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

Shuang Li[1†], Jian Gao[2†], Chulhong Kim[3], Seongwook Choi[3], 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, 4*]

[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*]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

Three-dimensional (3D) handheld photoacoustic tomography typically relies on bulky and expensive external positioning trackers to correct motion artifacts, which severely limits its clinical flexibility and accessibility. To address this challenge, we present PA-SfM, a tracker-free framework that leverages exclusively single-modality photoacoustic data for both sensor pose recovery and high-fidelity 3D reconstruction via differentiable acoustic radiation modeling. Unlike traditional Structure-from-Motion (SfM) methods that formulate pose estimation as a geometry-driven optimization over visual features, PA-SfM integrates the acoustic wave equation into a differentiable programming pipeline. By leveraging a high-performance, GPU-accelerated acoustic radiation kernel, the framework simultaneously optimizes the 3D photoacoustic source distribution and the sensor array pose via gradient descent. To ensure robust convergence in freehand

scenarios, we introduce a coarse-to-fine optimization strategy that incorporates geometric consistency checks and rigid-body constraints to eliminate motion outliers. We validated the proposed method through both numerical simulations and *in-vivo* rat experiments. The results demonstrate that PA-SfM achieves sub-millimeter positioning accuracy and restores high-resolution 3D vascular structures comparable to ground-truth benchmarks, offering a low-cost, software-defined solution for clinical freehand photoacoustic imaging. The source code is publicly available at https://github.com/JaegerCQ/PA-SfM.

Keywords: Freehand 3D photoacoustic imaging, tracker-free imaging, differentiable acoustic radiation, structure-from-motion

1 Introduction

Photoacoustic imaging (PAI) is a hybrid imaging modality that leverages ultrasound detection alongside optical absorption contrast, enabling non-invasive visualization of biological tissues at centimeter-scale depths with spatial resolutions superior to conventional optical imaging techniques. Due to these advantages, PAI has found wide application in both preclinical studies and clinical settings [1–6]. Advancements in three-dimensional (3D) PAI using 2D matrix arrays have enabled *in-vivo* imaging of peripheral vessels [7, 8], breast tissues [9, 10], and small animals [11] with promising results. Various array designs, including spherical, planar [12–19], and synthetically scanned arrays [8, 20], have been explored to enhance 3D imaging capabilities.

Despite the diagnostic advantages of 3D photoacoustic tomography (PACT), its integration into routine clinical practice is often hindered by hardware constraints. Conventional high-fidelity 3D reconstruction typically necessitates the use of large-scale 2D matrix arrays or specialized motorized scanning stages to ensure precise spatial registration of captured data [1, 8, 21, 22]. While these systems provide controlled trajectories, they are often bulky, expensive, and lack the mechanical flexibility required to navigate irregular anatomical surfaces or operate in resource-limited settings [23]. Handheld, freehand scanning has emerged as a compelling alternative, offering clinicians unrestricted flexibility to cover large regions of interest. However, the absence of a fixed mechanical reference introduces significant out-of-plane motion artifacts. To achieve high-fidelity volumetric reconstruction, the sensor's six-degree-of-freedom (6-DoF) pose must be accurately recovered for every frame [23]. Failure to account for these pose variations leads to severe structural blurring, vessel fragmentation, and loss of quantitative accuracy.

To compensate for manual motion, traditional solutions rely on external positioning sensors such as optical tracking systems (OTS), electromagnetic (EM) trackers, GPS sensors or inertial measurement units (IMUs) [24, 25]. Although OTS can provide sub-millimeter precision, it requires a continuous line-of-sight between the camera and the probe markers, which is often impractical in crowded clinical environments or surgical suites [24, 26]. EM tracking avoids line-of-sight issues but is highly susceptible to electromagnetic interference from metallic medical tools or nearby electronic

equipment [25]. IMU-based solutions are cost-effective and portable but suffer from inherent cumulative drift, where small measurement errors integrate over time to cause significant positioning offsets [24, 27]. These hardware-based methods increase system complexity, necessitate tedious calibration procedures, and significantly raise the overall cost of the imaging platform.

