This article has been accepted for publication in IEEE Transactions on Medical Imaging. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TMI.2026.3653667

High-Speed Volumetric Dual-Mode Ultrasound and Photoacoustic Tomography with a Single-Element Detector

Jigmi Basumatary, Yousuf Aborahama, Yang Zhang, Yide Zhang, Yushun Zeng, Cindy Z. Liu, Qifa Zhou, Lihong V. Wang

***Abstract*—Acoustic wave detection techniques like ultrasound (US) and photoacoustic (PA) tomography are widely used in biomedical imaging but often require expensive transducer arrays and complex setups for 3D imaging. To overcome these challenges, recent research has explored using an ergodic relay (ER) with a single-element transducer. This approach improves imaging speed at low cost and reduces complexity. This study presents a dual-mode system that utilizes an ER for single-shot, 3D PA and US imaging with a single-element detector. Each ultrasonic or optical excitation pulse generates a 1D signal, which is then used to reconstruct a 3D image in US or PA mode. While the PA mode provides optical absorption-based contrast, the US mode offers complementary acoustic scattering-based contrast. We demonstrate the system's capabilities through *in vivo* imaging of blood vessels and skin structure in human hands. The system is non-invasive, label-free, and ultrafast, enabling 4D imaging with simplified hardware requirements.**

***Index Terms*—Photoacoustic tomography, Ultrasound tomography, Ergodic relay, Single-shot imaging, Single-element detector.**

I. INTRODUCTION

ACOUSTIC wave-based imaging has revolutionized biomedical diagnostics and research by enabling high-resolution, non-invasive visualization of anatomical structures and functional processes. Among the most widely used modalities, ultrasound (US) imaging and photoacoustic (PA) imaging offer unique and complementary advantages, which are often in demand for clinical investigations [1]–[4]. Ultrasound imaging provides detailed structural information based on acoustic wave propagation, while photoacoustic imaging detects optical absorption contrast to capture functional and molecular information. Often dual-mode imaging with both ultrasound and photoacoustics is demanded for advanced biomedical investigation. Photoacoustic computed tomography (PACT) [5], [6] and ultrasound computed tomography techniques are typically implemented using multi-element transducer arrays, enabling three-dimensional (3D) imaging. To capture a full 3D view, it is necessary to either mechanically move or rotate a 1D array, or to use a 2D array that enables beam steering in two directions. This use of arrays has been shown to generate high-resolution 3D images of histology [7], [8] and molecular information [9]–[15]. However, such systems are often complex, costly, and require substantial computational resources. A promising approach leverages compressive sensing and single-pixel imaging, using acoustic scatterers to achieve PA or US computed tomography with a single detector element [16]–[22]. However, these methods are time-consuming, as they require multiple measurements with different mask configurations, which limits their speed. To overcome this limitation, researchers have developed techniques that take advantage of the spatiotemporal encoding properties of an Ergodic Relay (ER) or chaotic cavity [23]–[29]. These methods enable single-shot imaging with fewer detector elements, improving both speed and efficiency. However, the approaches reported so far are limited to 2D imaging and require object-dependent calibration. Recently, imaging using this technique has been demonstrated with object-independent calibration [30] and the capability for 3D imaging [31]. The ER uniquely encodes acoustic waves from each point in the imaging region of interest (ROI), preserving the spatial information and enabling image reconstruction with minimal hardware, making single-shot 3D imaging feasible with just a single-element ultrasound transducer.

In this work, we present a dual-mode acoustic computed tomography system that employs a single ER for single-shot 3D imaging using a single-element ultrasound detector in both PA and US modes. We implemented the proposed ER-based dual-mode system by integrating an unfocused single-element US source for widefield insonification in US mode with our previously reported ER-based PA imaging approach [31]. The proposed dual-mode system required minimal hardware compared to traditional dual-mode systems and supported seamless switching between PA mode and US mode. Similar to PA encoding, in US mode, the scattered or reflected signals from the subject were spatially and temporally encoded in

This work was supported in part by National Institutes of Health (NIH) grants U01 EB029823, R35 CA220436 (Outstanding Investigator Award), R01 EB028277, and R01 CA282505. Yide Z. was sponsored by the NIH grants K99 EB035645. Cindy Z. L. was sponsored by the NIH grant NIGMS T32 GM152342.

Jigmi Basumatary, Yousuf Aborahama, Yang Zhang, Yide Zhang, Cindy Z. Liu, and Lihong V. Wang are with the California Institute of Technology (e-mail : jigmibty@caltech.edu, y.aborahama@caltech.edu, zoengy@caltech.edu, yzhang34@caltech.edu, czliu@caltech.edu and lvw@caltech.edu). Yushun Zeng and Qifa Zhou are with the University of Southern California(e-mail : yushunze@usc.edu and qifazhou@usc.edu).

the ER, which were recorded as 1D temporal signals using a single-element detector. Using a calibration signal, these 1D spatiotemporally encoded signals enabled 3D ultrasound image reconstruction without the need for scanning. The US mode of the proposed system captured single-shot 3D information of cutaneous features, while its PA mode captured single-shot 3D molecular information from the same region of interest (ROI). The system required only a single, object-independent calibration for both modes, enabling comprehensive 3D imaging with reduced hardware complexity. The device encoded 6,400 virtual transducers and captured signals using a single-element transducer, enabling 3D image reconstruction with just a single laser/ultrasound pulse illumination/insonification. In both modes, the system achieved longitudinal volumetric imaging at kilohertz rates, allowing for the capture of fast events. We demonstrated the dual-mode capability of the system by imaging a human hand, using the US mode for acoustic scattering-based contrast imaging and the PA mode for optical absorption-based contrast imaging of blood vessels. This study presents an initial demonstration of an ER based compact dual-mode PA/US imaging system based on a single-element transducer, highlighting the feasibility of such an approach and its potential for further development toward accessible multimodal imaging platforms.

