PA-SfM: Tracker-free differentiable acoustic radiation for multi-view 3D photoacoustic imaging

Shuang Li[1†], Jian Gao[2†], Chulhong Kim[3], Seongwook Choi[3], Hao Huang[4, 6], Xuanhao Wang[4, 5], Junhui Shi[4, 5, 6], Qian Chen[1], Yibing Wang[1], Shuang Wu[2], Yu Zhang[1], Tingting Huang[1], Yucheng Zhou[1], Boxin Yao[1], Yao Yao[2*], Changhui Li[1, 7*]

[1*]Department of Biomedical Engineering, College of Future Technology, Peking University, Beijing, China.
[2*]School of Intelligence Science and Technology, Nanjing University, Suzhou, China.
[3]Department of Electrical Engineering, Convergence IT Engineering, Mechanical Engineering, and Medical Science and Engineering, Medical Device Innovation Center, Pohang University of Science and Technology, Pohang, Republic of Korea.
[4]Zhejiang Lab, Hangzhou, China.
[5]State Key Laboratory of Ocean Sensing, Hangzhou, China.
[6]College of Biomedical Engineering and Instrument Science, Zhejiang University, Hangzhou, China.
[7*]National Biomedical Imaging Center, Peking University, Beijing, China.

*Corresponding author(s). E-mail(s): yaoyao@nju.edu.cn; chli@pku.edu.cn;
†These authors contributed equally to this work.

Abstract

Sensor arrays in three-dimensional photoacoustic imaging (3D PAI) are often sparse because of physical constraints. Increasing spatial sampling density and expanding the effective aperture therefore commonly require measurements from multiple array positions. Conventional systems achieve this using mechanical rotation stages or external tracking devices, but these approaches increase system complexity and ultimately limit image resolution through the precision bounds of physical pose feedback. Here we present PA-SfM, an algorithmic framework that recovers sensor array pose directly from photoacoustic data. By combining

a differentiable acoustic radiation model with hierarchical optimization under rigid geometry constraints, PA-SfM enables accurate joint reconstruction across multiple views without external positioning hardware. We validated the method in numerical simulations, *in vivo* rat imaging with known relative geometry, and a 3D PAI system operated with mechanical rotation. PA-SfM achieved sub-millimeter pose accuracy and substantially improved volumetric reconstruction fidelity. In the mechanically rotated system, PA-SfM outperformed reconstruction based on motor encoder feedback, yielding sharper and more continuous vascular structures. These results establish PA-SfM as a universal robust computational strategy for multi-view 3D photoacoustic imaging and show that geometric calibration recovered directly from acoustic signals can surpass the precision limits of conventional physical feedback. The source code is publicly available at https://github.com/JaegerCQ/PA-SfM.

Keywords: Multi-view 3D photoacoustic tomography, pose recovery, differentiable acoustic radiation, tracker-free imaging, structure from motion

1 Introduction

Photoacoustic imaging (PAI) is a hybrid imaging modality that combines optical absorption contrast with ultrasound detection, enabling noninvasive visualization of biological tissues at centimeter scale depths with spatial resolution beyond that of purely optical imaging techniques [1–6]. Owing to these advantages, PAI has been widely explored in both preclinical and clinical settings. Recent advances in three-dimensional (3D) PAI using two-dimensional matrix arrays have enabled *in vivo* imaging of peripheral vessels [7, 8], breast tissue [9, 10], and small animals [11]. A variety of array geometries, including spherical and planar arrays [12–19], as well as synthetically scanned systems [8, 20], have been developed to improve volumetric imaging performance.

Despite these advances, sensor arrays in 3D photoacoustic imaging are often sparse because of physical and engineering constraints. As a result, a single acquisition pose may provide insufficient spatial sampling density or angular coverage for high quality volumetric reconstruction. Acquiring data from multiple views is therefore a common strategy to increase the effective aperture, expand the field of view, and mitigate limited view artifacts. In practice, these additional views are often generated through mechanical rotation, translational scanning, or freehand repositioning of the probe [1, 8, 21, 22]. However, the success of multi-view reconstruction critically depends on accurate knowledge of the relative sensor pose across views. Even small geometric errors can lead to structural blurring, vessel discontinuity, and loss of quantitative fidelity in the reconstructed volume.

Current solutions typically rely on physical pose feedback. In controlled systems, motorized stages and mechanical encoders are used to define the acquisition trajectory. In freehand settings, external positioning devices such as optical tracking systems, electromagnetic trackers, and inertial measurement units are commonly employed [23, 24]. Although these methods can provide useful geometric information, they also increase

system complexity, cost, and calibration burden. Optical tracking requires a clear line of sight and is often difficult to maintain in crowded experimental or clinical environments [23, 25]. Electromagnetic tracking is susceptible to field distortion from metallic objects and nearby electronics [24]. Inertial sensing is portable and inexpensive, but its estimates are vulnerable to drift accumulation over time [23, 26]. Mechanical scanning systems offer controlled motion, yet the final reconstruction remains limited by encoder precision, assembly tolerance, and motion induced vibration.

These limitations have motivated increasing interest in tracker-free or sensorless pose estimation directly from the acquired data [27–29]. However, this problem is fundamentally challenging in photoacoustic imaging. Unlike ultrasound, photoacoustic signals do not exhibit stable speckle patterns that support standard speckle tracking or cross-correlation based motion estimation [30–33]. Deep learning approaches can learn motion priors from reconstructed images, but they require large labeled datasets and may not generalize across imaging systems or acquisition conditions [34]. Other image based strategies, including frequency domain analysis of vascular patterns [35], can estimate relative probe motion under restricted conditions, but they typically require substantial overlap between views and become unreliable when displacement is large, structural features are weak, or probe motion changes abruptly. More broadly, these methods operate on reconstructed images rather than on the acoustic measurement process itself.

Here we present PA-SfM, a tracker-free framework for pose recovery and joint reconstruction in multi-view 3D photoacoustic imaging. PA-SfM estimates relative array geometry directly from photoacoustic data by combining a differentiable acoustic wave model with hierarchical optimization under rigid geometry constraints. The framework first reconstructs a reference volume from a known view, then localizes sensors in the new view, enforces rigid-array consistency, and finally refines the global pose through differentiable optimization before joint multi-view reconstruction. We validated the method in numerical simulations, *in vivo* rat experiments with known relative geometry, and a 3D PAI system operated with mechanical rotation. Across these settings, PA-SfM recovered cross-view geometry with high accuracy, improved reconstruction fidelity, and, in the mechanically rotated system, outperformed reconstruction based on motor encoder feedback. These results establish PA-SfM as a robust computational strategy for multi-view 3D photoacoustic imaging without external trackers or precise mechanical pose feedback.

2 Results

We first demonstrated that PA-SfM could replace, and even outperform, hardware-based pose feedback in a real multi-view photoacoustic tomography system. We then performed controlled validation in numerical simulations and in *in-vivo* rat experiments with known relative geometry to quantify localization and reconstruction accuracy. Unless otherwise stated, the term PA-SfM coarse pose refers to the intermediate pose estimate obtained after rigid-array estimation and before the final global differentiable fine-tuning step.