In response, recent research has shifted toward tracker-free or sensorless pose estimation directly from acquired data [28–31]. In ultrasound imaging, speckle tracking [32, 33] is a well-established technique for motion estimation; however, photoacoustic signals are inherently speckle-free. Because PAI signals originate from the thermoelastic expansion of discrete absorbers, all initial pressure rises are positive, lacking the phase-randomized interference patterns that create ultrasound speckle. Consequently, traditional cross-correlation algorithms [34, 35] often fail in the photoacoustic domain. Emerging deep learning methods have attempted to learn motion priors from 2D images, yet these models typically require massive, high-quality labeled datasets that are difficult to obtain and fail to transfer across different imaging systems, which lack a rigorous connection to the underlying physics of acoustic wave propagation [23]. Furthermore, some researchers have utilized the frequency-domain characteristics of vascular networks in photoacoustic images to estimate the relative motion and orientation of the imaging probe [36], thereby enabling freehand imaging. Although this approach expands the field of view to some extent, significant limitations remain. Because the estimation relies on images rather than the intrinsic properties of the photoacoustic equation, this method strictly requires a high overlap ratio between different poses. Translation estimation may fail if the displacement is excessively large or if the images themselves lack distinct features. Additionally, the scanning direction cannot change abruptly, and the moving speed must be restricted.

Inspired by the success of differentiable rendering and structure-from-motion (SfM) in computer vision [37], we present PA-SfM, a tracker-free framework that leverages exclusively single-modality photoacoustic data for both sensor pose recovery and high-fidelity 3D reconstruction. Unlike traditional SfM methods based on visual features, PA-SfM integrates the acoustic wave equation into a differentiable programming pipeline. By representing the initial pressure field as a superposition of continuous Gaussian kernels, we utilize a GPU-accelerated acoustic radiation kernel to simultaneously optimize the 3D source distribution and the sensor array pose via gradient descent [38]. To ensure robust convergence in freehand scenarios, we introduce a coarse-to-fine optimization strategy incorporating geometric consistency checks and rigid-body constraints to eliminate motion outliers. We validated the proposed method through numerical simulations and *in-vivo* rat experiments, demonstrating sub-millimeter positioning accuracy. Furthermore, by utilizing a sparse planar array at multiple poses to synthesize an equivalent array with an ultra-wide view with PA-SfM, we achieved high-quality three-dimensional reconstruction. PA-SfM offers a low-cost, software-defined solution that significantly improves the clinical flexibility and accessibility of freehand 3D photoacoustic imaging.

2 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. 1). 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.