II. METHODS

A. System setup and data acquisition

The ER of the system comprised a right-angle prism with ground edges and a custom optical rod, as shown in Figs. 1(a) and (f). The prism and rod were bonded with ultraviolet-curing adhesive. The single-element ultrasonic transducer (30 MHz, 0.4 × 0.4 mm² aperture), mounted at the bottom of the prism, was fabricated with a PMN-PT piezoelectric single crystal. Details of the fabrication of the ER and transducer can be found in [31]. The fabricated single-element receiving ultrasound transducer has a broadband bandwidth (−6 dB) of approximately 20 MHz, selected to ensure a compact size that minimizes interference with the acoustic scrambling process in the ER while maintaining adequate directivity.

For calibration, a 532 nm pulsed laser with a 5 ns pulse width and 1 kHz repetition rate (INNOSLAB IS8II-DE, EdgeWave) was directed to the ER system, with power controlled using a half-wave plate (WPH10M-532, Thorlabs) and a polarizing beam splitter (PBS25-532-HP, Thorlabs). Transmitted beams were expanded by lenses (ACN254-050-A and AC254-100-A, Thorlabs) and adjusted to 2 mm using an iris (SM1D12, Thorlabs). The adjusted beam was passed through an xy scanner driven by a motorized stage (PLS-85, PI) and then focused on the ER calibration surface using a lens (AC254-300-A, Thorlabs). During calibration mode, the calibration signal was recorded from each point in the FOV by stepwise scanning (using the xy scanner) of the focused beam spot. The signals detected by the UST were amplified using low-noise amplifiers (ZKL-1R5+, Mini-Circuits), filtered with a low-pass filter (BLP-70+, Mini-Circuits), and digitized by a data acquisition card (ATS9350, AlazarTech).

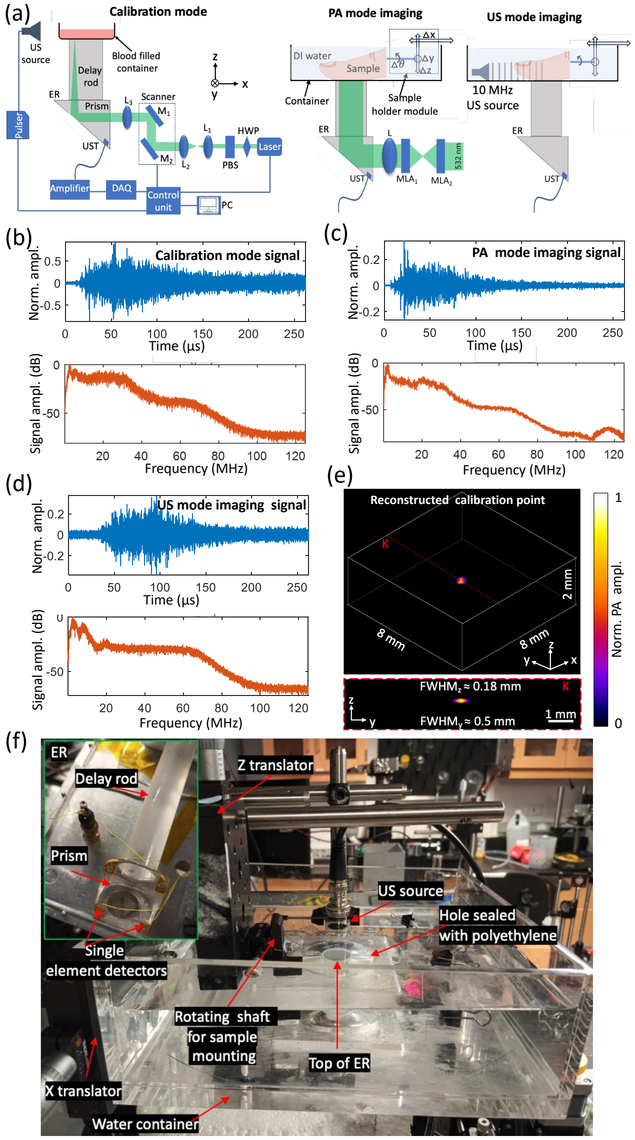


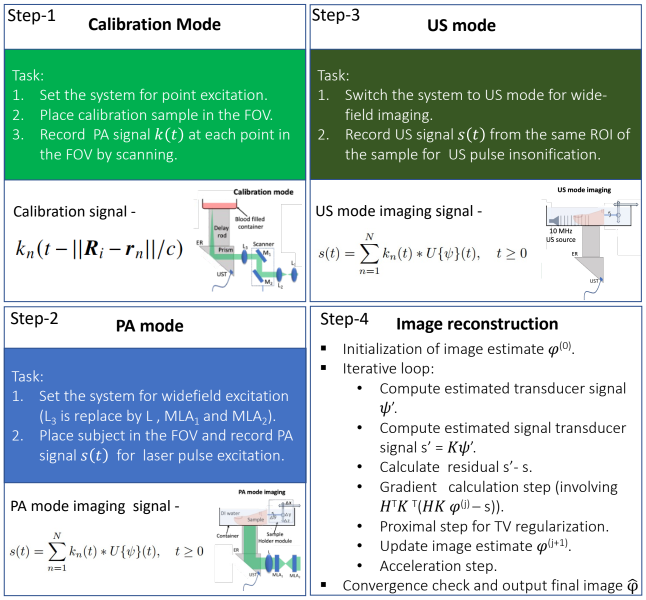
Fig. 1: **Schematic of the system and working modes.** (a) Calibration mode: A tiny focused spot of 532 nm pulse laser light is used to excite the PA signal from absorbing molecules (blood) at the base of imaging FOV and the emitted PA signal is recorded by the UST. The calibration signal at each point in the FOV is obtained by scanning the laser spot. PA mode: In this mode, the entire object in the FOV is illuminated with a homogenized beam of light. As a result, the PA signal emitted from the entire FOV is recorded by the UST. US mode: In this mode, the object in the FOV is insonified by unfocused US wave. The scattered US waves from the object in the FOV, passing through the ER, are recorded by the UST. Typical 1D recorded signal along with its corresponding frequency content in (b) calibration mode, (c) PA mode imaging, and (d) US mode imaging. (e) single-shot 3D volume view reconstructed from a 1D signal, alongside the cross-sectional view of a single calibration point. (f) Photography of imaging chamber. HWP: Half-wave plate; PBS: Polarizing beam splitter, DAQ: Data acquisition, UST: Ultrasound transducer, MLA: Micro lens array.

