the good suggestions. We have updated the manuscript to include all citations relating to the previously described works using single-slit diffraction for PA imaging. We have also emphasized the novelty of this work in comparison to the previous works. Namely:

1: The use of a diffractive slit to enhance US imaging (3D-DUST) has never been demonstrated, which represents a major novelty of this work.

2: The efficient focal line reconstruction (FFL) method improves the applicability of the 3D-DAT method for both preclinical and clinical settings.

3. The dynamic slit allows for a thorough experimental characterization of the 3D-DAT method for both PA and US and optimization of the system setup to achieve a balanced resolution and SNR enhancement.

4. The 3D-DAT method was also extensively validated and tested for in-vivo applications. This includes functional and molecular imaging, as well as high-speed imaging for dynamic processes.

5. The longitudinal evaluation of embryonic oxygenation throughout pregnancy, and the effect of PFAS exposure, provide new knowledge on this important question.

2. *Concerning the number of experiments:* The authors may want to consider reducing the number of experiments they present in the paper, but go much more in-depth with the ones they chose to include, focusing on quantitative image analysis (including image quality measures such as FWHM, CNR, sharpness). We believe that the authors could remove the following experiments from the main article: the frog images, the leaf phantom, the caged Gold nanoparticles, and the photoswitching experiment. In addition, we do not believe that the PFAS experiment makes for a good “validation” case. Instead, we propose that the authors change the change story of the paper for this experiment: “After extensive testing of the method, we now demonstrate its capabilities on a feasibility case study”, focussing on the interesting insights generated.

Response: We thank the reviewer for the insightful suggestion. We have added more image analysis, including CNR, in-vivo FWHM calculations, and analysis of the spatial frequency spectrum of the images. We have also included **Supplementary Note 1** and **Supplementary Figure 2**, which provide in-depth explanation to the improved resolution using 3D-DAT. Additionally, to better address the effect on SNR, we have added a new simulation study in **Supplementary Figure 8** and **Figure 2**, showing how signals are recovered in beamforming due to increased off-axis detection angle, even though we have decreased on-axis sensitivity. We have additionally removed several experiments from the main text, including the hypoxia challenge, the star phantom, and the tumor hypoxia study, to improve the manuscript flow. As suggested, we have also improved the writing of the PFAS experiment and highlighted its nature as a feasibility study.

3. *Concerning the trade-off between SNR and resolution:* We do not believe that the trade-off between resolution and SNR is well explained in the paper. Based on the k-wave simulations, the majority of the signal amplitude is lost when passing the diffraction slit, which is expected to have consequences on the signal amplitude and thus the SNR. SNR can of course be partially recovered by the increased resolution and averaging, etc., but this mechanism has not been made clear. Furthermore, the authors report relative signal amplitudes most of the time, which hides the decrease in absolute signal and there should be at least a few figures highlighting this effect (for instance in Figure 3d,g). This trade-off between decreased SNR and improved resolution should further be quantified as percentage changes relative to the “traditional PAI” method. Terminology used in the paper such as “Superior”, “improved”, etc. have to be substantiated by the authors by providing necessary details or best, by providing quantitative values for the changes,

along with paired grouped comparisons using statistical tests in technical or biological replicates.

Response: Per the reviewer's suggestion, we have added **Supplementary Note 1** and **Supplementary Figure 2** to explain the impact of slit width on resolution. Additionally, as mentioned in the previous comment, we have added a new simulation study in **Supplementary Figure 8** and **Figure 2.** This study shows that while there is a decrease in on-axis peak pressure of 3D-DAT compared with the traditional method, the off-axis signal amplitude increases. Thus, through 3D beamforming along the elevational axis, we are able to achieve similar SNR in 3D-DAT compared with traditional imaging. This analysis is substantiated by simulation, phantom, and in-vivo studies of both SNR and CNR comparisons between 3D-DAT and the traditional method.

Due to the 3D nature of the diffractive signal detection and beamforming, more positions are integrated in 3D-DAT compared to traditional imaging, and thus the raw reconstructed image in 3D-DAT has a much stronger voxel value than that in traditional imaging. For this reason, we often show the normalized signal comparisons between the two methods. It is never our intention to hide the signal difference, which would have defied the very purpose of this paper. In fact, as we have clarified above, 3D-DAT has comparable signal to noise ratio and contrast with the traditional imaging. Moreover, 3D-DAT has shown much less dependence on the target orientation than the traditional imaging, which has been demonstrated in detail in our manuscript (**Figure 2, Figure 3**).

4. *Concerning the novelty of the FFL reconstruction:* It does not become clear in the manuscript how the proposed FFL reconstruction algorithm is different from a conventional sparse-matrix-multiplication-based reconstruction, apart from the necessity to adjust the delay lengths because of the slit. Following up on this, the relevance of the characterisation of the FFL reconstruction speed in the context of this work is unclear. The main claim of the paper is that of increased elevational resolution, which is not changed by optimising the computational speed of the algorithm. The investigation of different synthetic aperture widths is interesting but could be further explained in the supplementary materials.

Response: We thank the reviewer for the suggestion. We refer the reviewer to **R2 Q4** for the elaboration on the novelty of FFL. Additionally, per the reviewer's suggestion to improve the focus of the paper, we have moved the discussion on the characterization of FFL to the supplementary materials.