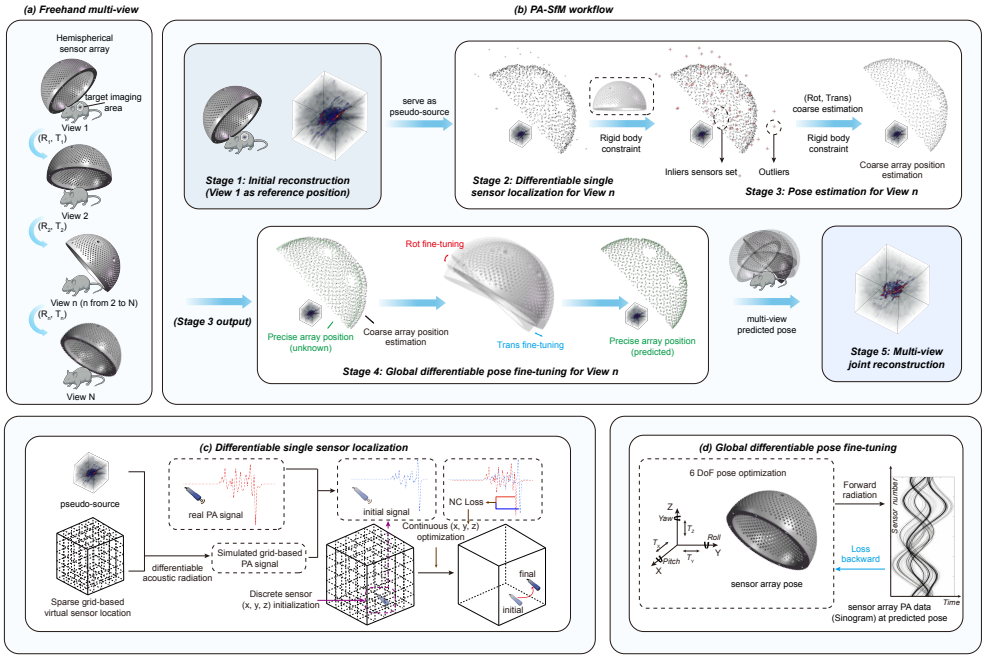

Fig. 1 The overview framework of PA-SfM algorithm for freehand multi-view 3D PAI. (a) The freehand multi-view 3D PAI. (b) The PA-SfM pipeline. (c) The differentiable single probe localization of PA-SfM. (d) The global differentiable pose fine-tuning of PA-SfM.

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**.

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 $\mathcal{L}_{\mathrm{TGV}}$ 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:

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

To circumvent local minima, a coarse-to-fine search strategy is employed, as demonstrated in Fig. 1(c):

1. **Global Coarse Search:** The target space is discretized into grids, and candidate locations are ranked based on $\mathcal{L}_{\mathrm{NC}}$ 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 $\hat{\mathbf{X}}_2$.

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) [39] 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 [40] 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. 1(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 B $(\mathbf{X}_2^{\text{fine}}, \mathbf{S}_2)$ 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.

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 $\mathbf{P}_{\text{final}}$, $\mathbf{X}_2^{\text{fine}}$

3 Results

3.1 3D PA image reconstruction under multi-view sparse hemispherical array

To evaluate the performance of the proposed PA-SfM algorithm, we first 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 during the simulation, Fig. 3 illustrates the PA signals of the target sensor at different optimization stages. The differentiable single-sensor localization effectively optimizes the sensor location, which is 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 persisted in the global pose. Consequently, the PA signal extracted from the coarsely predicted sensor pose exhibited a noticeable discrepancy compared to the actual ground-truth supervision signal (as indicated by the green dashed line for the coarse prediction and the solid gray line for the ground truth in Fig. 3(f)). At this coarse estimation stage, the maximum and minimum Euclidean distance errors for sensor localization were recorded as 0.002334 mm and 0.000265 mm, respectively. Through further Global Differentiable Fine-tuning, the alignment of the PA signals improves progressively; the fine-tuned pose signal closely matches the ground-truth signal compared to both the initial and coarse pose stages (Fig. 3(b-f)). Ultimately, the maximum and minimum sensor localization errors were substantially reduced to 0.001784 mm and 0.000135 mm, 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 Maximum Amplitude Projection (MAP) along the XY plane. While the multi-view joint reconstruction using the coarse pose estimation effectively expands the field of view, it still exhibits structural blurring due to 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 ground truth of the phantom liver, achieving a significantly improved PSNR of 38.90 dB and an SSIM of 0.9637.