This article has been accepted for publication in IEEE Transactions on Medical Imaging. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TMI.2026.3653667

A PCIe-6321 device (National Instruments) controlled the laser, motorized stages, and data acquisition. A custom synchronization LabVIEW program was utilized to trigger the pulsed laser, control motorized stages during calibration, and acquire data. Signals were sampled at 250 MHz with 65,532 data points per acquisition. We chose bovine blood for system calibration primarily because it generated a strong, stable, and highly consistent PA signal at our 532 nm laser wavelength over the extended period required for scanning and averaging across all 6,400 virtual transducer points [31]. Bovine blood in a polyethylene-sealed container, placed on top of the ER, served as the optical absorber. US gel (Aquasonic 100, ParkerLabs) was used for acoustic coupling between the ER and polyethylene (thickness $\approx$ 0.1 mm). For calibration, the signal-to-noise ratio was enhanced by averaging 500 acquisitions per calibrated virtual transducer, with data collected only during forward motor movement to avoid backlash.

Post calibration, the system was switched to PA mode imaging. In this mode, the illumination beam diameter was adjusted to 6 mm and homogenized using microlens arrays (64-480, Edmund Optics) and a lens (AC254-250-A, Thorlabs). In PA mode imaging, a 28 μs delay after the laser trigger accounted for the distance from the top of ER to the receiving ultrasonic transducer. During imaging, a laser fluence of $5\ \mathrm{mJ\ cm^{-2}}$ was employed, which was substantially below the American National Standards Institute (ANSI) safety limit of $20\ \mathrm{mJ\ cm^{-2}}$ for exposure at $532\ \mathrm{nm}$. Following the PA mode imaging, the US mode was activated. For US mode imaging, a single-element unfocused US transducer (10 MHz, 10 mm element diameter, HARISONIC) insonified the object orthogonally to the ER delay rod. This 10 MHz center frequency was selected to achieve sufficient penetration depth for superficial tissues with reasonable resolution, and its output frequency lies well within the effective detection bandwidth of the receiver. The transmitting US transducer was pulsed by a pulser (5072PR, PANAMETRICS) with an energy of 140 μJ, a damping of 20 Ω, and a gain of 59 dB. The scattered/reflected US wave from the object propagating through the ER was collected by a fabricated single-element receiving US transducer. In US mode imaging, a delay of 40 μs after the US source trigger accounted for the distance from US source to the receiving ultrasonic transducer.

Imaging for both modalities was conducted in deionized (DI) water within a custom container featuring a central aperture sealed with a polyethylene membrane. The polyethylene layer provided a flexible, waterproof barrier while maintaining sufficient acoustic transparency. A layer of ultrasound gel facilitated acoustic coupling between the polyethylene membrane and the ER. Consequently, the acoustic propagation path comprised sequential layers: the subject, DI water, polyethylene membrane, ultrasound gel, and the ER. This configuration ensured efficient transmission of both PA and US signals to the detector. To maximize acquisition speed, raw single-shot signals were acquired without hardware averaging. Noise suppression was subsequently achieved computationally during the image reconstruction process.


Step-1 **Calibration Mode**

Task:
1. Set the system for point excitation.
2. Place calibration sample in the FOV.
3. Record PA signal $k(t)$ at each point in the FOV by scanning.

Calibration signal -

$$k_n(t - ||\boldsymbol{R}_i - \boldsymbol{r}_n||/c)$$

Step-3 **US mode**

Task:
1. Switch the system to US mode for wide-field imaging.
2. Record US signal $s(t)$ from the same ROI of the sample for US pulse insonification.

US mode imaging signal -

$$s(t) = \sum_{n=1}^{N} k_n(t) * U\{\psi\}(t), \quad t \geq 0$$

Step-2 **PA mode**

Task:
1. Set the system for widefield excitation (L_3 is replace by L , MLA_1 and MLA_2).
2. Place subject in the FOV and record PA signal $s(t)$ for laser pulse excitation.

PA mode imaging signal -

$$s(t) = \sum_{n=1}^{N} k_n(t) * U\{\psi\}(t), \quad t \geq 0$$

Step-4 **Image reconstruction**

- Initialization of image estimate $\varphi^{(0)}$.
- Iterative loop:
  - Compute estimated transducer signal ψ'.
  - Compute estimated signal transducer signal s' = $K\psi'$.
  - Calculate residual s'- s.
  - Gradient calculation step (involving $H^{T}K^{T}(HK\varphi^{(j)} - s)$).
  - Proximal step for TV regularization.
  - Update image estimate $\varphi^{(j+1)}$.
  - Acceleration step.
- Convergence check and output final image $\hat{\varphi}$

Fig. 2: **System workflow.** Key steps of the proposed dual-mode system operation and the corresponding image reconstruction process.

To position the specimen within the imaging FOV, we employed a custom sample holder assembled from modular optomechanical components. The holder featured multiple degrees of freedom, allowing for precise linear translation and rotational adjustment. It was mounted horizontally within the imaging chamber to align the sample orientation with the system's imaging geometry. A photograph of the complete imaging system, highlighting these key components, is presented in Fig. 1(f).

B. System schematic and working modes

We present a high-speed dual-mode acoustic tomography system that uses an ER with a single-element detector for low-cost, high-throughput, snapshot wide-field imaging. The ER encodes spatial and temporal information of the object by employing randomized temporal signatures through ergodicity. The system requires only a single-element ultrasonic transducer to capture 3D wide-field information from a single-pulse wide-field illumination or insonification. We reconstructed 3D images in both PA and US modes from 1D signals recorded with a single transducer, using a one-time universal calibration. An overview of the system workflow is presented in Fig. 2. The schematic of the system during calibration is shown in Fig. 1(a). Calibration signals at each point in the field of view (FOV) were generated by exciting PA signals from blood molecules using a focused laser spot. After calibration, the system was switched to PA imaging mode and subsequently to US imaging mode. In PA imaging mode, the FOV was illuminated with a homogenized light beam, while in US imaging mode, the FOV was insonified with an unfocused ultrasound wave orthogonal to the ER. Figs. 1(b), (c), and (d) show the single-shot 1D time-domain signals along with their frequency content, while Fig. 1(e) presents the corresponding 3D reconstructed images for a point-PA-excited calibration.