5. *Concerning the use of the signal-to-noise ratio:* Across all sections of the paper, it did not become apparent how the authors defined and calculated the SNR on their images. Which regions of interest (ROI) were chosen as the signal and noise masks? Were the ROI the same for each of the methods (DPAT vs. TD PA, and DUST vs. TD US) knowing that the same structures were represented? Are they computed on 2D or 3D masks? If 2D: in the imaging plane or the y-z plane? Perhaps the authors should consider including these details along with supplementary metrics such as the CNR or gCNR for the

analyses, to differently include the noise floor, and provide the readership with a better understanding of the SNR-resolution trade-off made in their study.

Response: We thank the reviewer for pointing this out. We have included a new section on page 25, explaining how both SNR and newly added CNR calculations were performed. Additionally, all images with SNR and CNR calculations have the target and background regions depicted.

Due to the inherent linearity of the FFL reconstruction, we chose to use CNR rather than gCNR calculations. Additionally, we found that the trends in CNR were similar to SNR for phantom and in-vivo studies. This reveals that the change in the background signal from the traditional method to 3D-DAT is approximately proportional to the change in the target signal. Thus, for better consistency and to reduce redundancy, we have shown only SNR comparisons in the main figures (Fig. 2, 3) and have included corresponding CNR comparisons for all SNR plots in newly added supplementary figures (**Supplementary Figures 10 and 20**).

6. *Concerning the clinical translation of the method:* The authors make claims towards the promises of clinical translatability of the method but do not substantiate these with their experiments. The authors should either include a healthy volunteer clinical study or focus their storyline on the potential for preclinical discovery (that could eventually influence clinical work).

Response: We thank the reviewer for the good suggestion. Currently, we are unable to include human data in this study due to our delayed IRB approval. While we do believe that this method is clinically translatable with the linear transducer array, we have removed statements attesting to the clinical translatability of the method in the introduction and discussion.

7. *Concerning the accessibility of code and data for this study:* The authors should consider making the code and data that support this study openly available to the research community.

Response: We fully appreciate that open accessibility is essential for technology dissipation. We will share the data and code for non-commercial applications once the paper is finalized for publication.

Detailed comments

1. Title: The “in deep tissues” part of the title could be considered redundant. The diffraction slit does not change any depth-related performance of the device, and both ultrasound and photoacoustic imaging are inherently depth-resolved. The authors may want to consider changing the title, making it more indicative of the paper content (e.g. by highlighting the extensive testing and validation efforts).

Response: We thank the reviewer for the good suggestion. We have modified the title of the paper to remove the redundancy.

2. Line 60: “imaging depth”. To our understanding, the limiting factor of imaging depth in PAI is not primarily the acoustic aperture but rather the light fluence decay. This is a confusing addition to the list and should be omitted.

Response: We thank the reviewer for pointing this out. The statement was made to describe the effect of detector choice on both PA and US imaging, however, due to the potential confusion this may cause, we have modified the statement by removing “imaging depth”.

3. Lines 62-66: We do not feel that the hemispherical array-based methods have found limited clinical translation. There have been successful initial studies both from the PAMMOTH project, as well as pioneered by Tomowave. The authors mention cost as a major point, which might not do the work justice, especially in the area of breast imaging. Perhaps the authors can elaborate on this a bit more? What are the advantages of these devices?

Response: We thank the reviewer for the good point. We have modified the discussion on clinical translatability of different arrays to reflect the reviewer’s comment. While there are exciting clinical translations with ring- and hemispherical-array based PA systems, linear-array based PA systems have the advantages of lower cost, off-the-shelf availability, and have found broader clinical applications, particularly for handheld use.

4. Lines 74-76: The deep learning sentence stands on its own and does not yield any benefit to the paper the way it is phrased. The authors should put it into context by specifying how and why these methods claimed to alleviate linear array imaging problems. We agree that “hallucinations” can be a problem as we cannot invent what isn't there, but a citation to one relevant paper that shows this would benefit the readership).

Response: We thank the reviewer for pointing this out. We have reworded the statement to be more specific to the challenges of 3D PA/US imaging, as ‘However, the lack of high-quality training data makes these approaches vulnerable to the ‘hallucination effect’, particularly for *in vivo* applications where the ground truth is not always available for high-fidelity training’. Additionally, we have included three new citations that are more specific to the ‘hallucination’ effect in deep-learning PA that aim to improve linear-array imaging.

5. 78-80: The authors claim that they are measuring in isotropic resolution, but later in the methods they reveal that the resolution is only approximately isotropic. Please correct the sentence and perhaps state the x, y, and z resolutions here already.

Response: We thank the reviewer for the good point. We have adjusted the manuscript to reflect that the resolution becomes approximately isotropic in 3D. Depending on the slit width, the lateral/elevational (x/y) resolutions can be isotropic (see resolution plots in **Figure 1**).The axial (z) resolution is independent of the slit width. To further clarify the

resolutions of 3D-DAT with the 0.8 mm slit width, we have included **Supplementary Figure 1**.

6. 82-84: See high-level comment on FFL. The current phrasing makes it sound like the authors have invented a new method for image reconstruction but it does not appear so from the methods. Previous work on conventional sparse-matrix-multiplication-based reconstruction should be cited.