2.1 PA-SfM surpasses hardware-tracked multi-view reconstruction using a mechanically rotating array

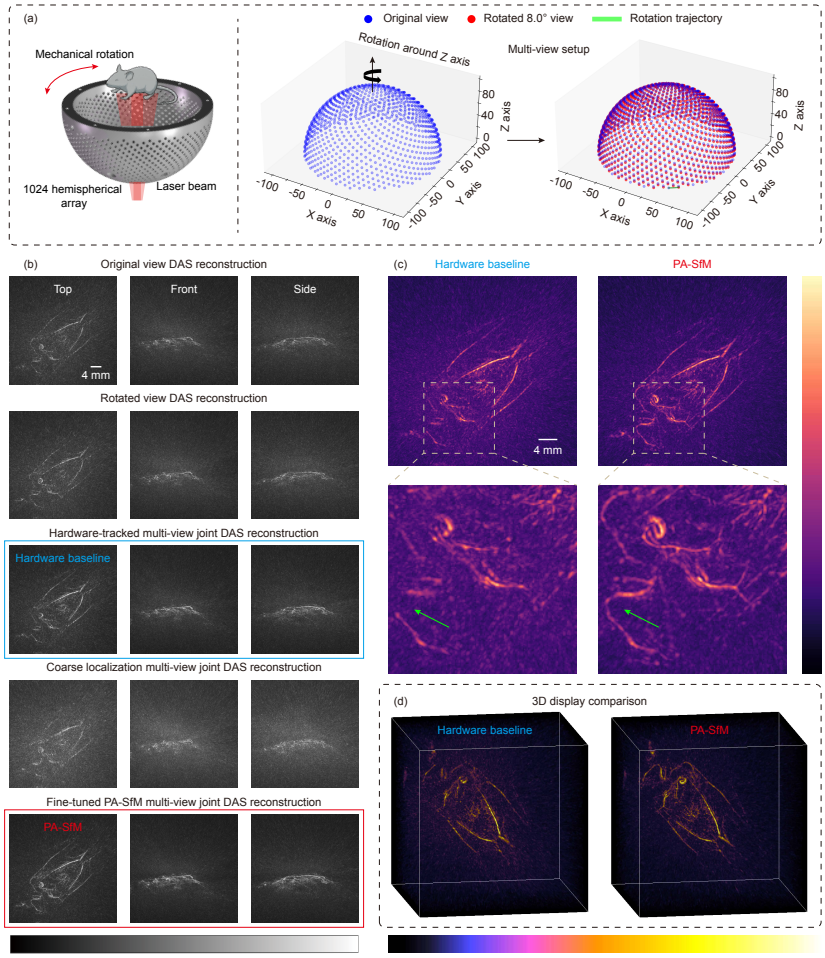

Fig. 1 Multi-view *in-vivo* reconstructions of mouse hepatic vasculature. (a) Multi-view 3D PAI imaging setup using a mechanically rotating hemispherical array. (b) Top-, front-, and side-view maximum amplitude projections (MAP) of reconstructions from single views (original and rotated) and multi-views (hardware-tracked baseline, PA-SfM coarse pose, and PA-SfM fine-tuned result) (Scale: 4 mm; grayscale; reconstruction with DAS). (c) Locally magnified top-view MAPs comparing the hardware-tracked baseline and PA-SfM fine-tuned reconstruction (Scale: 4 mm; colormap: magma). (d) 3D display comparing hardware-tracked and PA-SfM reconstructions.

To evaluate the practical utility and robustness of PA-SfM in a representative hardware system for sparse 3D photoacoustic tomography, we first tested the method using a mechanically rotating hemispherical transducer array (3D-PanoPACT [21]). The experimental data were acquired by Professor Junhui Shi's team at Zhejiang Lab.

More details regarding the 3D-PanoPACT system configuration and performance can be found in literature [21]. In these systems, rotational acquisition is commonly used to increase effective spatial sampling density and expand the field of view. However, this strategy also introduces a strong dependence on mechanical feedback, and the resulting multi-view reconstruction remains vulnerable to encoder inaccuracies and stochastic micro-vibrations during motor movement. These pose errors induced by hardware can degrade spatial resolution and introduce misalignment artifacts, thereby limiting the performance of conventional hardware-tracked reconstruction.

As illustrated in Fig. 1(a), the imaging target was the *in-vivo* mouse hepatic vasculature. We acquired two views, including an original view and a second view mechanically rotated by 8.0° around the Z-axis. To benchmark PA-SfM against encoder-based tracking while also isolating the contribution of the final optimization stage, we compared reconstructions from the two single views with multi-view joint reconstructions using the hardware-tracked baseline (strictly determined by motor encoder feedback), the intermediate PA-SfM coarse pose, and the final PA-SfM fine-tuned pose. The classical delay-and-sum (DAS) was used to perform reconstruction for both single view and multi-view.

As shown in Fig. 1(b), both the original and rotated single-view reconstructions provide only incomplete vascular information because of sparse spatial sampling. The hardware-tracked joint reconstruction partially alleviates this sparsity, but still exhibits noticeable blurring and vessel discontinuities, indicating that mechanical feedback alone is insufficient for precise multi-view alignment. The intermediate PA-SfM coarse-pose reconstruction recovers the relative pose at a gross level and restores the overall vascular outline in the fused image; however, residual pose errors still lead to substantial blurring and loss of fine structural detail. In contrast, the final PA-SfM fine-tuned reconstruction estimates the relative transformation directly from the photoacoustic data and yields markedly sharper and more continuous vascular structures. In the locally magnified top-view maximum amplitude projections (MAPs) in Fig. 1(c), the green arrows highlight misregistered and fragmented microvessels in the hardware-tracked reconstruction, whereas the final PA-SfM fine-tuned reconstruction resolves these local discontinuities and recovers the microvascular morphology with much higher fidelity. The 3D comparison in Fig. 1(d) further confirms that PA-SfM not only removes the need for explicit mechanical feedback, but also achieves higher-quality multi-view reconstruction than the hardware-tracked baseline.

Because this mechanically rotated experiment does not provide strict ground truth (GT) pose labels for numerical error analysis, we next turned to controlled simulation and *in-vivo* experiments with known relative geometry to quantify the accuracy of pose recovery and joint reconstruction.

2.2 Quantitative validation in multi-view sparse-array simulations

To quantitatively evaluate pose recovery and image reconstruction under fully controlled conditions, we next conducted numerical simulations using a phantom liver model under a multi-view setup. As shown in Fig. 2(c), the simulation utilized a multi-view configuration to acquire sparse hemispherical array data. Specifically, the array

consisted of a 33-element sparse planar array distributed across a spherical cap. View 1 served as the reference array pose, while View 2 and View 3 represented novel array poses generated by translating and rotating around the simulated source.

To demonstrate the pose optimization mechanism, Fig. 3 illustrates the PA signals of the target sensor at different optimization stages. The differentiable single-sensor localization effectively optimized the sensor location, as evidenced by a clear convergence in the distance error curve (Fig. 3(c)). After filtering out aberrant single-sensor localizations via rigid constraints, a minor positional deviation still persisted in the global pose. Consequently, the PA signal extracted from the intermediate PA-SfM coarse pose exhibited a noticeable discrepancy from the GT supervision signal (green dashed line versus solid gray line in Fig. 3(f)). At this intermediate coarse-pose stage, the maximum and minimum Euclidean distance errors for sensor localization were 0.002334 m and 0.000265 m, respectively. Through further global differentiable fine-tuning, the alignment of the PA signals progressively improved, and the fine-tuned pose signal matched the GT signal much more closely than either the initial or coarse-pose estimates (Fig. 3(b-f)). Ultimately, the maximum and minimum sensor localization errors were substantially reduced to 0.001784 m and 0.000135 m, respectively, demonstrating the high accuracy of the proposed fine-tuning mechanism.

Based on this precise pose recovery, we systematically compared the 3D reconstruction results. As depicted in Fig. 2(a-b), the single-view reconstruction suffers from severe limited-view artifacts, yielding quantitative metrics of 25.51 dB for the peak signal-to-noise ratio (PSNR) and 0.6775 for the structural similarity index measure (SSIM) in the MAP along the XY plane. While the multi-view joint reconstruction using the intermediate PA-SfM coarse pose effectively expands the field of view, it still exhibits structural blurring due to residual cross-view misalignment, resulting in a degraded PSNR of 21.35 dB and an SSIM of 0.4552. However, after applying the PA-SfM global differentiable pose fine-tuning module, the refined multi-view reconstruction demonstrates a substantial improvement. It successfully mitigates motion artifacts and restores high-resolution 3D structures that are highly comparable to the GT of the phantom liver, achieving a significantly improved PSNR of 38.90 dB and an SSIM of 0.9637.

2.3 In-vivo rat validation of multi-view PA-SfM joint reconstruction with known relative geometry

We then validated PA-SfM on real biological data while preserving a known relative geometry between views. Specifically, we performed *in-vivo* multi-view 3D photoacoustic imaging of rat kidney (Fig. 4) and rat liver (Fig. 5), which were acquired by Professor Kim's lab using a hemispherical ultrasound (US) transducer array with 1024 elements and a radius of 60 mm [36]. More details regarding the 3D PAI system and animal experiments can be found in the literature [37–43]. To emulate a multi-view acquisition scheme under limited-view conditions while retaining known relative geometry, we deliberately partitioned the full 1024-element spherical array into two interleaved 512-element sparse subsets. These two subsets, designated as View A and View B, represent two independent imaging poses separated by a 90° relative rotation