C. Sample preparation and imaging

For calibration, refrigerated lysed bovine blood (910, Quad Five) was restored to room temperature and transferred to a customized container (Figs. 1(a) and 3(a)) containing a polyethylene layer at the bottom to acquire calibration signals for image reconstruction. The blood-containing container was mounted on top of the ER, and ultrasound gel was applied between the polyethylene layer and the ER to ensure acoustic coupling.

For Phantom 1–5, 7, and 8 experiments, we used black wire (8251T9, McMaster-Carr), which consisted of multiple strands enclosed in a rubber coating. The rubber insulation was removed, and a single strand ($\approx$ 0.1 mm diameter) was selected to construct the desired phantoms. To enhance the PA signal during PA mode imaging, all wire phantoms were painted with black ink. The letters in Phantom 1 (Fig. 4(a)) were constructed from single-strand wire segments, which were cut and glued using Fevistik to form the desired letter shapes. The point Phantom 2 (Fig. 4(c)) was made of a single strand of black wire. The strand was cut into approximately 0.4 mm segments and attached to a human hair using Fevistick adhesive. We observed that US scattering from the human hair was negligible compared to the scattering from the point Phantom 2. During US imaging, the point Phantom 2 was freely moved by hand using the attached hair. Phantom 3 (Fig. 5(a)) was composed of two wire strands glued together to form a cross. Similarly, Phantom 4 (Fig. 5(b)) was composed of two wire strands glued together to form a cross and painted with black ink to enhance the photoacoustic (PA) signal. Phantom 5 was a single straight wire coated with black ink and served as the reference object in the co-registration experiment (Fig. 5(d)). Phantom 7 (Fig. 11) was made of two straight, single-strand wires glued together using Fevistick to form a cross. Line Phantom 8 (Movie 5) was made of a straight, single-strand wire that was moved freely by hand within the field of view (FOV) during the experiment.

Phantom 6 was designed for the dual-mode black ink flushing experiment (Fig. 5(e)). A tube (Braintree Scientific, 0.3 mm inner diameter) was positioned in the FOV, and scattered US signals (using US mode imaging, Fig. 1(a)) from the tube were recorded. Subsequently, a syringe pump (NE-300, New Era) was used to flush black ink (X-1, Tamiya) through the tube, and PA signals (using PA mode imaging, Fig. 1(a)) were recorded. The customized sample holder allowed XYZ translational movement and rotation, enabling the phantom to be tilted at the desired angle. For the experiment, these phantoms were mounted on the customized sample holder and positioned at the desired orientation within the imaging FOV, which was filled with DI water.

Human imaging experiments were conducted in accordance with relevant guidelines and regulations. The experiments followed a protocol approved by the Institutional Review Board (IRB) of the California Institute of Technology. After calibration, participants were enrolled solely for the imaging procedure. To image the vasculature and cutaneous features of the human hand, participants were instructed to place their hands on top of the ER, which was immersed in DI water. To image the hemodynamics of the human hand in response to cuffing, a sphygmomanometer was wrapped around the participant's upper arm (Fig. 6(a)). Blood vessel occlusion was induced by inflating the sphygmomanometer to a high pressure (200 mmHg), maintained for a short period (approximately 20 seconds), and then quickly released. Subsequently, the system was switched to US mode imaging, and data were collected. The total imaging time was approximately 60 seconds. All dual-mode imaging experiments, both on phantoms and *in vivo*, started with PA mode imaging (Fig. 1(a)), followed by US mode imaging.

D. Image Reconstruction

In this work, the signal $s(t)$ detected by the ultrasonic transducer at time t (in both PA and US modes) in a homogeneous medium is expressed as:

$$s(t) = \sum_{n=1}^{N} k_n(t) * U\{\psi\}(t), \quad t \geq 0, \tag{1}$$

where N is the number of virtual transducers, $k_n(t)$ is the normalized impulse response extracted from calibration at the n-th virtual transducer through integration, $*$ is a 1D convolution, and $U\{\psi\}(t)$ is the normal derivative of the pressure field (at the relay surface) generated by an initial pressure ψ in the PA mode or due to the scattering of the US source from the target ψ in the US mode. The scattered field is treated here as independent sources that are excited at specific time instances based on the duration it takes the excitation wave to reach each point. The images are reconstructed by solving the regularized least squares problem:

$$\hat{\varphi} = \arg\min_{\varphi \in V} \|KH\varphi - s\|^2 + \lambda |\varphi|_{\mathrm{TV}}, \tag{2}$$

where $H : \mathbb{R}^{N_s} \to \mathbb{R}^{N \times N_t}$ is a discrete approximation for the operator U described in Eq. 1 that maps the discretized vesion of ψ which is denoted here as φ (acoustic sources at N_s points) to N intermediate signals (at the virtual transducers) of length N_t, $K : \mathbb{R}^{N \times N_t} \to \mathbb{R}^{N_T}$ is the convolution operator that maps the intermediate signal at the virtual transducers to the output signal from the transducer consisting of N_T time samples, s is the data measured by the transducer, V is the set of admissible images which is chosen to be the set of N-dimensional real vectors for the US mode and the positive real numbers for the PA mode, and $\lambda |\varphi|_{\mathrm{TV}}$ is a total variation (TV) regularization term. Using TV regularization allows the iterative reconstruction process to leverage the inherent piecewise smoothness of biological tissues, significantly enhancing stability and accuracy.

We solve the above optimization problem using an accelerated proximal gradient method [36]–[38], where the step size

This article has been accepted for publication in IEEE Transactions on Medical Imaging. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TMI.2026.3653667

can be estimated using power iterations [39]. The actions of the operator H and its adjoint (used to compute the gradient) are evaluated using the pseudo-spectral time-domain-based k-Wave toolbox [40]. The optimization process was terminated after 10 iterations. The effect of iteration count and TV regularization, as well as the key steps of the image reconstruction process, are presented in Figs. 11 and 2, respectively. Post-reconstruction, the volume stack is background-subtracted using the rolling ball algorithm [35], followed by median filtering to localize the features. The background subtraction step is not applied to the US images. The data processing and image reconstruction were performed using MATLAB R2024b on an Ubuntu 24.04.1 LTS system equipped with dual Intel Xeon 6248R CPUs (24 cores per CPU, 3.0 GHz base, 4.0 GHz Turbo), 12 × 128 GB DDR4-2933 ECC memory, and dual NVIDIA A100 GPUs (PCI-E 4.0, 40GB HBM2 memory each).