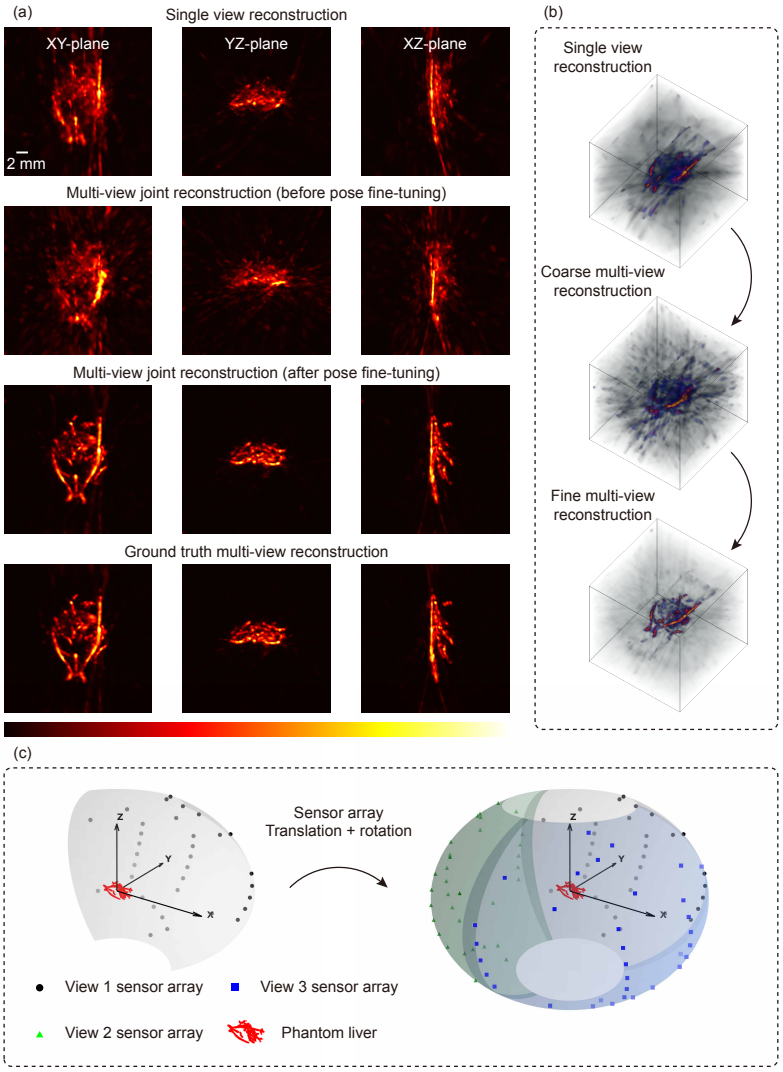

Fig. 2 The reconstruction results of multi-view simulation data with PA-SfM algorithm. (a) The single view reconstruction, multi-view coarse pose reconstruction, multi-view fine-tuned pose reconstruction and ground truth of the phantom liver (Scale: 2 mm). (b) The 3D results of single view, coarse multi-view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

3.2 In-vivo rat results of multi-view PA-SfM joint reconstruction

We further validated the clinical feasibility of the PA-SfM framework through *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)