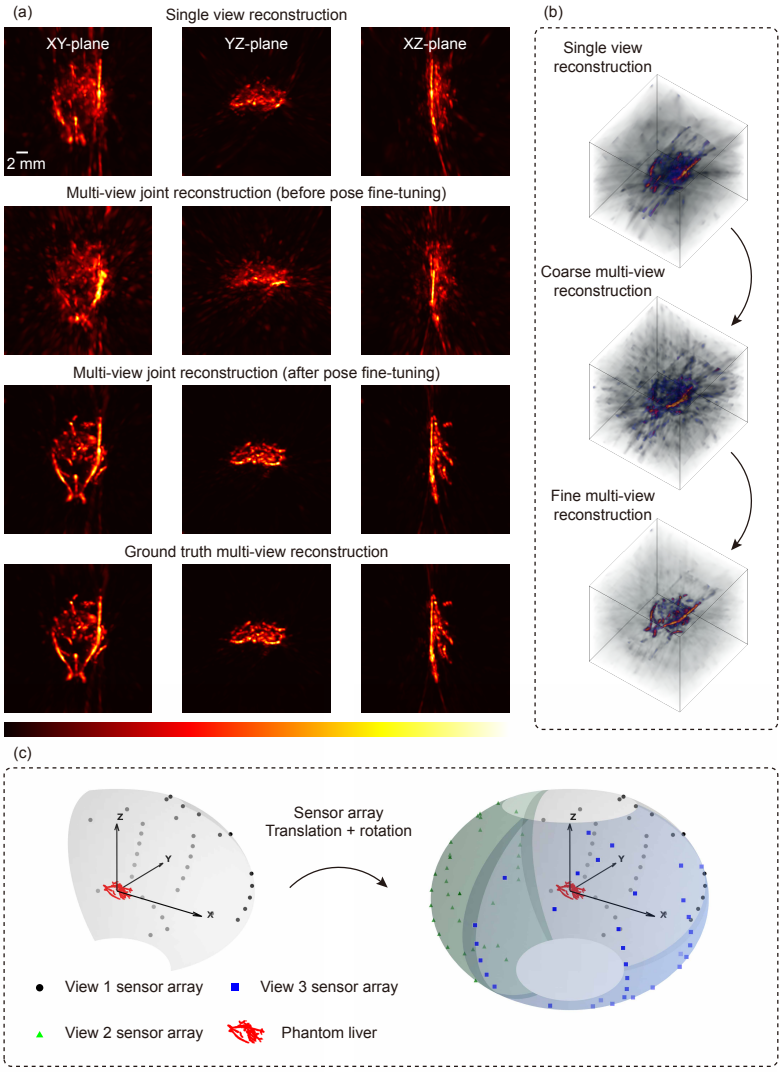

Fig. 2 The reconstruction results of multi-view simulation data with PA-SfM algorithm. (a) The single-view reconstruction, multi-view reconstruction using the PA-SfM coarse pose, multi-view reconstruction using the PA-SfM fine-tuned pose, and GT of the phantom liver (Scale: 2 mm). (b) The 3D results of single-view, PA-SfM coarse-pose multi-view, and PA-SfM fine-tuned multi-view reconstruction. (c) The multi-view setup of the simulation experiment.

along the azimuthal axis (Fig. 4(a-b)). The final *in-vivo* 3D reconstruction results are presented in Fig. 4(c-d) and Fig. 5(a-b).

To quantify the contribution of the final optimization stage within PA-SfM, we compared the single-view iterative reconstruction against multi-view joint reconstructions using the intermediate PA-SfM coarse pose, the final PA-SfM fine-tuned pose, and the GT pose. As detailed in Fig. 4(c), the respective rows display the MAPs of the

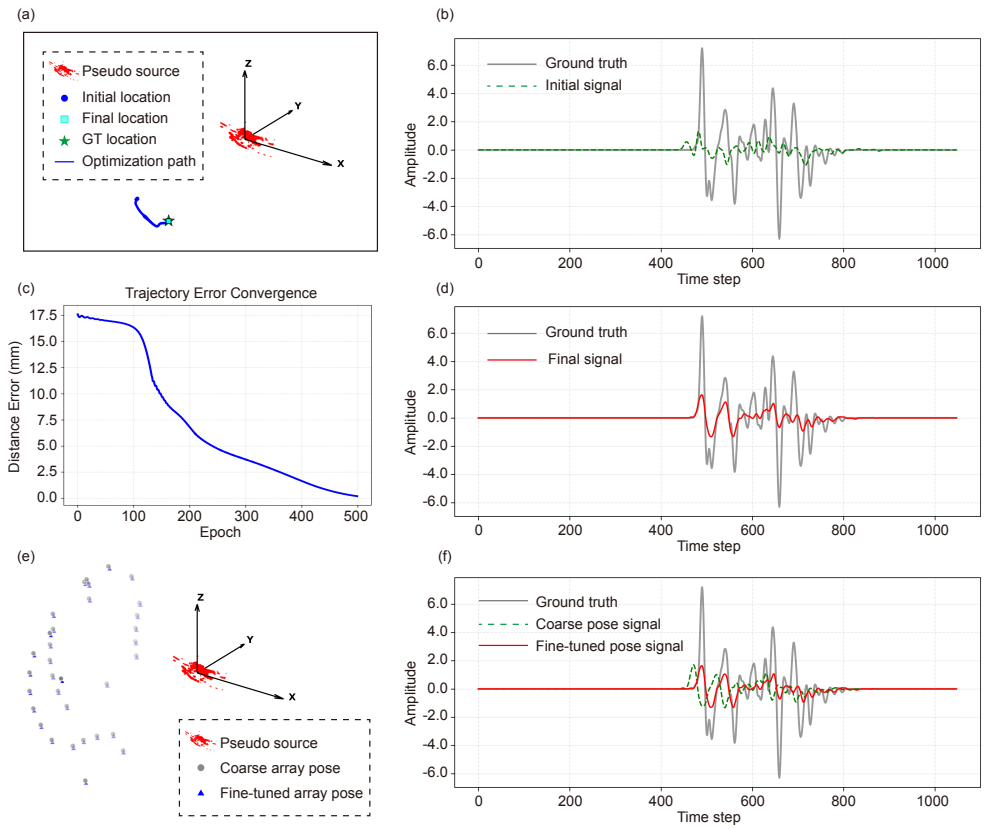

Fig. 3 The PA signal of the target sensor at different stages of the PA-SfM algorithm. (a) Differentiable single-sensor localization optimization. (b) The PA signal of the target sensor at the initial location. (c) The distance error curve of the single-sensor localization optimization. (d) The PA signal of the target sensor at the final location. (e) Global differentiable pose fine-tuning of the new pose. (f) The PA signal of the target sensor before and after the global differentiable pose fine-tuning.

reconstructed volumes for the rat kidney from the top, side, and front viewpoints using a hot colormap. Furthermore, a top-view MAP using a turbo colormap is explicitly provided to better visualize background noise and reconstruction artifacts.

As shown in Fig. 4(c-d), the single-view method provides limited structural information with prominent artifacts. In contrast, the multi-view fine-tuned-pose reconstruction successfully fuses the dual-pose data (View A + View B), effectively overcoming limited-view constraints. Consequently, the PA-SfM framework yields a highly accurate, artifact-free 3D representation of the *in-vivo* rat kidney vascular network, demonstrating high fidelity and structural integrity. Quantitative assessments based on the top-view MAP further corroborate these visual improvements. When evaluated against the GT multi-view reconstruction, the single-view approach yielded a PSNR of 30.20 dB and an SSIM of 0.7019. Similar to the simulation phase, the intermediate PA-SfM coarse-pose reconstruction still suffered from residual cross-view

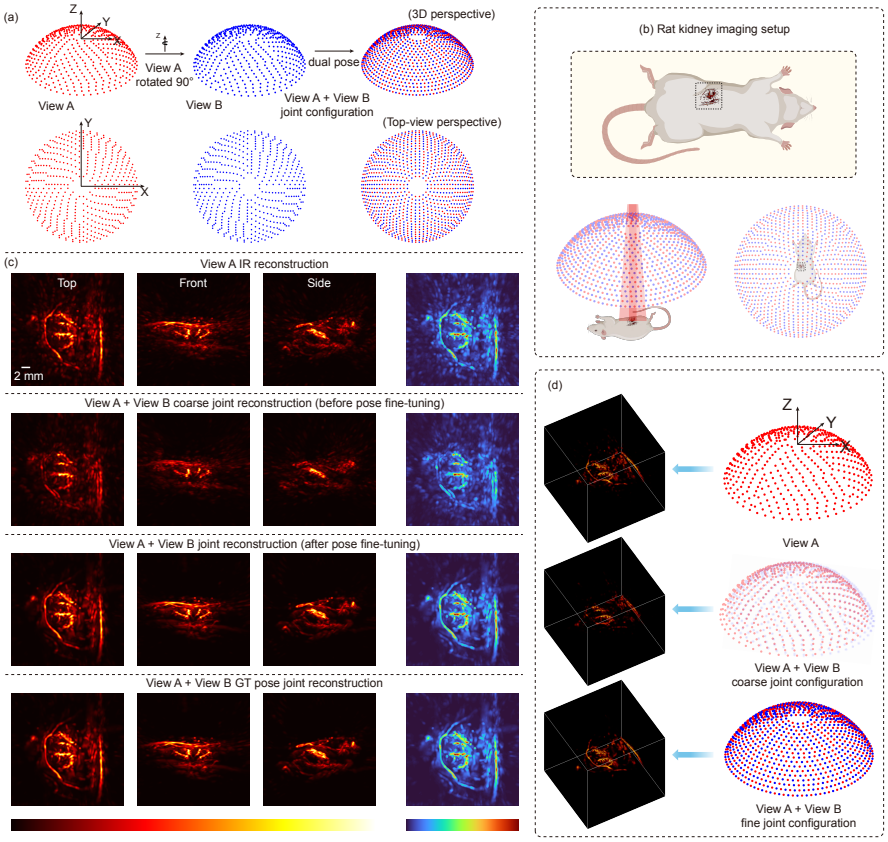

Fig. 4 The reconstruction results of multi-view *in-vivo* data with PA-SfM algorithm. (a) The dual-view 3D PAI for *in-vivo* rat kidney imaging. (b) The rat kidney imaging setup. (c) The single-view reconstruction, multi-view reconstruction using the PA-SfM coarse pose, multi-view reconstruction using the PA-SfM fine-tuned pose, and GT of the *in-vivo* rat kidney (Scale: 2 mm). (d) The 3D results of single-view, PA-SfM coarse-pose multi-view, and PA-SfM fine-tuned multi-view reconstruction.

misalignment, leading to degraded metrics (PSNR: 23.58 dB, SSIM: 0.6098). Remarkably, after applying the PA-SfM fine-tuning module, the quantitative performance surged to a PSNR of 41.42 dB and an SSIM of 0.9864.