III. RESULTS

A. Calibration signal

The calibration signals (Figs. 3(a) and (b)) generated at the base of the imaging volume, denoted as $(k_1, k_2, \ldots, k_n)$, where n is the total number of virtual transducers, are object-independent. These signals are determined by the geometry of the ER, its acoustic properties, the transducer's properties, and the laser-induced excitation parameters. As a result, a single universal calibration can be used for 3D imaging, with calibration data remaining effective for more than one year [31]. Due to the delay line attached to the prism, the calibration signals generated at the base of the imaging volume are object-independent. Any acoustic signal passing through any virtual transducer follows identical acoustic paths within the ER, meaning that the same calibration signals can be used for both PA and US mode image reconstruction. The focused laser beam was scanned in an 80 × 80 grid with 0.10 mm steps, resulting in 6,400 virtual transducers in the imaging FOV. To ensure distinct calibration signals, the step size ($\sim$ 0.10 mm) was set to half of the full width at half maximum (FWHM) of the cross-correlation profile (Fig. 3(b)). While the system is calibrated for 6,400 virtual transducer positions, the effective number of independent detection channels realized in practice is influenced by the bandwidth of the acoustic signals generated in PA or US mode. A narrower signal bandwidth may lead to increased correlation between the responses of closely spaced virtual transducers, thereby reducing the degrees of freedom in the measurement below the nominal 6,400. Nevertheless, the current configuration provides sufficient independent information for 3D volumetric reconstruction, as demonstrated.

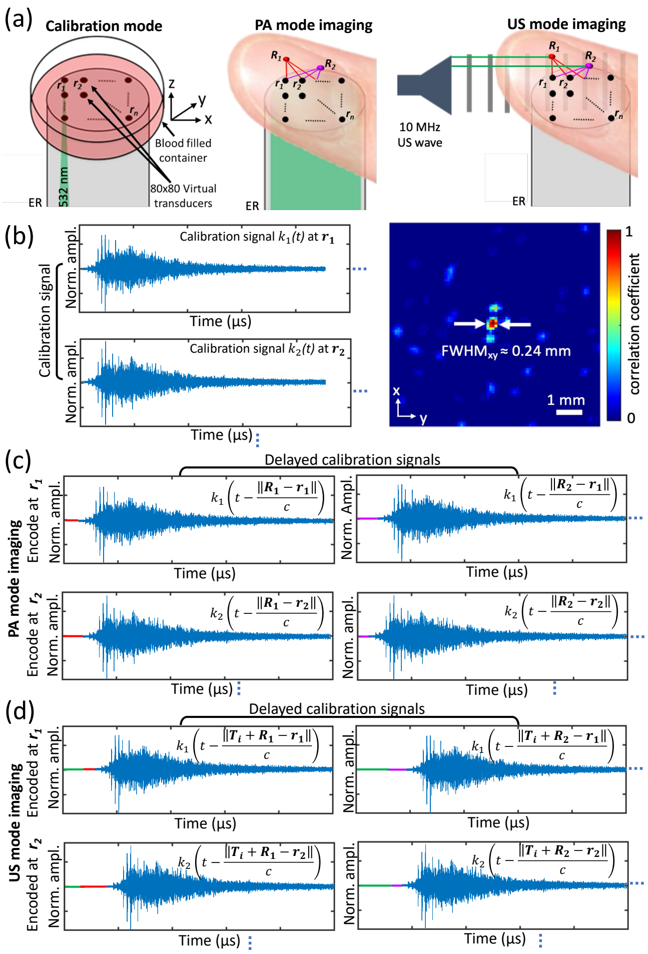


Fig. 3: **Spatial and temporal illustrations of signals**. (a) In calibration mode, signals k_1, k_2 $\ldots$ k_2 at the virtual transducer positions $\boldsymbol{r_1}$, $\boldsymbol{r_2}$ $\ldots$ $\boldsymbol{r_n}$ are obtained using blood. In PA mode, homogenized broad beam shown in green provides wide-field illumination. $\boldsymbol{R_1}$ and $\boldsymbol{R_2}$ indicated by red and yellow spheres represent two optical absorbers that emit PA signals in the object. In US mode, US waves illuminate the FOV orthogonally, and $\boldsymbol{R_1}$ and $\boldsymbol{R_2}$ represent the positions of two acoustic scattering points in the object, indicated by red and magenta spheres, respectively. (b) In calibration mode, typical 1D(t) calibration signals k_1 and k_2 detected by the ultrasonic transducer at virtual transducer positions $\boldsymbol{r_1}$ and $\boldsymbol{r_2}$. The cross-correlation map of the calibration signals shows a FWHM of $\sim$ 0.24 mm. (c) In PA mode, the PA signals from points $\boldsymbol{R_1}$ (red lines) and R_2 (magenta lines) reach the virtual transducers (e.g., $\boldsymbol{r_1}$ and $\boldsymbol{r_2}$) with different delays. PA signals from $\boldsymbol{R_1}$ and $\boldsymbol{R_2}$, which are proportional to k_1 and k_2 with delays based on the distances between the calibrated virtual transducers and the absorbers. (d) In ultrasound mode, the US signals from points $\boldsymbol{R_1}$ (red lines) and $\boldsymbol{R_2}$ (magenta lines) reach the virtual transducers (e.g., $\boldsymbol{r_1}$ and $\boldsymbol{r_2}$) with different delays. US signals from $\boldsymbol{R_1}$ and $\boldsymbol{R_2}$, which are proportional to k_1 and k_2 with delays based on the distances between the calibrated virtual transducers and the source (green line) via the scattering points.