Response: We have revised the phrasing to better reflect the novelty of the FFL reconstruction. We have also included a citation on previous work in sparse-matrix-multiplication-based reconstructions [1].

[1] Perrot, Vincent, et al. "So you think you can DAS? A viewpoint on delay-and-sum beamforming." Ultrasonics 111 (2021): 106309.

7. 86: 3D-DPAT has achieved “optimal balance” -> this is not substantiated by the results. The 0.8mm slit width has achieved the optimal balance for the proposed specific hardware design. The way it is written here, it sounds like this is the optimal hardware for PAI.

Response: We thank the reviewer for the good comment. We have adjusted the wording to better reflect the optimal balance in the context of 3D-DAT with 0.8 mm slit width. Our work does provide a useful pipeline for testing the system’s performance with different slit settings. The system setup for different detection geometries and ultrasound frequencies should be tested accordingly by other groups.

8. 117: The authors should mention here that their device can do multispectral imaging and perhaps briefly state the wavelengths and lasers set-up, even though they will be detailed again later in the methods section.

Response: We thank the reviewer for the good suggestion. We have included a comment at the specified point in the manuscript describing the multispectral imaging of 3D-DAT. Additionally, the multispectral imaging capability is described in **Figure 1**.

9. 121-124: It should be clarified how the beam broadening increases the resolution.

Response: We thank the reviewer for the good suggestion. We have added a supplementary note (**Supplementary Note 1**) on single-slit diffraction. Additionally, we have added a new supplementary figure (**Supplementary Figure 2**) with schematics to explain the corresponding concepts described in **Supplementary Note 1**.

10. 127ff: See high-level comment on FFL. Please make sure to highlight how the FFL algorithm differentiates from a conventional sparse-array reconstruction scheme.

Response: We thank the reviewer for the good suggestion. FFL is an efficient implementation of the 3D focal line reconstruction [1]. The reconstruction is formulated as a sparse beamforming matrix that is directly multiplied with the RF data to generate the beamformed image, additionally accelerated by GPU processing. This sparse

beamforming matrix in 3D-DAT is different from previous works [2-3], in which we generate a synthetic matrix array by scanning the single-slit with the linear array transducer. To improve parallelization, the nonzero entries of the columns represent each of the elements of the transducer array (n_e) at each of the positions that the transducer was scanned (n_{sa}). Additionally, our investigation of the synthetic aperture width (**Supplementary Figure 6**) further improved the efficiency of the reconstruction, because it revealed that only a synthetic aperture width of a few millimeters was necessary to achieve the near-optimal SNR and resolution enhancement. These factors allow FFL to achieve nearly two orders of magnitude reduction in computation time compared to the previously published 3D focal line reconstruction. Per the reviewer's suggestion, we have expanded on the novelty of the FFL reconstruction in the revised manuscript (pages 4,6), as well as moved the characterization of the reconstruction to supplementary materials.

[1] Xia, Jun, et al. "Three-dimensional photoacoustic tomography based on the focal-line concept." *Journal of biomedical optics* 16.9 (2011): 090505-090505.

[2] Perrot, Vincent, et al. "So you think you can DAS? A viewpoint on delay-and-sum beamforming." Ultrasonics 111 (2021): 106309.

[3] Hou, Gary Y., et al. "Sparse matrix beamforming and image reconstruction for 2-D HIFU monitoring using harmonic motion imaging for focused ultrasound (HMIFU) with in vitro validation." IEEE transactions on medical imaging 33.11 (2014): 2107-2117.

11. 150: It is unclear why the authors state here that the advantage mainly comes from the slit? Isn't it the only hardware component that is different?

Response: We thank the reviewer for pointing this out. We have revised this statement by removing the word "mainly".

12. 150ff: At which wavelength(s) is the characterisation done?

Response: The characterizations using the cross-hair target, the star target, the leaf phantom target, and the graphite depth targets were all performed using 1064 nm excitation for PA. We have included this information in the manuscript in the "validation of 3D-DAT by phantom imaging" section of the methods.

13. 182-190: The authors do not state the relative SNR decrease vs. resolution gain compared to not using the diffraction slit at all. This is important to decide how useful this will be for clinical translation. Loss of SNR can be detrimental as it significantly lowers the possible imaging depth and discernibility of image-derived biomarkers. Data points of SNR obtained in the absence of the designed slit should be added on all graphs quantifying SNR.

Response: Figure 2 shows the relationship between SNR and slit width. With an L7-4 probe, the elevational beam size at the focus (where the slit is) is approximately 1.1 mm. We investigate the SNR and resolution with slit widths up to ~2 mm, at which point the slit has little effect on the PA signals (i.e., the acoustic focus regime described in results

section). All SNR and CNR plots in the manuscript include the corresponding traditional imaging results as well.

14. 194: We suggest that there should be quantification of the resolution on a line plot through the star pattern. Ideally, this should be done as a function from the outside of the star shape to the centre, quantifying the modulation transfer function (MTF) of their system. The leaf phantom is believed to be made redundant here with the inclusion of this data. We suggest the authors adjust Figure 2 accordingly.

Response: We thank the reviewer for the good point. We have removed the redundancy, according to the reviewer's suggestion. We have moved the star phantom to supplementary material and kept the leaf phantom in the main figure. This is because the star phantom, while a good phantom for characterizing PA imaging, cannot be imaged with US. Instead, the leaf phantom provides a better omnidirectional target for both US and PA. We have additionally added frequency spectrum of the leaf phantom to better demonstrate the resolution and contrast difference between traditional imaging and 3D-DAT.