To further demonstrate the robustness and generalizability of the proposed algorithm, we extended our *in-vivo* validation to the rat liver. Following an identical visualization scheme, Fig. 5(a-b) presents the liver reconstructions. Visually, the single-view liver reconstruction suffers from incomplete vessel connectivity and elevated background noise, which is clearly observable in the turbo colormap. While the intermediate PA-SfM coarse-pose multi-view fusion introduces additional vessel segments into the field of view, residual pose errors still disrupt vascular continuity, causing distinct ghosting artifacts and structural blurring. Conversely, the PA-SfM fine-tuned reconstruction eliminates these misalignment-induced artifacts, seamlessly bridges fragmented vessels, and accurately recovers the intricate vascular branching of the liver.

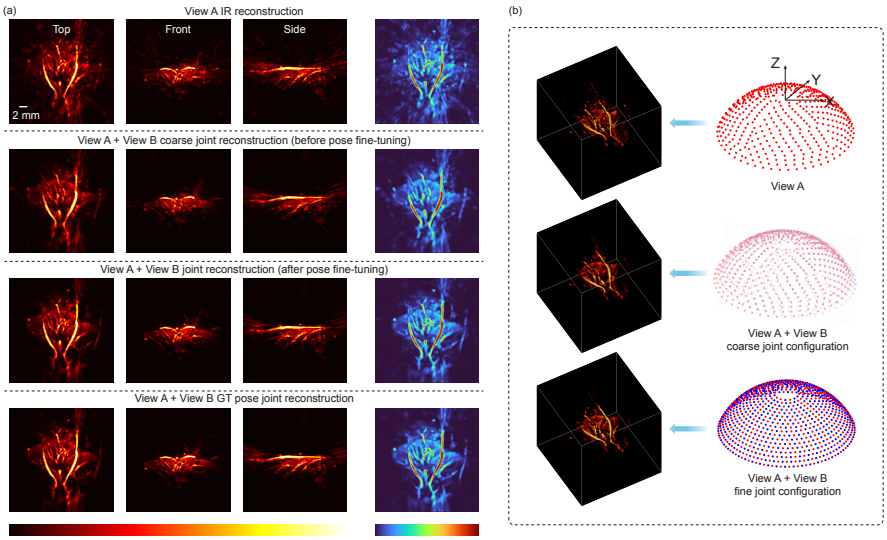

Fig. 5 The reconstruction results of multi-view *in-vivo* rat liver with PA-SfM algorithm. (a) The single-view reconstruction, multi-view reconstruction using the PA-SfM coarse pose, multi-view reconstruction using the PA-SfM fine-tuned pose, and GT of the *in-vivo* rat liver (Scale: 2 mm). (b) The 3D results of single-view, PA-SfM coarse-pose multi-view, and PA-SfM fine-tuned multi-view reconstruction.

These morphological observations are firmly supported by the quantitative metrics. Compared to the GT dual-pose reconstruction, the single-view liver MAP yielded a PSNR of 28.01 dB and an SSIM of 0.6947. Before the final global fine-tuning step, the intermediate PA-SfM coarse-pose reconstruction exhibited a significant drop in PSNR to 24.13 dB, although the SSIM slightly increased to 0.7900 due to the expanded spatial coverage. Ultimately, the PA-SfM fine-tuning process dramatically elevated the reconstruction accuracy, achieving a PSNR of 40.87 dB and an SSIM of 0.9841, thereby confirming its exceptional capability in mitigating motion artifacts and recovering complex biological structures across different organs.

3 Discussion

In this study, we developed PA-SfM, a tracker free framework that recovers array pose directly from photoacoustic data and enables high fidelity joint reconstruction across multiple views. Through simulations, *in vivo* rat experiments with known relative geometry, and validation in a mechanically rotated 3D PAI system, we showed that PA-SfM can accurately recover cross view geometry, suppress motion induced misalignment, and improve volumetric reconstruction quality. Most importantly, in the mechanically rotated array experiment, PA-SfM not only removed the need for explicit hardware based pose feedback, but also produced sharper and more continuous vascular structures than the hardware tracked reconstruction, indicating that pose estimation from the photoacoustic signals themselves can exceed the practical precision limit imposed by mechanical feedback. This finding is particularly relevant

for 3D photoacoustic imaging, where many systems require multiple poses to increase effective array density, enlarge the field of view, and mitigate limited view artifacts. At present, these additional poses are often provided by mechanical rotation, scanning stages, or external trackers, but such solutions increase hardware complexity, calibration burden, cost, and acquisition effort, while their final performance remains constrained by the accuracy of the tracking or actuation system. PA-SfM offers a different route by integrating a differentiable acoustic model, single sensor localization, rigid array estimation, and global fine tuning into one data driven pipeline, making multi pose reconstruction possible without relying on external tracking devices or mechanical encoder readout. The controlled simulation and *in vivo* experiments further showed that the intermediate PA-SfM coarse pose could recover the gross relation between views, whereas the final global fine tuning step was essential for removing residual pose error and reaching near GT reconstruction fidelity. These results suggest that PA-SfM is not only a motion correction method for freehand imaging, but also a general software based solution for multi pose 3D photoacoustic tomography. It provides researchers using two-dimensional matrix arrays for 3D photoacoustic imaging with an open source, low cost, and high precision multi-pose imaging solution, and offers a practical strategy for freehand 3D photoacoustic imaging.

4 Methods

The proposed PA-SfM framework establishes a complete closed-loop pipeline from reference reconstruction to unknown pose estimation, culminating in joint multi-view reconstruction (Fig. 6). The system is formulated upon a custom-built differentiable acoustic radiation engine and constrained by rigid-body geometry. The detailed algorithmic workflow is structured into the following interacting modules.

Differentiable Acoustic Radiation

At the core of the framework is a high-performance differentiable acoustic radiation engine that models forward acoustic wave propagation and computes gradients. To map the 3D photoacoustic initial pressure distribution $\mathbf{P}_c \in \mathbb{R}^{N_x \times N_y \times N_z}$ to the time-resolved sensor signals $p(\mathbf{x}_s, t)$ at sensor location $\mathbf{x}_s \in \mathbb{R}^3$, we employ a Gaussian kernel-based propagation model. The forward model is defined as:

$$p(\mathbf{x}_s, t) = \sum_{\mathbf{x} \in \Omega} \mathbf{P}_c(\mathbf{x}) \cdot \mathcal{G}(\|\mathbf{x} - \mathbf{x}_s\|, t; \sigma) \tag{1}$$

where Ω denotes the reconstructed volume, and $\mathcal{G}$ is the spatio-temporal Gaussian kernel parameterized by width σ.

The physical validity and mathematical rigor of this Gaussian kernel-based discrete model are rooted in the fundamental photoacoustic wave equation under far-field approximations. A comprehensive analytical derivation, which starts from the generalized spherically symmetric initial conditions and culminating in the exact GPU-accelerated far-field discrete model used in our engine, is provided in **Appendix A**.