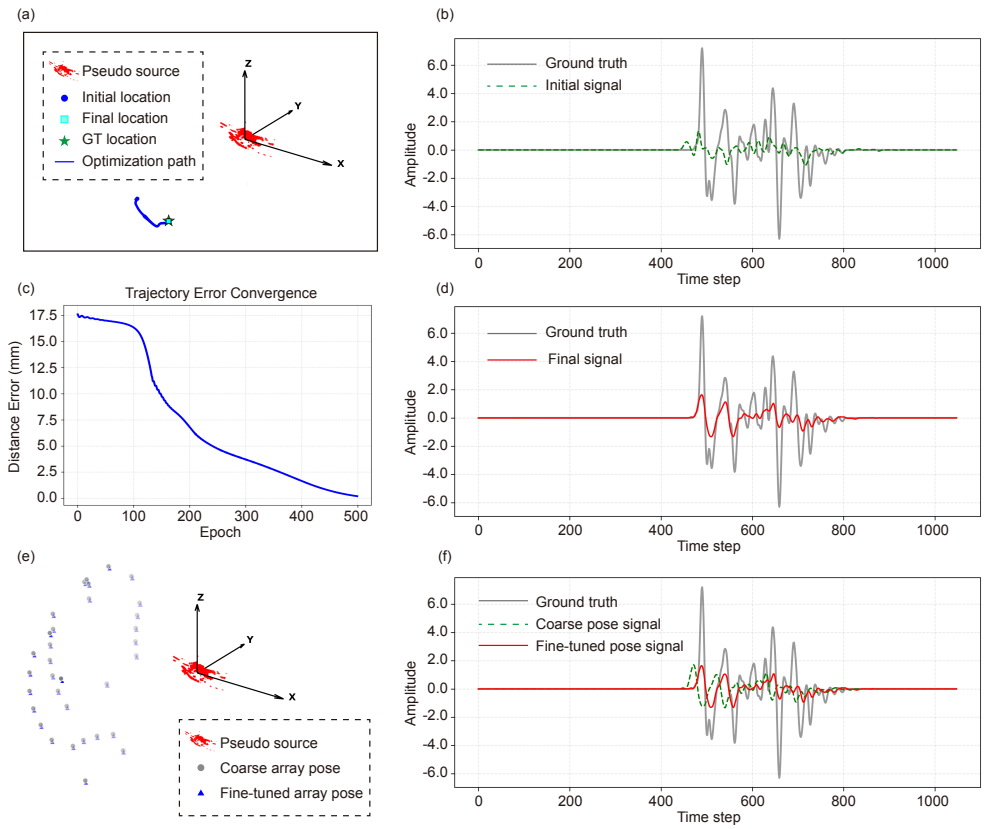

Fig. 3 The PA signal of the target sensor at different stages of the PA-SfM algorithm. (a) Differentiable single sensor location optimization. (b) The PA signal of the target sensor at the initial location. (c) The distance error curve of the single sensor location optimization. (d) The PA signal of 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.

transducer array with 1024 elements and a radius of 60 mm [41]. More details regarding the 3D PAI system and animal experiments can be found in literatures [42–48]. Specifically, to emulate a multi-view acquisition scheme under limited-view conditions, we deliberately partitioned the full 1024-element spherical array into two interleaved, 512-element sparse subsets. These two distinct subsets, designated as View A and View B, represent two independent imaging poses separated by a 90° relative rotation 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).

For a comprehensive evaluation, we compared the single-view Iterative Reconstruction against multi-view joint reconstructions utilizing the coarsely predicted pose, our PA-SfM fine-tuned pose, and the ground-truth (GT) pose. As detailed in Fig. 4(c), the respective rows display the MAPs of the reconstructed volumes for the rat kidney from the top (above), side, and front viewpoints utilizing a hot colormap. Furthermore,