B. PA mode imaging signal

Spatial and temporal encoding of PA mode imaging signals is illustrated through Figs. 3(a) and (c). When source points in the 3D volume with positions $\boldsymbol{R_i}$, $i = 1, 2, ...m$ are wide-field illuminated by a laser pulse from the bottom of ER as shown in Fig. 3(a), the PA signals generated would propagate to the calibrated virtual transducers $\boldsymbol{r_n}$, $n = 1, 2, ..n$ after time $t_{i,n} = ||\boldsymbol{R_i} - \boldsymbol{r_n}||/c$, where c is the speed of sound in the medium. Then, these signals would follow the same acoustic path inside the ER to the ultrasonic transducer as that of the calibration signals. From the perspective of the transducer, compared with the calibration signal $k_n(t)$ acquired at $\boldsymbol{r}_n$ (Fig. 3(b)), the signal from the source point $\boldsymbol{R}_i$ relayed through $\boldsymbol{r}_n$ is proportional to $k_n(t)$ delayed by $t_{i,n}$, that is, $k_n(t - ||\boldsymbol{R}_i - \boldsymbol{r}_n||/c)$ (Fig. 3(c)). The signal is modulated by both the initial pressure $P_{0,i}$ at $\boldsymbol{R}_i$, and a weighting factor dependent on the angle and distance. Therefore, the recorded wide-field imaging signal is essentially the superposition of delayed calibration signals from all absorbers in the illuminated field of view. Accordingly, we developed an algorithm to reconstruct the initial pressure in the 3D volume (Methods section).

C. US mode imaging signal

Spatial and temporal encoding of US mode imaging signals is illustrated through Figs. 3(a) and (d). When acoustically scattering points in a 3D volume with positions $\boldsymbol{R}_i$, where $i = 1, 2, \ldots, M$ are wide-field insonified by a pulse of US wave orthogonal to the delay rod (Fig. 3(a)), the scattered US wave from $\boldsymbol{R}_i$ would propagate to the calibrated virtual transducers at positions $\boldsymbol{r}_n$, where $n = 1, 2, \ldots, N$ after time $t_{i,n} = ||\boldsymbol{T}_i + \boldsymbol{R}_i - \boldsymbol{r}_n||/c$, where c is the speed of sound in the medium and $\boldsymbol{T}_i$ is the distance between US source and $\boldsymbol{R}_i$ scattering point. Then, these scattered signals would follow the same acoustic path inside the ER to the ultrasonic transducer as that of the calibration signals. From the perspective of the transducer, compared with the calibration signal $k_n(t)$ acquired at $\boldsymbol{r}_n$ (Fig. 3(b)), the signal from the scattering point $\boldsymbol{R}_i$ relayed through r_n is proportional to $k_n(t)$ delayed by $t_{i,n}$, that is, $k_n(||\boldsymbol{T}_i + \boldsymbol{R}_i - \boldsymbol{r}_n||/c)$ (Fig. 3(d)). The signal is modulated by both scattered US pressure $U_i(t)$ at R_i, and a weighting factor dependent on the angle and distance. Therefore, the recorded wide-field imaging signal is essentially the superposition of delayed calibration signals from all scattering points in the insonified field of view. Accordingly, we developed an algorithm to reconstruct the initial pressure in the 3D volume (Methods section).

D. In vitro characterization with phantoms

To validate that the 10 MHz US source insonifies the imaging FOV, we imaged a moving line Phantom 8 positioned at a height of 5 mm. The motion of the line phantom was visible in the reconstructed images (Movie 5 and Appendix B), confirming coverage of the 8×8 mm FOV. Furthermore, we were able to image the cross phantom up to a height of 10 mm (Figs. 5(a)-(c)), which confirms that the US source insonifies at least an $8 \times 8 \times 8$ mm FOV. Next, to ensure that the 3D volumes reconstructed in US mode accurately represent the objects imaged, we conducted a few controlled experiments by imaging several phantoms with known ground truths. Fig. 4(a) shows maximum amplitude projections (MAPs) of the 3D reconstruction of Phantom 1, which consists of the letters “CALTECH” made of black wire. The images were averaged over 100 single-shot acquisitions to maximize the signal-to-noise ratio (SNR). Note that the letters were positioned at approximately a 45-degree incline