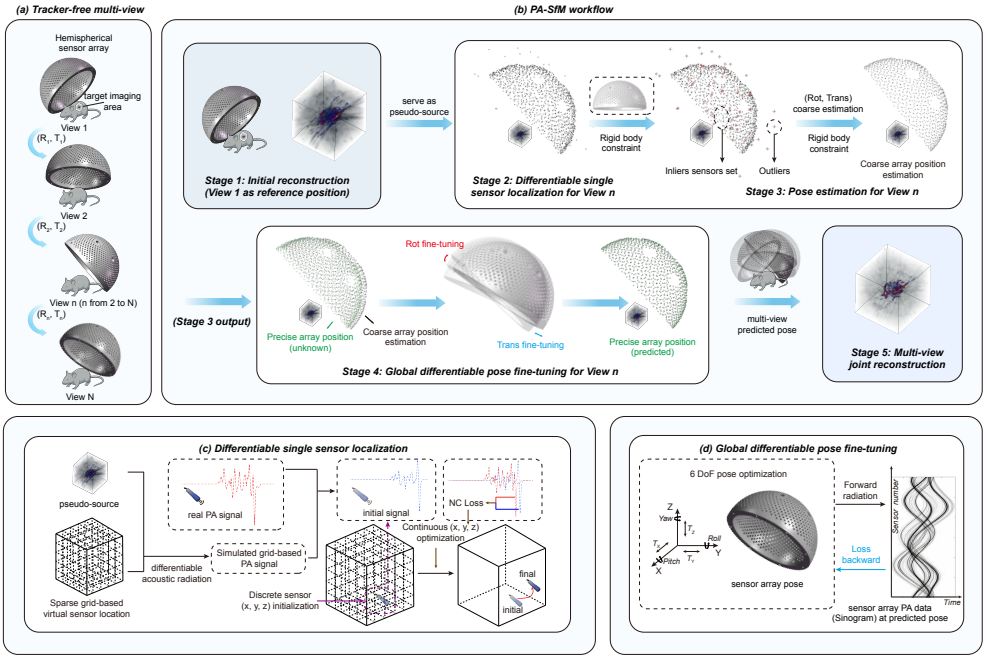

Fig. 6 The overview framework of PA-SfM algorithm for tracker-free multi-view 3D PAI. (a) The tracker-free multi-view 3D PAI. (b) The PA-SfM pipeline. (c) The differentiable single probe localization of PA-SfM. (d) The global differentiable pose fine-tuning of PA-SfM.

To achieve scalable and highly concurrent GPU acceleration, the forward and backward propagation operators are implemented as custom CUDA kernels using OpenAI Triton. By encapsulating these kernels within `torch.autograd.Function`, the engine seamlessly supports bidirectional gradient flow with respect to both the source intensity $\mathbf{P}_c$ and the sensor spatial coordinates $\mathbf{x}_s$. This automatic differentiation capability provides a unified mathematical foundation for both image reconstruction and sensor localization.

Stage 1: Initial Reference Reconstruction

The first stage establishes a structural reference map (coarse model) using data acquired at a known, fixed pose (Pose 1 or View 1). Given the known sensor array coordinates $\mathbf{X}_1$ and the corresponding measured signals $\mathbf{S}_1$, the initial source distribution $\mathbf{P}_c$ is initialized on a 256^3 voxel grid. The reconstruction is formulated as an inverse optimization problem:

$$\hat{\mathbf{P}}_c = \arg\min_{\mathbf{P}_c} \left(\|\mathcal{F}(\mathbf{P}_c, \mathbf{X}_1) - \mathbf{S}_1\|_2^2 + \lambda \mathcal{L}_{\mathrm{TGV}}(\mathbf{P}_c) \right) \tag{2}$$

where $\mathcal{F}$ represents the forward simulation engine, and $\mathcal{L}_{\text{TGV}}$ is the Total Generalized Variation (TGV) regularization term utilized to preserve structural edges while suppressing background noise. The objective function is iteratively minimized using the Adam optimizer, yielding the initial 3D anatomical structure serving as the reference map.

Stage 2: Differentiable Single Sensor Localization

To estimate the unknown sensor pose (Pose 2 or View 2), we independently localize each transducer element using the reference map $\hat{\mathbf{P}}_c$ obtained in Stage 1. By fixing $\hat{\mathbf{P}}_c$, the individual sensor coordinates $\mathbf{x}_{s,i}$ become the optimizable parameters. The localization is driven by a Negative Correlation (NC) loss:

$$\mathcal{L}_{\text{NC}}(\mathbf{x}_{s,i}) = -\frac{\text{Cov}(\mathcal{F}(\hat{\mathbf{P}}_c, \mathbf{x}_{s,i}), \mathbf{S}_{2,i})}{\sigma_{\text{sim}}\sigma_{\text{meas}}} \tag{3}$$

To circumvent local minima, a coarse-to-fine search strategy is employed, as demonstrated in Fig. 6(c):

1. **Global Coarse Search:** The target space is discretized into grids, and candidate locations are ranked based on $\mathcal{L}_{\text{NC}}$ to select the Top-K priors.
2. **Gradient-based Refinement:** Starting from the candidate priors, gradient descent is utilized to optimize the (x, y, z) coordinates.
3. **Dynamic Smoothing:** The Gaussian kernel width σ in Eq. (1) is dynamically decayed (from large to small) during optimization, acting as a simulated annealing mechanism to ensure convergence to the global optimum.

This stage outputs a set of independently predicted, albeit noisy, sensor coordinates $\tilde{\mathbf{X}}_2$.

Stage 3: Robust Rigid Array Estimation

To eliminate the localization noise introduced in Stage 2, the inherent geometric constraints of the rigid sensor array are leveraged. Given the predefined geometric template of the array $\mathbf{X}_{\text{template}}$, the problem is transformed into finding the optimal global rotation matrix $\mathbf{R} \in SO(3)$ and translation vector $\mathbf{t} \in \mathbb{R}^3$.

A modified Random Sample Consensus (RANSAC) [44] algorithm is implemented. To drastically improve search efficiency under low inlier ratios, a *geometric consistency pre-check* is introduced: randomly sampled point sets are evaluated based on their pairwise edge distances before performing Singular Value Decomposition (SVD). For valid subsets, the Kabsch algorithm [45] is applied to solve the rigid transformation. This stage yields a corrected global array coordinate set $\mathbf{X}_2^{\text{corr}}$ and reliably identifies a subset of inlier sensors $\mathcal{I}$.

Stage 4: Global Differentiable Fine-tuning

To further enhance spatial precision, a `RigidArrayOptimizer` module is constructed. The optimization variables are globally parameterized as Euler angles $\boldsymbol{\theta}$ and the translation vector $\mathbf{t}$ (Fig. 6(d)), ensuring that the array strictly adheres to rigid body kinematics: $\mathbf{X}_2(\boldsymbol{\theta}, \mathbf{t}) = \mathbf{R}(\boldsymbol{\theta})\mathbf{X}_{\text{template}} + \mathbf{t}$.

We introduce an inlier-masked gradient flow mechanism. The global loss is computed exclusively on the highly reliable signals from the inlier set $\mathcal{I}$ (identified in Stage 3):

$$\mathcal{L}_{\text{global}}(\boldsymbol{\theta}, \mathbf{t}) = \sum_{i \in \mathcal{I}} \mathcal{L}_{\text{NC}} \left(\mathbf{R}(\boldsymbol{\theta})\mathbf{x}_{\text{template},i} + \mathbf{t} \right) \tag{4}$$

Crucially, while the loss is masked, the gradients backpropagate to update the global parameters $(\boldsymbol{\theta}, \mathbf{t})$. This mechanism physically compels the unreliable (outlier) sensors to be guided into their correct anatomical positions by the reliable (inlier) sensors, yielding highly precise Pose 2 coordinates.

Stage 5: Joint Multi-view Reconstruction

In the final stage, the data from the known Pose 1 $(\mathbf{X}_1, \mathbf{S}_1)$ and the calibrated Pose 2 $(\mathbf{X}_2^{\text{fine}}, \mathbf{S}_2)$ are seamlessly fused. The objective function defined in Stage 1 is re-executed utilizing the augmented synthetic aperture. By coherently combining the dual-view data, limited-view artifacts are effectively mitigated, resulting in the final high-resolution, artifact-free 3D photoacoustic volume.

Acknowledgements

This research was supported by the following grants: the National Key R&D Program of China (No. 2023YFC2411700, No. 2017YFE0104200); the Beijing Natural Science Foundation (No. 7232177); the National Natural Science Foundation of China (62441204, 62472213); the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2020R1A6A1A03047902).

Declarations

Competing interests

All authors declare no competing interests.

Data availability

The data supporting the findings of this study are provided within the Article and its Supplementary Information. The raw and analysed datasets generated during the study are available for research purposes from the corresponding authors on reasonable request.

Algorithm 1 Differentiable Photoacoustic Structure from Motion (PA-SfM)

Input: Known Pose 1 data: $\mathbf{X}_1, \mathbf{S}_1$; Unknown Pose 2 signals: $\mathbf{S}_2$; Rigid array template: $\mathbf{X}_{\text{template}}$; Simulation engine $\mathcal{F}$ with TGV weight λ.

Output: High-resolution joint 3D volume $\mathbf{P}_{\text{final}}$; Fine-tuned Pose 2 array coordinates $\mathbf{X}_2^{\text{fine}}$.