15. 216-241: We suggest to remove the entire paragraph on the computation time of the algorithm as this is not relevant to the main messages of the paper. It might be interesting in the supplementary materials.

Response: We thank the reviewer for the good suggestion. We have moved the paragraph on computation time of the algorithm to **Supplementary Note 2**. Additionally, we have moved the graph of computation time shown in **Figure 1** of the original submission to **Supplementary Figure 6**.

16. 242: We do not agree that all of the experiments are necessarily validation experiments, especially those, where the authors only show qualitative images. The authors should consider replacing the word "validation" with "testing".

Response: We have included several new quantifications for the in-vivo data of **Figure 2**, including FWHM, SNR, and CNR in several regions, and images of the frequency spectrum of the frog data in **Supplementary Figure 17**.

17. 262-263: Based on the phantom figure, it seems that well over 50% of the signal amplitude is lost using the diffraction slit. How does it maintain the same depth imaging capabilities? Isn't the system therefore approaching the noise floor much sooner?

Response: We thank the reviewer for the good question. As we have detailed above, while the received on-axis peak pressure decreases for 3D-DAT compared to the traditional method at each scanning position, the off axis peak pressure increases (due to the broader detection angle). The integration of signals over the elevational axis in 3D FFL reconstruction provides comparable SNR, CNR, and imaging depth between traditional and 3D-DAT methods. The newly added k-Wave simulation addresses this question, shown in the revised **Figure 2** and **Supplementary Figure 8**.

18. 264: The authors should quantify their claim with the inclusion of FWHM measurements.

Response: We thank the reviewer for the good suggestion. We have included the measurement of FWHM in the profile plots in **Figure 3** for the frog images, and in **Supplementary Figure 19** for the mouse images. Additionally, we have included **Supplementary Figure 17** with the spatial frequency spectrum of the frog images, clearly showing 3D-DAT achieves improved spatial frequency isotropy.

19. 269-270: Why did the authors use 1064nm as the O2 reference marker? We believe using the multispectral images to linearly unmix blood oxygenation to will better support the reduction in sO2 claim.

Response: We agree with the reviewer. However, in this case we used only the 1064 nm laser for higher imaging speed, which has a PRF of 100 Hz (10 times higher than the OPO laser). At 1064 nm, the oxy-hemoglobin has 10-fold stronger optical absorption than the deoxy-hemoglobin. To achieve fast imaging of hemodynamics, we relied on only the 1064 nm laser and showed the relative change in the oxy-hemoglobin under hypoxia challenge. We have included this reasoning in the text of the revised manuscript on page 11.

20. 446-448: Did the authors characterise the absorption coefficient of the glass and UV glue in the vis and NIR? How does the optical refraction change the light paths in their setup?

Response: In the visible and NIR-I regime, borosilicate glass and UV glue have little optical attenuation (https://www.glsciences.com/product/cells/about_a_cell/00003.html, https://www.masterbond.com/articles/optical-transmission-properties-adhesives). Since light delivered by the fiber bundle is quickly diverging, the refraction by the glass also has negligible effect on the illumination pattern. We have included a brief discussion in the revised manuscript.

21. 461-462: This is not isotropic. Between y and z resolution, there is nearly a 2-fold difference.

Response: We thank the reviewer for the good comment. The 2-fold voxel size difference used in the reconstruction grid does not represent the spatial resolution of the system. These values represent the effective spatial sampling we use when reconstructing the volume. However, the reviewer is correct that the axial resolution of the system is ~20% better than the lateral and elevational resolution. Thus, we have updated the manuscript to reflect that the resolution may approach isotropic in the x/y plane, while axial resolution is independent of the slit. Additionally, we have included **Supplementary Figure 1,** showing similar axial resolution (~400 µm) for PA and US using the crosshair target.

22. 468: Please add a reference to the original k-wave publication [Treeby et al, 2010].

Response: We thank the reviewer for pointing this out. We have included the original k-wave publication reference.

23. 490-491: How was the SNR computed? What were the ROIs used?

Response: SNR was computed as the average signal of a target divided by the standard deviation of a background region. We have included a section describing how both SNR and the newly added CNR were calculated. Additionally, we have added the ROIs used for calculating SNR and CNR to all relevant images.

24. 519ff: How many animals were enrolled in each study? How many controls? Where are the details about the frog images?

Response: We thank the reviewer for pointing this out. We modified the manuscript to include this information in the referenced section on page 21.

25. 521-522: Are ID numbers for the ethics approvals available? If so, they should be disclosed for transparency.

Response: We thank the reviewer for pointing this out. We have included the protocol numbers for both the mouse imaging and frog imaging.

26. 532-535: Did the authors take multiple frames at each position or only a single frame? Were the frames corrected for the pulse laser energy?

Response: Only a single frame was acquired at each position. The frames were corrected for laser pulse energy, as indicated in lines 443-445 of the original submission. More information has been added to page 18.

27. 565-566: Is there a citation that details the fluorescence imaging setup? Otherwise, details should be added here.

Response: We thank the reviewer for pointing this out. We have included more information on fluorescence imaging in the referenced section on page 23.