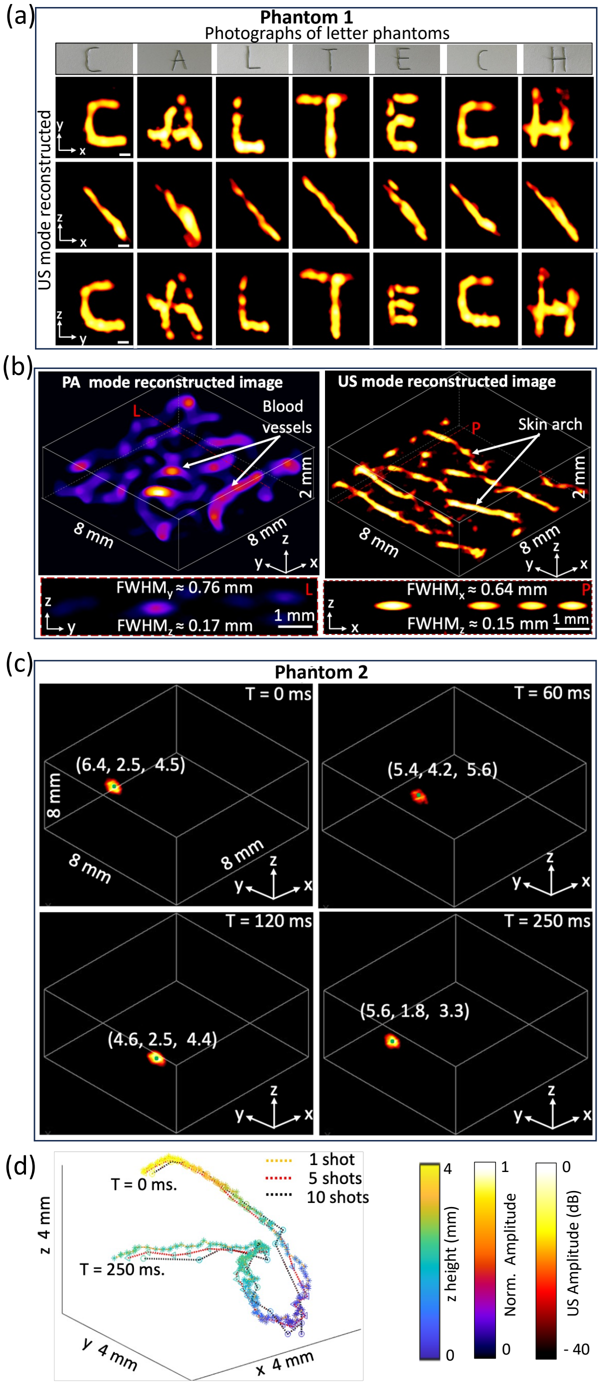


Fig. 4: **US mode 3D spatialtemporal imaging.** (a) Phantom 1: Orthogonal maximum amplitude projection images of letters made of a wire, positioned at a 45-degree angle relative to the US source insonification, along with a photograph of the phantoms. (b) Single-shot 3D volume view reconstructed from a 1D signal, alongside the cross-sectional view of the vasculature in the fingertip using PA mode, and the cutaneous features on the fingertip using US mode. (c) Phantom 2: Single-shot 3D images of an acoustically scattering point phantom moving in the FOV acquired at time $T = 0$ ms, 60 ms, 120 ms, and 250 ms (d) 3D movement path of the point phantom reconstructed with varying shots of measurement.

This article has been accepted for publication in IEEE Transactions on Medical Imaging. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TMI.2026.3653667

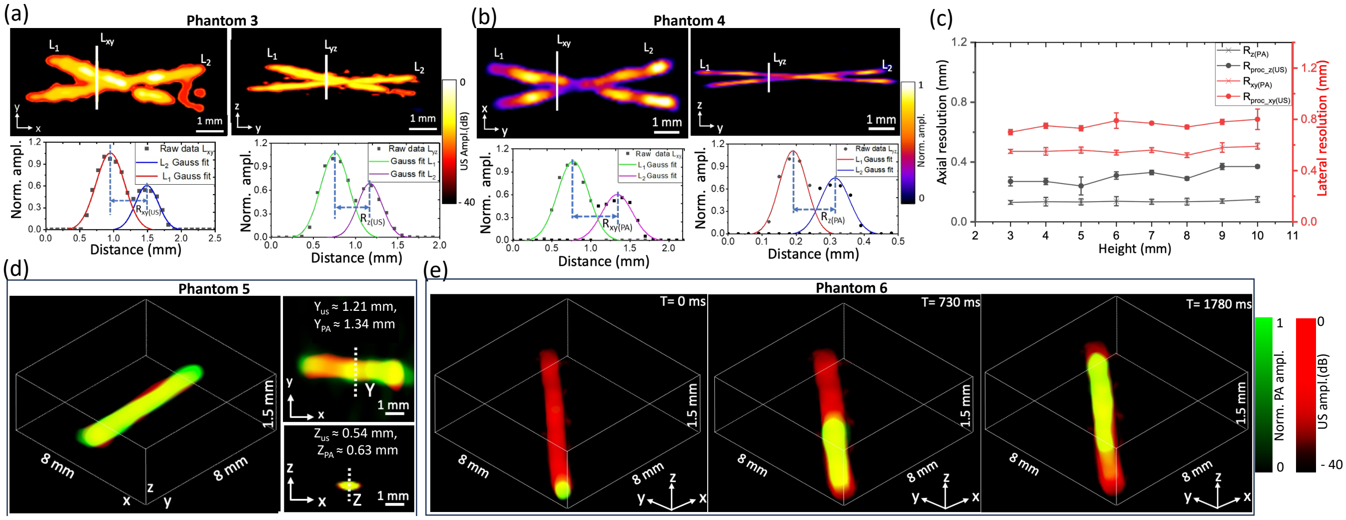


Fig. 5: **Dual-mode resolution and co-registration.** (a) Phantom 3: US mode image of two crossing wires in the xy plane and the yz plane. The profiles along L_{xy} and L_{yz} are used to estimate the lateral $R_{xy(\text{US})}$ and axial $R_{z(\text{US})}$ resolutions. (b) Phantom 4: PA mode image of two crossing wires in the xy plane and the yz plane. The profiles along L_{xy} and L_{yz} are used to quantify the lateral $R_{xy(\text{PA})}$ and axial $R_{z(\text{PA})}$ resolutions. (c) Post-processed US mode and PA mode resolutions at varying imaging heights in the FOV. (d) Phantom 5: 3D reconstructed co-registered PA and US images of the same straight wire along with the cross-sectional views. (e) Phantom 6: Co-registered dual-mode single-shot 3D reconstructed images of an acoustically scattering tube (green color) and flowing black ink (red color) acquired at varying time points. Yellow indicates the overlap of red and green.

relative to the transmitting US source. A complete 3D view of the reconstructed phantoms can be found in Movie 4. Subsequently, single-shot imaging was performed on a real subject using PA mode to visualize blood vessels and US mode to image the cutaneous fingertip (Fig. 4(b)). To assess the reliability of the system, additional imaging was conducted on different parts of the human body (knuckle and wrist) as well as on an adult zebrafish, as presented in Fig. 10. Additionally, to investigate the effect of signal averaging on the apparent position of a moving particle, and to illustrate the trade-off between enhanced image quality and reduced temporal fidelity during dynamic tracking, we performed imaging of a freely moving point Phantom 2. Fig. 4(c) shows the single-shot 3D reconstructed images along with the position of a freely moving point Phantom 2 in the FOV at a 1 kHz data acquisition rate, at time points $T = 0$ ms, 60 ms, 120 ms, and 250 ms. The full movement path of the phantom in the FOV, with single-shot, 5-shot averaging, and 10-shot averaging is depicted in Fig. 4(d). More details of the moving trajectory can be found in Movie 1 and Appendix B.