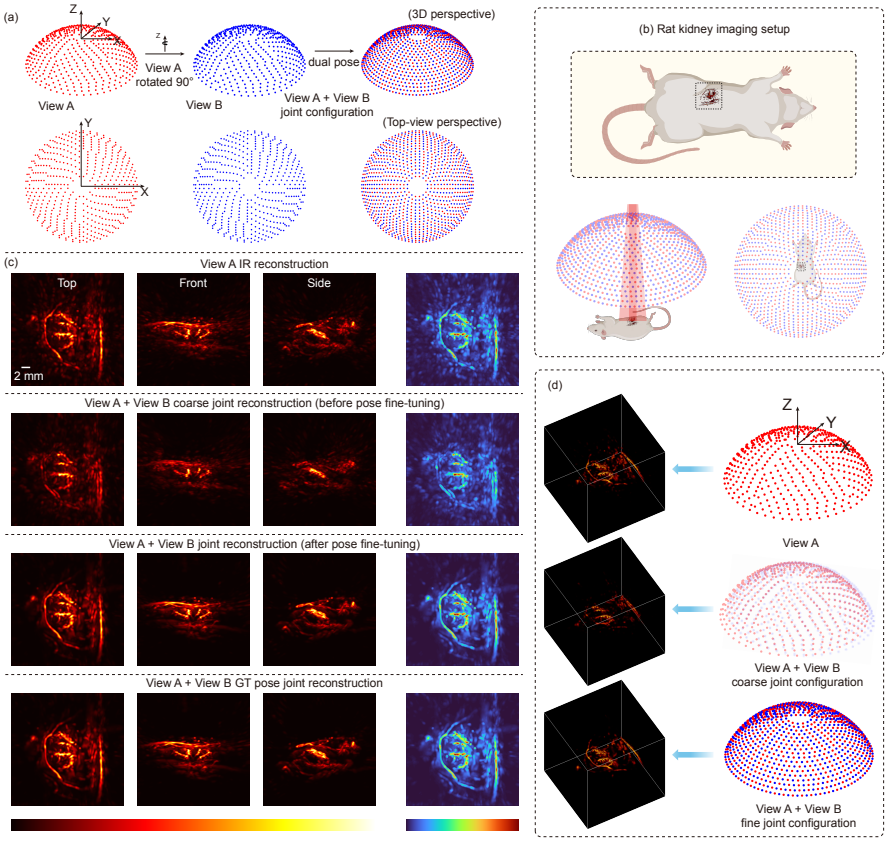

Fig. 4 The reconstruction results of multi-view *in-vivo* data with PA-SfM algorithm. (a) The free-hand multi-view 3D PAI for *in-vivo* rat kidney imaging. (b) The rat kidney imaging setup. (c) The single view reconstruction, multi-view coarse pose reconstruction, multi-view fine-tuned pose reconstruction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

a top-view MAP utilizing a turbo colormap is explicitly provided to better visualize background noise and reconstruction artifacts.

As shown in Fig. 4(c) and 4(d), single-view methods provide 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 uncalibrated coarse-pose multi-view reconstruction suffered from severe misalignment blurring, leading to degraded metrics (PSNR: 23.58 dB, SSIM: 0.6098). Remarkably,

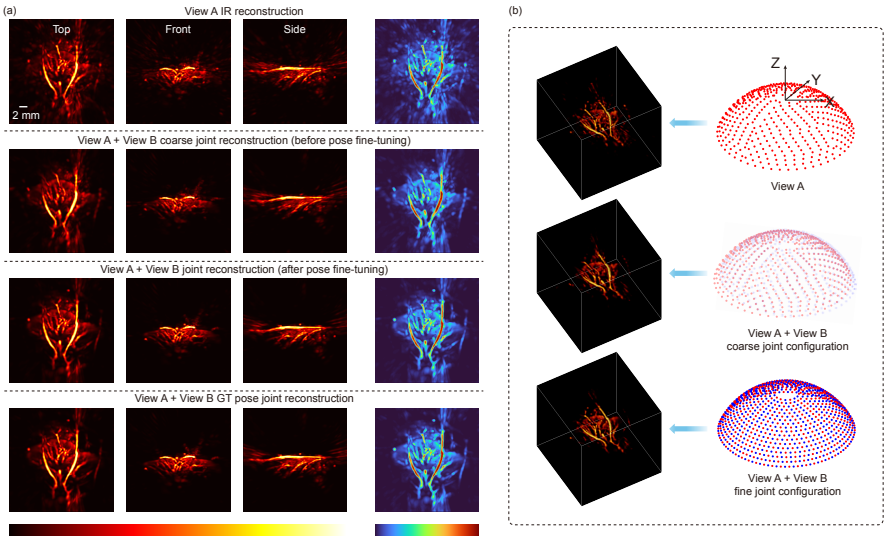

Fig. 5 The reconstruction results of multi-view *in-vivo* rat liver with PA-SfM algorithm. (a) The single view reconstruction, multi-view coarse pose reconstruction, multi-view fine-tuned pose reconstruction and ground truth of the *in-vivo* rat liver (Scale: 2 mm). (b) The 3D results of single view,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

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) and 5(b) present the liver reconstructions. Visually, the single-view liver reconstruction suffers from incomplete vessel connectivity and elevated background noise (clearly observable in the turbo colormap). While the coarse-pose multi-view fusion introduces additional vessel segments into the field of view, the uncalibrated misalignment severely disrupts vascular continuity, causing distinct ghosting artifacts and structural blurring. Conversely, the PA-SfM fine-tuned reconstruction eliminates these misalignment-induced artifacts, seamlessly bridging fragmented vessels and accurately recovering the intricate vascular branching of the liver.