28. 618: Details on the used scanning magnification should be provided.

Response: We have included the magnification in the referenced section. (20x for large image, 63x for close-up image).

29. 619ff: More elaboration on the image quantification strategies used should be included, along with metrics in a separate section, elaborating on SNR and further measures included in response to the reviews.

Response: We thank the reviewer for the good suggestion. We have included a new section in the methods that details how SNR and CNR were computed on page 25.

30. 622: How were the vessels segmented?

Response: Vessel segmentations were done manually, using the two main landmark skin vessels in the mice (see image below). For the pregnancy study, embryos were segmented manually, and the segmentations were validated from the corresponding high-frequency US data. We have updated the manuscript to include this information in the referenced section on page 25.

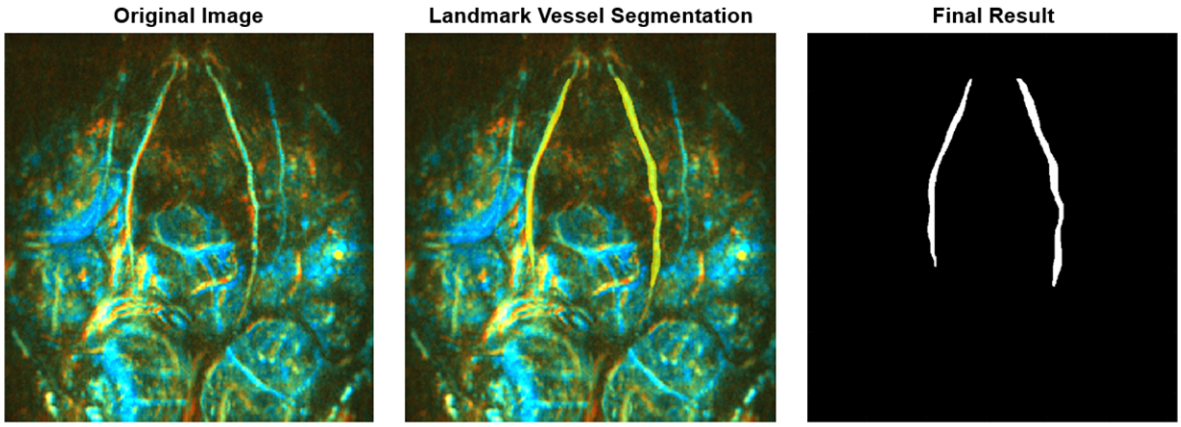

Figure R3. Segmentation of vessels*. A control pregnant mouse image is shown on the left, the segmented vessels overlaid on the image in the middle, and the final binary segmentation of the vessels on the right.*

Figure comments

31. In Figure 2c: how is the SNR computed? Where are the masks, are they the same for traditional vs non-traditional configurations? How does the SNR change when computing over the imaging plane or even better, in 3D?

Response: We thank the reviewer for the good questions. The SNR is computed as the average signal of the target region divided by the standard deviation of the background region. The masks are identical for corresponding point targets, manually chosen as a box of 6x6 pixels centered on the target in the 3D-DAT images, which is also centered on the point targets in the traditional images. **Figure 2e** shows a comparison of corresponding point targets at different depths for both PA and US. We chose to display the close-up data in the main figure for a better side-by-side comparison. As such, the region chosen as background is outside of the FOV of the displayed images in the main figure. We have created a new **Supplementary Figure 12** showing the entire depth phantom images, with the corresponding target and background regions.

32. Fig 3k: why not use spectral unmixing here, like in the coil phantom?

Response: As stated in response to comment 19, only the 1064 nm laser was used for its high PRF of 100 Hz, while the other two lasers we have are only 10 Hz PRF.

33. Figure 3 d-h: Unnormalised signal amplitudes should be provided along with quantification metrics such as FWHM.

Response: We thank the reviewer for the good suggestion. We have included quantifications of FWHM for the profile plots, as well as SNR and corresponding CNR of select regions in both the mouse and frog images. However, due to the nature of the reconstructions used, the un-normalized scale of each reconstruction is very different (FFL sums over many more datapoints than the traditional reconstruction), and thus we believe a plot of the unnormalized reconstructed signals would not provide useful information. By contrast, the new quantifications of SNR and CNR in the image should give the readers an understanding of the comparison of relative signal strengths between the two datasets. Additionally, the newly added k-Wave simulations address the reduction in in-focus signal strength in 3D-DAT, and how the SNR is recovered by the improved off-axis detection sensitivity (included in **Figure 1** and **Supplementary Figure 3 and 8**).

34. Fig 4c: How do the large error bars come to be with only N=3 points? Is this the right visualisation for only N=3 points? Please plot the points individually.

Response: We thank the reviewer for the good suggestion. We have reformatted the plot to be a bar plot with the individual data points overlaid, and error bars representing SD. Additionally, for the reader's clear understanding of the data, all images at all timepoints of the cGNS-injected mice are shown in **Supplementary Figure 23**.

35. Fig 4f: Boxplots cannot be used with N=3 values. For them to statistically make sense as a visualisation strategy, at least 5 data points are needed per boxplot. The visualisation should be changed if average values are plotted, except if every single pixel value is represented, not the average. What was the p-value computed on?