Stage 1: Initial Reference Reconstruction
Initialize source distribution $\mathbf{P}_c$ on a 3D grid.
while not converged **do**
  $\hat{\mathbf{P}}_c \leftarrow \arg\min_{\mathbf{P}_c} \left(\|\mathcal{F}(\mathbf{P}_c, \mathbf{X}_1) - \mathbf{S}_1\|_2^2 + \lambda \mathcal{L}_{\text{TGV}}(\mathbf{P}_c) \right)$
end while

Stage 2: Differentiable Single Sensor Localization
Initialize empty noisy coordinate set $\tilde{\mathbf{X}}_2 \leftarrow \emptyset$
for each sensor $i = 1, 2, \ldots, N$ in Pose 2 **do**
  $\mathbf{x}_{s,i}^0 \leftarrow$ Top-1 candidate from Global Coarse Search using $\mathcal{L}_{\text{NC}}$
  for $\sigma = \sigma_{\max}$ down to $\sigma_{\min}$ **do** ▷ Dynamic Gaussian smoothing
    $\mathbf{x}_{s,i} \leftarrow \mathbf{x}_{s,i} - \eta \cdot \nabla_{\mathbf{x}_{s,i}} \mathcal{L}_{\text{NC}} \left(\mathcal{F}(\hat{\mathbf{P}}_c, \mathbf{x}_{s,i}), \mathbf{S}_{2,i}; \sigma \right)$
  end for
  $\tilde{\mathbf{X}}_2 \leftarrow \tilde{\mathbf{X}}_2 \cup \{\mathbf{x}_{s,i}\}$
end for

Stage 3: Robust Rigid Array Estimation
$(\mathbf{R}_{\text{init}}, \mathbf{t}_{\text{init}}, \mathcal{I}) \leftarrow$ Modified RANSAC$(\tilde{\mathbf{X}}_2, \mathbf{X}_{\text{template}})$
▷ $\mathcal{I}$ is the robust inlier subset passing geometric consistency pre-check

Stage 4: Global Differentiable Fine-tuning
Initialize Euler angles $\boldsymbol{\theta}$ from $\mathbf{R}_{\text{init}}$, and translation $\mathbf{t} \leftarrow \mathbf{t}_{\text{init}}$
while not converged **do**
  $\mathcal{L}_{\text{global}} \leftarrow \sum_{i \in \mathcal{I}} \mathcal{L}_{\text{NC}} \left(\mathbf{R}(\boldsymbol{\theta}) \mathbf{x}_{\text{template},i} + \mathbf{t} \right)$ ▷ Inlier-masked loss
  $\boldsymbol{\theta} \leftarrow \boldsymbol{\theta} - \alpha \cdot \nabla_{\boldsymbol{\theta}} \mathcal{L}_{\text{global}}$ ▷ Update global rotation
  $\mathbf{t} \leftarrow \mathbf{t} - \alpha \cdot \nabla_{\mathbf{t}} \mathcal{L}_{\text{global}}$ ▷ Update global translation
end while
$\mathbf{X}_2^{\text{fine}} \leftarrow \mathbf{R}(\boldsymbol{\theta}) \mathbf{X}_{\text{template}} + \mathbf{t}$ ▷ Kinematic constraint applied to ALL sensors

Stage 5: Joint Multi-view Reconstruction
$\mathbf{X}_{\text{joint}} \leftarrow \mathbf{X}_1 \cup \mathbf{X}_2^{\text{fine}}$
$\mathbf{S}_{\text{joint}} \leftarrow \mathbf{S}_1 \cup \mathbf{S}_2$
$\mathbf{P}_{\text{final}} \leftarrow \arg\min_{\mathbf{P}} \left(\|\mathcal{F}(\mathbf{P}, \mathbf{X}_{\text{joint}}) - \mathbf{S}_{\text{joint}}\|_2^2 + \lambda \mathcal{L}_{\text{TGV}}(\mathbf{P}) \right)$
return $\mathbf{P}_{\text{final}}$, $\mathbf{X}_2^{\text{fine}}$

Code availability

The source code for PA-SfM, along with supplementary materials and demonstration videos, has been made publicly available in the following GitHub repository: https://github.com/JaegerCQ/PA-SfM.

Author contributions

Shuang Li and Jian Gao conceived and designed the study and the algorithms. Chulhong Kim, Seongwook Choi, Hao Huang, Xuanhao Wang and Junhui Shi provided the *in-vivo* experimental data. Qian Chen, Yibing Wang, Shuang Wu and Yu Zhang contributed to the simulation experiments. Tingting Huang, Yucheng Zhou and Boxin Yao contributed to the metrics calculation. Changhui Li and Yao Yao supervised the study. All of the authors contributed to writing the paper.

A Analytical Expression for Spherically Symmetric Photoacoustic Sources: theoretical foundation for the differentiable acoustic radiation

In this appendix, we present a comprehensive analytical derivation of the spatiotemporal acoustic pressure generated by photoacoustic sources, which serves as the theoretical foundation for the differentiable acoustic radiation engine proposed in Section 2. We start from the fundamental photoacoustic wave equation, derive closed-form solutions for spherically symmetric distributions (including the Gaussian model used in our framework), and systematically formulate the far-field approximations that enable the ultrafast GPU-accelerated forward simulation.

A.1 Theoretical Background

A.1.1 Photoacoustic Wave Equation

The pressure wave generated by the photoacoustic effect satisfies the following wave equation [46, 47]:

$$\left(\nabla^2 - \frac{1}{v_s^2}\frac{\partial^2}{\partial t^2}\right) p(\mathbf{r}, t) = -\frac{\beta}{C_p}\frac{\partial H(\mathbf{r}, t)}{\partial t} \tag{5}$$

where $p(\mathbf{r}, t)$ is the acoustic pressure, v_s is the speed of sound, β is the thermal expansion coefficient, C_p is the specific heat capacity, and $H(\mathbf{r}, t)$ is the heating function. For instantaneous heating, the initial conditions are:

$$p(\mathbf{r}, 0) = p_0(\mathbf{r}), \quad \frac{\partial p}{\partial t}(\mathbf{r}, 0) = 0 \tag{6}$$

A.1.2 Fundamental Integral Formula

The solution to the photoacoustic wave equation can be expressed using Green's function as [46, 47]:

$$p(\mathbf{r},t) = \frac{1}{4\pi v_s^2}\frac{\partial}{\partial t}\left[\frac{1}{v_s t}\int d\mathbf{r}'\, p_0(\mathbf{r}')\,\delta\!\left(t - \frac{|\mathbf{r}-\mathbf{r}'|}{v_s}\right)\right] \tag{7}$$

where $p_0(\mathbf{r}')$ is the initial pressure distribution.

A.2 Derivation for Spherically Symmetric Case

Assume the initial pressure distribution is spherically symmetric about the origin, $p_0(\mathbf{r}') = p_0(r')$, where $r' = |\mathbf{r}'|$ and $r = |\mathbf{r}|$. Using the scaling property of the delta function, $\delta(t - |\mathbf{r}-\mathbf{r}'|/v_s) = v_s\,\delta(|\mathbf{r}-\mathbf{r}'| - v_s t)$, and substituting it into Eq. 7, we get:

$$p(\mathbf{r},t) = \frac{1}{4\pi v_s^2}\frac{\partial}{\partial t}\left[\frac{1}{t}\int d\mathbf{r}'\, p_0(r')\,\delta(|\mathbf{r}-\mathbf{r}'| - v_s t)\right] \tag{8}$$

By transforming to spherical coordinates (r', θ, ϕ) and applying the law of cosines $|\mathbf{r}-\mathbf{r}'| = \sqrt{r^2 + r'^2 - 2rr'\cos\theta}$, the delta function can be transformed (see Section A.6) as:

$$\delta(|\mathbf{r}-\mathbf{r}'| - v_s t) = \frac{v_s t}{rr'}\,\delta\!\left(\cos\theta - \frac{r^2 + r'^2 - (v_s t)^2}{2rr'}\right) \tag{9}$$

Integrating over solid angles yields $\int d\Omega\,\delta(\dots) = \frac{2\pi v_s t}{rr'}$ under the condition $|r - v_s t| \le r' \le r + v_s t$. Substituting this back into the pressure expression and simplifying yields:

$$p(r,t) = \frac{1}{2v_s r}\frac{\partial}{\partial t}\left[\int_{|r-v_s t|}^{r+v_s t} r' p_0(r')\,dr'\right] \tag{10}$$

Applying Leibniz's rule for differentiating under the integral sign gives the unified analytical expression for the spherically symmetric case:

$$\boxed{p(r,t) = \frac{1}{2r}\left[(r+v_s t)p_0(r+v_s t) + (r-v_s t)p_0(|r-v_s t|)\right]} \tag{11}$$

A.3 Analytical Solutions for Typical Initial Distributions

Based on Eq. 11, explicit analytical solutions can be derived for several commonly encountered distributions:

1. Uniform Spherical Source ($p_0(r) = p_0\,U(a_0 - r)$): For an observation point outside the sphere ($r > a_0$), $p(r,t) = \frac{p_0}{2r}(r - v_s t)$ when $r - a_0 \le v_s t \le r + a_0$, and 0 otherwise.