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. Burdened by spatial misalignment, the coarse-pose multi-view approach 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.

4 Discussion

In this work, we propose the PA-SfM framework, a novel tracker-free approach that leverages exclusively single-modality photoacoustic data for simultaneous sensor pose recovery and high-fidelity 3D reconstruction. Traditional handheld 3D photoacoustic tomography systems typically rely on bulky and expensive external tracking hardware, such as optical tracking systems, electromagnetic trackers, or inertial measurement units, to correct manual motion artifacts. However, these conventional methods often suffer from strict line-of-sight requirements, susceptibility to electromagnetic interference, or inherent cumulative drift, which significantly limit their clinical accessibility and flexibility. To overcome these hardware constraints, PA-SfM integrates the forward acoustic wave equation directly into a differentiable programming pipeline. Unlike conventional SfM methods based on visual features or traditional cross-correlation algorithms that fail in the inherently speckle-free photoacoustic domain, our method is built upon a physical understanding of acoustic propagation. By utilizing a high-performance, GPU-accelerated acoustic radiation kernel, the framework simultaneously optimizes the 3D photoacoustic source distribution and the 6-DoF sensor array pose via gradient descent. Furthermore, ensuring robust convergence in unconstrained freehand scenarios poses a significant challenge. To address this, we implemented a comprehensive coarse-to-fine optimization strategy. By incorporating a modified RANSAC algorithm with geometric consistency pre-checks and rigid-body kinematics, the system effectively identifies and eliminates localization outliers. The global differentiable fine-tuning module utilizes the highly reliable signals from an inlier subset to physically guide unreliable sensors into their correct anatomical positions, thus guaranteeing high spatial precision. Our validations, including both numerical simulations and *in-vivo* rat kidney experiments, demonstrate that PA-SfM successfully achieves sub-millimeter positioning accuracy. By coherently fusing multi-view data to augment the synthetic aperture, the framework mitigates limited-view artifacts and restores high-resolution 3D vascular structures that are comparable to ground-truth benchmarks. Ultimately, PA-SfM provides a low-cost, software-defined solution that significantly improves the viability of routine clinical freehand photoacoustic imaging without the need for complex external positioning sensors.

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.

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 and Seongwook Choi 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 [49, 50]:

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

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 [49, 50]:

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

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 $|r - v_s t| \leq r' \leq r + v_s t$. Substituting this back into the pressure expression and simplifying yields:

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

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] 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)

[24] 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)

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

[26] 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)

[27] 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)

[28] 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

[29] 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

[30] 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)

[31] Ren, H., Du, C., Yuan, Z., Park, K., Volkow, N.D., Pan, Y.: Cocaine-induced cortical microischemia in the rodent brain: clinical implications. Molecular psychiatry **17**(10), 1017–1025 (2012)

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

[33] 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)

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

[35] 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)

[36] 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)

[37] Schonberger, J.L., Frahm, J.-M.: Structure-from-motion revisited. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4104–4113 (2016)

[38] Tillet, P., Kung, H.-T., Cox, D.: Triton: an intermediate language and compiler for tiled neural network computations. In: Proceedings of the 3rd ACM SIGPLAN International Workshop on Machine Learning and Programming Languages, pp. 10–19 (2019)

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

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

[41] 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)

[42] 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)

[43] 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)

[44] 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)

[45] 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)

[46] 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)

[47] 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)

[48] 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)

[49] 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)

[50] Xu, M., Wang, L.V.: Photoacoustic imaging in biomedicine. Review of scientific instruments **77**(4) (2006)