Response: We thank the reviewer for pointing this out. We have changed the plot to be a bar plot with the individual data points. The p-value was computed on the 3 average values between the vessel and tumor groups. Additionally, this experiment has been moved to supplementary materials to reduce the number of experiments, as suggested by the reviewer.

Three-Dimensional Diffractive Acoustic Tomography

Luca Menozzi [a*], Tri Vu [a*], Aidan J Canning [a*], Harshal Rawtani [b], Carlos Taboada [a,c], Marie Elise Abi Antoun [d], Chenshuo Ma [a], Jesse Delia [e], Van Tu Nguyen [a], Soon-Woo Cho [a], Jianing Chen [a], Theresa Charity [f,g], Yirui Xu [a], Phuong Tran [a,h], Jun Xia [i], Gregory M. Palmer[f,g], Tuan Vo-Dinh [a,j], Liping Feng [b], Junjie Yao [a,k]

[a] *Department of Biomedical Engineering, Duke University, Durham, North Carolina, USA*

[b] *Duke University School of Medicine, Durham, North Carolina, USA*

[c] *Department of Biological Sciences, Vanderbilt University, Nashville, Tennessee, USA*

[d] *Tufts Medical Center, Boston, Massachusetts, USA*

[e] *American Museum of Natural History, New York City, New York, USA*

[f] *Department of Radiation Oncology, Duke University School of Medicine, Durham, NC 27710, USA*

[g] *Department of Surgery, Duke University School of Medicine, Durham, NC 27710, USA*

[h] *Department of Biomedical Engineering, North Carolina State University, Raleigh, North Carolina, USA*

[i] *Department of Biomedical Engineering, University at Buffalo, Buffalo, New York, USA*

[j] *Department of Chemistry, Duke University, Durham, NC 27708, USA*

[k] *Department of Neurology, Duke University of School of Medicine, Durham, NC 27710, USA*

The authors sincerely appreciate the time and efforts of the editor and all reviewers for providing helpful comments and suggestions that have improved the manuscript.

Response to Reviewer 1:

The authors well address all of my previous comments. Further, the revised manuscript shows compelling results which have never done before with the conventional PA and US methods. I strongly recommend this manuscript for publication.

I have one suggestion: the authors show a series structural, functional, and molecular imaging of 3D DAT in various animals. Can you perform any human imaging like in human forearm or palm? Then, the potential of clinical translation would be highly feasible.

Response: We sincerely thank the reviewer for their suggestions and support. Regarding the human imaging, while we agree with the point that human data would show the clinical translatability of the technique, we are still in the process of having our lab's institutional review board (IRB) for human imaging approved, and as such are unable to include human data in this study.

Response to Reviewer 2:

All comments are addressed. The manuscript can be accepted for publication.

Response: We sincerely thank the reviewer for their suggestions and support.

Response to Reviewer 3:

The authors have addressed all reviewers' comments well and the paper can be accepted.

Response: We sincerely thank the reviewer for their suggestions and support.

Response to Reviewer 4:

I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

Response: We sincerely thank the reviewer for their suggestions and support.

Review of NCOMMS-24-34632-T: Three-Dimensional Diffractive Acoustic Tomography in Deep Tissues

Summary of the paper

The authors present the use of a diffraction slit, created by two opposing glass slides that move along the centre of a 2D US detection array. A photoacoustically (PA) or ultrasonically (US) induced sound wave is re-focussed through the slit, travelling directly towards the detection elements afterwards. In the US configuration, the sound waves pass the diffraction slit twice, leading to an initial broadening of the sound waves. It is an effective idea which successfully increases the lateral resolution of acquired PA and US images.

The authors present a whole suite of experiments to demonstrate the capabilities of their device design, especially focusing on the improved elevational resolution of their system. For this, they include phantom experiments, but also pre-clinical small animal experiments that cover a range of purposes: characterising the resolution of the device and showing that it can characterise time-resolved phenomena as well as functional parameters; potentially with higher quantitative accuracy due to the higher resolution and therefore less pronounced partial volume/blurring effects.

Summary of the main strengths and weaknesses of the paper

The main strength of the paper is the comprehensive suite of testing experiments. Not only did they reproduce experiments from the literature, but also included a completely novel study on the investigation of PFAS and their influence on embryonal development.
The manuscript is well written and the introduction strikes a great balance of catering to both a general and specialised audience by introducing the main concepts in an appropriate level of detail. The authors also (generally) go into great detail to describe their methods, leading to a high chance that many aspects of this work will be reproducible in the future.

The main drawback of the work is, that the claims and conclusions are at times not clearly supported. Examples:

- The idea of using a diffraction slit is not novel but has already been proposed by [Wang, Y et al. 2017], though the authors claim that they *“develop a novel technique for 3D PA/US imaging using an off-the-shelf linear-array transducer—3D diffractive acoustic tomography (3D-DAT)”*.
- The authors compare the computation time of the sparse array reconstruction method with a conventional delay-and-sum approach, which is inherently not as optimised.
- Traditional method images look confounded by artefacts and of poor quality, not just in the lateral resolution, but also in the imaging plane, where many discontinuities can be visualised (*e.g.* in Figures 3 b and f) which were not seen in past work published by the authors.
- The authors use imprecise language in the text, *e.g.* “optimal”, “greatly”, “superior”, etc., without providing quantifications and details how and why this was achieved.
- In the introduction and conclusion, the authors claim the potential for clinical translation of the method, but none of the experiments directly support this claim.