2. Gaussian Distribution ($p_0(r) = p_c\,\exp\!\left(-r^2/2\sigma^2\right)$):

$$p(r,t) = \frac{p_c}{2r}\left[(r+v_s t)e^{-\frac{(r+v_s t)^2}{2\sigma^2}} + (r-v_s t)e^{-\frac{(r-v_s t)^2}{2\sigma^2}}\right] \tag{12}$$

3. Exponential Distribution ($p_0(r) = p_c\, e^{-r/a}$):

$$p(r,t) = \frac{p_c}{2r}\left[(r+v_s t)e^{-(r+v_s t)/a} + (r-v_s t)e^{-|r-v_s t|/a}\right] \tag{13}$$

4. Power-Law Distribution ($p_0(r) = A/(r^2+a^2)^\nu, \nu > 1/2$):

$$p(r,t) = \frac{A}{2r}\left[\frac{r+v_s t}{\left((r+v_s t)^2+a^2\right)^\nu} + \frac{r-v_s t}{\left((r-v_s t)^2+a^2\right)^\nu}\right] \tag{14}$$

A.4 Far-Field Approximations

Under far-field conditions ($r \gg$ characteristic source dimension and $t \approx r/v_s$), the first term in Eq. 11 (representing an inward converging wave) is exponentially or power-law suppressed. The dominant contribution comes from the second term, leading to the unified far-field approximation:

$$p(r,t) \approx \frac{1}{2r}(r-v_s t)p_0(|r-v_s t|) \tag{15}$$

Applying this to the **Gaussian distribution** yields the outward propagating Gaussian pulse:

$$p(r,t) \approx \frac{p_c}{2r}(r-v_s t)\exp\left(-\frac{(r-v_s t)^2}{2\sigma^2}\right) \tag{16}$$

which provides the direct mathematical justification for the Gaussian kernel $\mathcal{G}$ defined in Eq. 1 of the main text.

A.5 GPU-Accelerated Forward Simulation Framework

Although the analytical expressions provide closed-form solutions, practical PA-SfM simulations involve a very large number of discrete source elements (N_s) and detector locations (N_d). To address this, we discretize the far-field approximation (Eq. 15) to build the GPU-accelerated forward model. The acoustic pressure measured at a detector position $\mathbf{r}_d$ becomes a superposition of outgoing waves emitted by all source elements $\mathbf{r}_k$ with amplitudes $P_c^{(k)}$:

$$p(\mathbf{r}_d,t) \approx \sum_{k=1}^{N_s} \frac{1}{2r_k}\,(r_k - v_s t)\,P_c^{(k)}\,\mathcal{K}(|r_k - v_s t|)\,, \tag{17}$$

where $r_k = \|\mathbf{r}_d - \mathbf{r}_k\|$ and $\mathcal{K}(\cdot)$ denotes the designated kernel function. Based on Eq. 16, the Gaussian kernel is defined as $\mathcal{K}_{\mathrm{G}}(\Delta r) = \exp\!\left(-\Delta r^2/2\sigma^2\right)$.

This discrete model exhibits two independent dimensions of parallelism: the detector index and the temporal sampling index. In our PyTorch-Triton implementation, a two-dimensional GPU execution grid is employed. Distances r_k and time delays are computed on-the-fly, avoiding the explicit construction of large distance matrices. This design enables the efficient evaluation of $\mathcal{O}(N_s \times N_d \times N_t)$ operations, seamlessly

supporting bidirectional gradient flow for the PA-SfM framework. Open-source implementations for this generalized ultrafast forward simulation are also publicly available at https://github.com/JaegerCQ/PA-SfM.

A.6 Detailed Delta Function Derivations

Starting from $\delta(\sqrt{A}-R)$ where $A = r^2+r'^2-2rr'\cos\theta$ and $R = v_s t$. Let $g(A) = \sqrt{A}-R$, then $\frac{dg}{dA} = \frac{1}{2\sqrt{A}}$. At $A = R^2$, $\left|\frac{dg}{dA}\right| = \frac{1}{2R}$. Using the delta function transformation formula:

$$\delta(g(A)) = \frac{\delta(A-R^2)}{|g'(R^2)|} = 2R\,\delta(A-R^2) \tag{18}$$

Next, let $h(\cos\theta) = r^2 + r'^2 - R^2 - 2rr'\cos\theta$. The zero point is $\cos\theta_0 = \frac{r^2+r'^2-R^2}{2rr'}$ and the derivative is $\left|\frac{dh}{d(\cos\theta)}\right| = 2rr'$. Therefore:

$$\delta(h(\cos\theta)) = \frac{1}{2rr'}\,\delta\left(\cos\theta - \frac{r^2+r'^2-R^2}{2rr'}\right) \tag{19}$$

Combining these factors, we obtain:

$$\delta(|\mathbf{r}-\mathbf{r}'|-R) = \frac{R}{rr'}\,\delta\left(\cos\theta - \frac{r^2+r'^2-R^2}{2rr'}\right) \tag{20}$$

A.7 Nomenclature

Symbol	Description
$p(\mathbf{r},t)$	Acoustic pressure at position $\mathbf{r}$ and time t
$p_0(\mathbf{r})$	Initial pressure distribution
v_s	Speed of sound
β	Thermal expansion coefficient
C_p	Specific heat capacity
r	Radial distance from origin
t	Time
$\delta(x)$	Dirac delta function
σ	Gaussian width parameter
a_0	Radius of uniform sphere

Please refer to our separate Supplementary Information file for further details.

References

[1] Park, J., Choi, S., Knieling, F., Clingman, B., Bohndiek, S., Wang, L.V., Kim, C.: Clinical translation of photoacoustic imaging. Nature Reviews Bioengineering **3**(3), 193–212 (2025)

[2] Wang, L.V., Hu, S.: Photoacoustic tomography: in vivo imaging from organelles to organs. science **335**(6075), 1458–1462 (2012)

[3] Assi, H., Cao, R., Castelino, M., Cox, B., Gilbert, F.J., Gröhl, J., Gurusamy, K., Hacker, L., Ivory, A.M., Joseph, J., *et al.*: A review of a strategic roadmapping exercise to advance clinical translation of photoacoustic imaging: from current barriers to future adoption. Photoacoustics **32**, 100539 (2023)

[4] Deán-Ben, X., Gottschalk, S., Mc Larney, B., Shoham, S., Razansky, D.: Advanced optoacoustic methods for multiscale imaging of in vivo dynamics. Chemical Society Reviews **46**(8), 2158–2198 (2017)

[5] Lin, L., Wang, L.V.: The emerging role of photoacoustic imaging in clinical oncology. Nature Reviews Clinical Oncology **19**(6), 365–384 (2022)

[6] Ntziachristos, V.: Addressing unmet clinical need with optoacoustic imaging. Nature Reviews Bioengineering **3**(3), 182–184 (2025)

[7] Wray, P., Lin, L., Hu, P., Wang, L.V.: Photoacoustic computed tomography of human extremities. Journal of biomedical optics **24**(2), 026003–026003 (2019)

[8] Li, S., Zhang, G., Wang, Y., Li, W., Sun, Y., Li, C.: Photoacoustic imaging of peripheral vessels in extremities by large-scale synthetic matrix array. Journal of Biomedical Optics **29**(S1), 11519–11519 (2024)

[9] Wang, M., Zhao, L., Wei, Y., Li, J., Qi, Z., Su, N., Zhao, C., Zhang, R., Tang, T., Liu, S., *et al.*: Functional photoacoustic/ultrasound imaging for the assessment of breast intraductal lesions: preliminary clinical findings. Biomedical Optics Express **12**(3), 1236–1246 (2021)

[10] Han, T., Yang, M., Yang, F., Zhao, L., Jiang, Y., Li, C.: A three-dimensional modeling method for quantitative photoacoustic breast imaging with handheld probe. Photoacoustics **21**, 100222 (2021)

[11] Sun, Y., Wang, Y., Li, W., Li, C.: Real-time dual-modal photoacoustic and fluorescence small animal imaging. Photoacoustics **36**, 100593 (2024)

[12] Matsumoto, Y., Asao, Y., Yoshikawa, A., Sekiguchi, H., Takada, M., Furu, M., Saito, S., Kataoka, M., Abe, H., Yagi, T., *et al.*: Label-free photoacoustic imaging of human palmar vessels: a structural morphological analysis. Scientific reports **8**(1), 786 (2018)

[13] Matsumoto, Y., Asao, Y., Sekiguchi, H., Yoshikawa, A., Ishii, T., Nagae, K.-i., Kobayashi, S., Tsuge, I., Saito, S., Takada, M., *et al.*: Visualising peripheral arterioles and venules through high-resolution and large-area photoacoustic imaging. Scientific reports **8**(1), 14930 (2018)