While the impressive number of experiments makes the testing of the methodology quite comprehensive, it simultaneously makes the paper unfocussed and the purpose of each individual experiment does not always come across clearly and does not always align with the general storyline of the paper. For the reader, the experiments can thus feel quite disconnected from the extensive methods section. The image analysis, performance quantification and statistical analysis should also be extensively improved.

Evaluation of the paper

While the authors present an extensive validation effort for the *Diffractive Acoustic Tomography* method, we believe that less is more in the case of this paper. The authors should consider reducing the sheer number of experiments and instead focus on improving the quality but primarily the significance of each, by focussing on more quantitative image analysis strategies. The authors should also carefully revise the claims they are making and the conclusions they are drawing from the experiments and only state claims that can be directly supported.

As such, the paper is not fit for publication at this stage and requires major revisions before being considered.

Major high-level comments

1. *Concerning the novelty of the work:* The authors have to acknowledge that the diffraction slit has already been presented in the literature and cite the relevant paper [Wang, Y et al. 2017]. The authors should state if and how this work differentiates from the previously reported state-of-the-art methods. The authors should clearly outline the novelty of their work: the extensive experimental validation of the 3D DPAT/DUST method.
2. *Concerning the number of experiments:* The authors may want to consider reducing the number of experiments they present in the paper, but go much more in-depth with the ones they chose to include, focusing on quantitative image analysis (including image quality measures such as FWHM, CNR, sharpness). We believe that the authors could remove the following experiments from the main article: the frog images, the leaf phantom, the caged Gold nanoparticles, and the photoswitching experiment. In addition, we do not believe that the PFAS experiment makes for a good “validation” case. Instead, we propose that the authors change the change story of the paper for this experiment: “After extensive testing of the method, we now demonstrate its capabilities on a feasibility case study”, focussing on the interesting insights generated.
3. *Concerning the trade-off between SNR and resolution:* We do not believe that the trade-off between resolution and SNR is well explained in the paper. Based on the k-wave simulations, the majority of the signal amplitude is lost when passing the diffraction slit, which is expected to have consequences on the signal amplitude and thus the SNR. SNR can of course be partially recovered by the increased resolution and averaging, etc., but this mechanism has not been made clear. Furthermore, the authors report relative signal amplitudes most of the time, which hides the decrease in absolute signal and there should be at least a few figures highlighting this effect

(for instance in Figure 3d,g). This trade-off between decreased SNR and improved resolution should further be quantified as percentage changes relative to the “traditional PAI” method. Terminology used in the paper such as “Superior”, “improved”, etc. have to be substantiated by the authors by providing necessary details or best, by providing quantitative values for the changes, along with paired grouped comparisons using statistical tests in technical or biological replicates.

4. *Concerning the novelty of the FFL reconstruction:* It does not become clear in the manuscript how the proposed FFL reconstruction algorithm is different from a conventional sparse-matrix-multiplication-based reconstruction, apart from the necessity to adjust the delay lengths because of the slit. Following up on this, the relevance of the characterisation of the FFL reconstruction speed in the context of this work is unclear. The main claim of the paper is that of increased elevational resolution, which is not changed by optimising the computational speed of the algorithm. The investigation of different synthetic aperture widths is interesting but could be further explained in the supplementary materials.
5. *Concerning the use of the signal-to-noise ratio:* Across all sections of the paper, it did not become apparent how the authors defined and calculated the SNR on their images. Which regions of interest (ROI) were chosen as the signal and noise masks? Were the ROI the same for each of the methods (DPAT vs. TD PA, and DUST vs. TD US) knowing that the same structures were represented? Are they computed on 2D or 3D masks? If 2D: in the imaging plane or the y-z plane? Perhaps the authors should consider including these details along with supplementary metrics such as the CNR or gCNR for the analyses, to differently include the noise floor, and provide the readership with a better understanding of the SNR-resolution trade-off made in their study.
6. *Concerning the clinical translation of the method:* The authors make claims towards the promises of clinical translatability of the method but do not substantiate these with their experiments. The authors should either include a healthy volunteer clinical study or focus their storyline on the potential for preclinical discovery (that could eventually influence clinical work).
7. *Concerning the accessibility of code and data for this study:* The authors should consider making the code and data that support this study openly available to the research community.

Detailed comments

1. Title: The “in deep tissues” part of the title could be considered redundant. The diffraction slit does not change any depth-related performance of the device, and both ultrasound and photoacoustic imaging are inherently depth-resolved. The authors may want to consider changing the title, making it more indicative of the paper content (e.g. by highlighting the extensive testing and validation efforts).
2. Line 60: “imaging depth”. To our understanding, the limiting factor of imaging depth in PAI is not primarily the acoustic aperture but rather the light fluence decay. This is a confusing addition to the list and should be omitted.
3. Lines 62-66: We do not feel that the hemispherical array-based methods have found limited clinical translation. There have been successful initial studies both from the PAMMOTH project, as well as pioneered by Tomowave. The authors mention cost as a major point, which might not do the work justice, especially in the area of breast

imaging. Perhaps the authors can elaborate on this a bit more? What are the advantages of these devices?

4. Lines 74-76: The deep learning sentence stands on its own and does not yield any benefit to the paper the way it is phrased. The authors should put it into context by specifying how and why these methods claimed to alleviate linear array imaging problems. We agree that “hallucinations” can be a problem as we cannot invent what isn't there, but a citation to one relevant paper that shows this would benefit the readership).
5. 78-80: The authors claim that they are measuring in isotropic resolution, but later in the methods they reveal that the resolution is only approximately isotropic. Please correct the sentence and perhaps state the x, y, and z resolutions here already.
6. 82-84: See high-level comment on FFL. The current phrasing makes it sound like the authors have invented a new method for image reconstruction but it does not appear so from the methods. Previous work on conventional sparse-matrix-multiplication-based reconstruction should be cited.
7. 86: 3D-DPAT has achieved “optimal balance” -> this is not substantiated by the results. The 0.8mm slit width has achieved the optimal balance for the proposedspecific hardware design. The way it is written here, it sounds like this is the optimal hardware for PAI.
8. 117: The authors should mention here that their device can do multispectral imaging and perhaps briefly state the wavelengths and lasers set-up, even though they will be detailed again later in the methods section.
9. 121-124: It should be clarified how the beam broadening increases the resolution.
10. 127ff: See high-level comment on FFL. Please make sure to highlight how the FFL algorithm differentiates from a conventional sparse-array reconstruction scheme.
11. 150: It is unclear why the authors state here that the advantage mainly comes from the slit? Isn’t it the only hardware component that is different?
12. 150ff: At which wavelength(s) is the characterisation done?
13. 182-190: The authors do not state the relative SNR decrease vs. resolution gain compared to not using the diffraction slit at all. This is important to decide how useful this will be for clinical translation. Loss of SNR can be detrimental as it significantly lowers the possible imaging depth and discernibility of image-derived biomarkers. Data points of SNR obtained in the absence of the designed slit should be added on all graphs quantifying SNR.
14. 194: We suggest that there should be quantification of the resolution on a line plot through the star pattern. Ideally, this should be done as a function from the outside of the star shape to the centre, quantifying the modulation transfer function (MTF) of their system. The leaf phantom is believed to be made redundant here with the inclusion of this data. We suggest the authors adjust Figure 2 accordingly.
15. 216-241: We suggest to remove the entire paragraph on the computation time of the algorithm as this is not relevant to the main messages of the paper. It might be interesting in the supplementary materials.
16. 242: We do not agree that all of the experiments are necessarily validation experiments, especially those, where the authors only show qualitative images. The authors should consider replacing the word “validation” with “testing”.
17. 262-263: Based on the phantom figure, it seems that well over 50% of the signal amplitude is lost using the diffraction slit. How does it maintain the same depth

imaging capabilities? Isn't the system therefore approaching the noise floor much sooner?

18. 264: The authors should quantify their claim with the inclusion of FWHM measurements.
19. 269-270: Why did the authors use 1064nm as the O2 reference marker? We believe using the multispectral images to linearly unmix blood oxygenation to will better support the reduction in sO2 claim.
20. 446-448: Did the authors characterise the absorption coefficient of the glass and UV glue in the vis and NIR? How does the optical refraction change the light paths in their setup?
21. 461-462: This is not isotropic. Between y and z resolution, there is nearly a 2-fold difference.
22. 468: Please add a reference to the original k-wave publication [Treeby et al, 2010].
23. 490-491: How was the SNR computed? What were the ROIs used?
24. 519ff: How many animals were enrolled in each study? How many controls? Where are the details about the frog images?
25. 521-522: Are ID numbers for the ethics approvals available? If so, they should be disclosed for transparency.
26. 532-535: Did the authors take multiple frames at each position or only a single frame? Were the frames corrected for the pulse laser energy?
27. 565-566: Is there a citation that details the fluorescence imaging setup? Otherwise, details should be added here.
28. 618: Details on the used scanning magnification should be provided.
29. 619ff: More elaboration on the image quantification strategies used should be included, along with metrics in a separate section, elaborating on SNR and further measures included in response to the reviews.
30. 622: How were the vessels segmented?

Figure comments

31. In Figure 2c: how is the SNR computed? Where are the masks, are they the same for traditional vs non-traditional configurations? How does the SNR change when computing over the imaging plane or even better, in 3D?
32. Fig 3k: why not use spectral unmixing here, like in the coil phantom?
33. Figure 3 d-h: Unnormalised signal amplitudes should be provided along with quantification metrics such as FWHM.
34. Fig 4c: How do the large error bars come to be with only N=3 points? Is this the right visualisation for only N=3 points? Please plot the points individually.
35. Fig 4f: Boxplots cannot be used with N=3 values. For them to statistically make sense as a visualisation strategy, at least 5 data points are needed per boxplot. The visualisation should be changed if average values are plotted, except if every single pixel value is represented, not the average. What was the p-value computed on?

Literature References

[Wang, Y et al. 2017] Wang, Yuehang, et al. "Second generation slit‐based photoacoustic tomography system for vascular imaging in human." Journal of biophotonics 10.6-7 (2017): 799-804.

[Treeby, B et al. 2010] Treeby, Bradley E., and Benjamin T. Cox. "k-Wave: MATLAB toolbox for the simulation and reconstruction of photoacoustic wave fields." Journal of biomedical optics 15.2 (2010): 021314-021314.