[14] Ivankovic, I., Merčep, E., Schmedt, C.-G., Deán-Ben, X.L., Razansky, D.: Real-time volumetric assessment of the human carotid artery: handheld multispectral optoacoustic tomography. Radiology **291**(1), 45–50 (2019)

[15] Deán-Ben, X.L., Razansky, D.: Portable spherical array probe for volumetric real-time optoacoustic imaging at centimeter-scale depths. Optics express **21**(23), 28062–28071 (2013)

[16] Nagae, K., Asao, Y., Sudo, Y., Murayama, N., Tanaka, Y., Ohira, K., Ishida, Y., Otsuka, A., Matsumoto, Y., Saito, S., et al.: Real-time 3d photoacoustic visualization system with a wide field of view for imaging human limbs. F1000Research **7** (2018)

[17] Kim, W., Choi, W., Ahn, J., Lee, C., Kim, C.: Wide-field three-dimensional photoacoustic/ultrasound scanner using a two-dimensional matrix transducer array. Optics Letters **48**(2), 343–346 (2023)

[18] Piras, D., Xia, W., Steenbergen, W., Van Leeuwen, T.G., Manohar, S.: Photoacoustic imaging of the breast using the twente photoacoustic mammoscope: present status and future perspectives. IEEE Journal of Selected Topics in Quantum Electronics **16**(4), 730–739 (2009)

[19] Heijblom, M., Piras, D., Xia, W., Van Hespen, J., Klaase, J., Engh, F., Van Leeuwen, T., Steenbergen, W., Manohar, S.: Visualizing breast cancer using the twente photoacoustic mammoscope: what do we learn from twelve new patient measurements? Optics express **20**(11), 11582–11597 (2012)

[20] Wang, Y., Li, C.: Comprehensive framework of gpu-accelerated image reconstruction for photoacoustic computed tomography. Journal of Biomedical Optics **29**(6), 066006–066006 (2024)

[21] Wang, X., Meng, Y., Sun, M., Gao, X., Wang, Y., Wang, S., Wang, K., Wang, R., Ren, D., Yin, Y., *et al.*: Cross-regional real-time visualization of systemic physiology and dynamics with 3d panoramic photoacoustic computed tomography (3d-panopact). Nature Communications **16**(1), 10077 (2025)

[22] Zhang, Y., Na, S., Russin, J.J., Sastry, K., Lin, L., Zheng, J., Luo, Y., Tong, X., An, Y., Hu, P., et al.: Rotational ultrasound and photoacoustic tomography of the human body. Nature Biomedical Engineering, 1–12 (2026)

[23] Dou, Y., Mu, F., Li, Y., Varghese, T.: Sensorless end-to-end freehand three-dimensional ultrasound reconstruction with physics guided deep learning. IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control (2024)

[24] Lee, C., Kim, C., Park, B.: Review of three-dimensional handheld photoacoustic and ultrasound imaging systems and their applications. Sensors **23**(19), 8149 (2023)

[25] Lasso, A., Heffter, T., Rankin, A., Pinter, C., Ungi, T., Fichtinger, G.: Plus: open-source toolkit for ultrasound-guided intervention systems. IEEE transactions on biomedical engineering **61**(10), 2527–2537 (2014)

[26] Frey, B.N., Hyun, D., Simson, W., Zhuang, L., Hashemi, H.S., Schneider, M., Dahl, J.J.: Ultraflex: Iterative model-based ultrasonic flexible-array shape calibration. IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control **72**(11), 1462–1475 (2025)

[27] El Hadramy, S., Verde, J., Beaudet, K.-P., Padoy, N., Cotin, S.: Trackerless volume reconstruction from intraoperative ultrasound images. In: International Conference on Medical Image Computing and Computer-assisted Intervention, pp. 303–312 (2023). Springer

[28] Li, Q., Shen, Z., Li, Q., Barratt, D.C., Dowrick, T., Clarkson, M.J., Vercauteren, T., Hu, Y.: Trackerless freehand ultrasound with sequence modelling and auxiliary transformation over past and future frames. In: 2023 IEEE 20th International Symposium on Biomedical Imaging (ISBI), pp. 1–5 (2023). IEEE

[29] He, W., Qiu, R.L., Zhang, C., Shah, K.D., Wang, T., Jani, A.B., Liu, T., Patel, P., Yang, X.: Trackerless 3d ultrasound volume reconstruction from 2d freehand scans using a hybrid transformer–cnn framework. Medical Physics **52**(11), 70103 (2025)

[30] Blessberger, H., Binder, T.: Two dimensional speckle tracking echocardiography: basic principles. Heart **96**(9), 716–722 (2010)

[31] Mondillo, S., Galderisi, M., Mele, D., Cameli, M., Lomoriello, V.S., Zacà, V., Ballo, P., D'Andrea, A., Muraru, D., Losi, M., *et al.*: Speckle-tracking echocardiography: a new technique for assessing myocardial function. Journal of ultrasound in medicine **30**(1), 71–83 (2011)

[32] Yoo, J.-C., Han, T.H.: Fast normalized cross-correlation. Circuits, systems and signal processing **28**(6), 819–843 (2009)

[33] Eng, J.K., Fischer, B., Grossmann, J., MacCoss, M.J.: A fast sequest cross correlation algorithm. Journal of proteome research **7**(10), 4598–4602 (2008)

[34] Lee, S., Kim, S., Seo, M., Park, S., Imrus, S., Ashok, K., Lee, D., Park, C., Lee, S., Kim, J., et al.: Enhancing free-hand 3d photoacoustic and ultrasound reconstruction using deep learning. IEEE Transactions on Medical Imaging (2025)

[35] Knauer, N., Dean-Ben, X.L., Razansky, D.: Spatial compounding of volumetric data enables freehand optoacoustic angiography of large-scale vascular networks. IEEE Transactions on Medical Imaging **39**(4), 1160–1169 (2019)

[36] Choi, S., Yang, J., Lee, S.Y., Kim, J., Lee, J., Kim, W.J., Lee, S., Kim, C.: Deep learning enhances multiparametric dynamic volumetric photoacoustic computed tomography in vivo (dl-pact). Advanced Science **10**(1), 2202089 (2023)

[37] Choi, W., Park, B., Choi, S., Oh, D., Kim, J., Kim, C.: Recent advances in contrast-enhanced photoacoustic imaging: overcoming the physical and practical challenges. Chemical Reviews **123**(11), 7379–7419 (2023)

[38] Kim, J., Lee, J., Choi, S., Lee, H., Yang, J., Jeon, H., Sung, M., Kim, W.J., Kim, C.: 3d multiparametric photoacoustic computed tomography of primary and metastatic tumors in living mice. ACS nano (2024)

[39] Kim, J., Kim, G., Li, L., Zhang, P., Kim, J.Y., Kim, Y., Kim, H.H., Wang, L.V., Lee, S., Kim, C.: Deep learning acceleration of multiscale superresolution localization photoacoustic imaging. Light: science & applications **11**(1), 131 (2022)

[40] Choi, S., Kim, J., Jeon, H., Lee, Y., Lim, G., Ju, W.-M., Kim, K., Lee, D.S., Lee, Y.-S., Lee, J.S., et al.: Photoacoustic computed tomography monitors cerebrospinal fluid dynamics and glymphatic function. Nature Communications (2026)

[41] Park, J., Oh, D., Yoo, J., Ha, H., Kim, D., Kim, H.H., Myung, Y., Kim, C.: Clinical ultrasound, photoacoustic, and fluorescence image-guided lymphovenous anastomosis microsurgery via a transparent ultrasound transducer array. Nature Communications **16**(1), 9853 (2025)

[42] Yang, J., Choi, S., Kim, J., Lee, J., Kim, W.J., Kim, C.: Multiplane spectroscopic whole-body photoacoustic computed tomography of small animals in vivo. Laser & Photonics Reviews **19**(3), 2400672 (2025)

[43] Kim, J., Kweon, J.Y., Choi, S., Jeon, H., Sung, M., Gao, R., Liu, C., Kim, C., Ahn, Y.J.: Non-invasive photoacoustic cerebrovascular monitoring of early-stage ischemic strokes in vivo. Advanced Science **12**(4), 2409361 (2025)

[44] Chum, O., Matas, J., Kittler, J.: Locally optimized ransac. In: Joint Pattern Recognition Symposium, pp. 236–243 (2003). Springer

[45] Kabsch, W.: Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. Applied Crystallography **26**(6), 795–800 (1993)

[46] Xu, M., Wang, L.V.: Universal back-projection algorithm for photoacoustic computed tomography. Physical Review E—Statistical, Nonlinear, and Soft Matter Physics **71**(1), 016706 (2005)

[47] Xu, M., Wang, L.V.: Photoacoustic imaging in biomedicine. Review of scientific

instruments **77**(4) (2006)