E. Imaged resolution

To quantify the resolution of US-mode imaging, Phantom 3, consisting of crossed wires, was imaged and reconstructed from 100 single-shot acquisitions (Fig. 5(a)). The wires were intentionally positioned so that the cross pattern was visible in both the z- and y-projections. Defining the spatial resolution as the minimum distance that can distinguish the peaks of the two wires, we found the lateral ($R_{xy(\text{US})}$) and axial ($R_{z(\text{US})}$) resolutions of single-shot US mode imaging to be 0.60 mm and 0.38 mm, respectively (Fig. 5(a)). As an alternative, we assessed the resolutions based on the contrast-to-noise ratio (CNR) (Fig. 7(c)), and the resolution values were consistent in both cases. However, due to orthogonal insonification, the reconstructed PSF in the xz-plane is inclined relative to the US source (Fig. 8(a)). To visualize and co-align with the PA PSF (Fig. 8(c)), we rotated the US reconstructed image by 45 degrees (inclined angle). After rotation, lateral resolution $R_{xy(\text{US})}$ scales up, and the axial resolution $R_{z(\text{US})}$ scales down by $\cos\theta$, where θ is the inclined angle (Figs. 8(d) and (b)). We measured the process resolutions $R_{\text{proc\_}xy((\text{US}))}$ and $R_{\text{proc\_}z(\text{US})}$, which are rotated versions of the raw resolutions $R_{xy(\text{US})}$ and $R_{z(\text{US})}$ at multiple z-positions in the FOV, to investigate variation along the imaging height. The results show a maximum error of 6.25% in resolution within an imaging height of 10 mm, as shown in Fig. 5(c). These experimentally calculated resolutions were found to be very close to the theoretical resolution values. More details on the theoretical resolution calculation can be found in Appendix A and Fig. 7(a).

Similarly, to evaluate the resolution in PA mode, Phantom 4, consisting of black ink-coated crossed wires, was imaged using the same method as that applied in US mode and reconstructed from 100 single-shot acquisitions (Fig. 5(b)). We found the lateral $R_{xy(\text{PA})}$ and axial $R_{z(\text{PA})}$ resolutions in PA mode to be 0.56 mm and 0.13 mm, respectively, with consistent resolution observed throughout the imaging FOV (Figs. 5(c), 7(b), 7(d), and Appendix A). Next, Phantom 5 was imaged to co-register the independently reconstructed US and PA images. Fig. 5(d) shows both the volume and cross-sectional views of

This article has been accepted for publication in IEEE Transactions on Medical Imaging. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TMI.2026.3653667

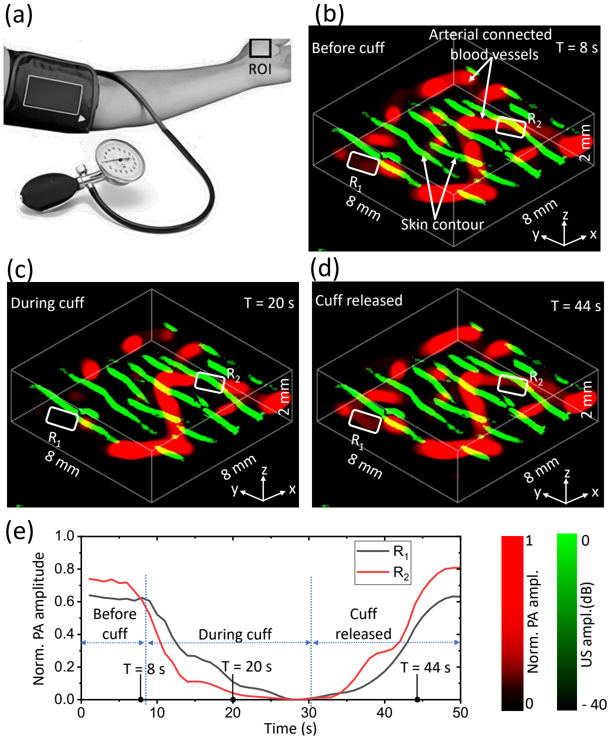


Fig. 6: **Dual-mode *in vivo* imaging of the human hand: participant-1.** (a) Photograph of the arm and hand illustrating the *in vivo* experiment. To vary blood flow in the thenar region of the hand, a sphygmomanometer cuff is applied to the arm. Images are captured before inflation, during inflation, and after the cuff release. Dual-mode images acquired (b) before the initiation of cuffing, (c) during inflation, and (d) after cuff release. Green color-coded structures represent the US images of cutaneous features in the thenar region. In contrast, red color-coded structures represent the PA images of blood vessels in the same region. (e) Mean PA amplitude variations observed during the experiment at regions R_1 and R_2.

the co-registered US and PA images of Phantom 5 at the same position, reconstructed from 100 single-shot acquisitions. For co-registration and alignment with the PA mode, we performed a rotational transformation and minor translational adjustments (via cross-correlation) on the US mode images. After co-registration, the Target Registration Error (TRE) was computed as the Euclidean distance between corresponding identifiable points ($P_{\mathrm{PA}} = (x_{\mathrm{PA}}, y_{\mathrm{PA}}, z_{\mathrm{PA}})$ in the PA image and $P_{\mathrm{US}} = (x_{\mathrm{US}}, y_{\mathrm{US}}, z_{\mathrm{US}})$ in the US image) in the co-registered volumes: $\mathrm{TRE} = \sqrt{(x_{\mathrm{PA}} - x_{\mathrm{US}})^2 + (y_{\mathrm{PA}} - y_{\mathrm{US}})^2 + (z_{\mathrm{PA}} - z_{\mathrm{US}})^2}$. Three identifiable points on the line phantom yielded a mean TRE of 0.15 mm with a standard deviation of 0.080 mm. The optimal co-registration parameters were then used for further data processing. Subsequently, we imaged the flushing of black ink through a tube (Phantom 6) at a speed of 1000 μL/hr. The tube structure was reconstructed from scattered signals in US mode, while the black ink flow was captured in PA mode. Fig. 5(e) shows the dual-mode, single-shot image reconstruction at time points $T = 0$ ms, $T = 730$ ms, and $T = 1780$ ms. Full blood flushing can be seen in Movie 2 and Appendix B.

F. Dual-mode imaging of structure and haemodynamics in human hands

To demonstrate the effectiveness of the system in dual-mode imaging, we monitored the hemodynamics of the thenar region of the hand using PA mode and used US mode to visualize cutaneous features in the same ROI. In this experiment, the US images served as a stable anatomical reference for the dynamic PA data. In PA mode, our focus was on imaging the thenar vasculature and its response to cuffing, induced by a sphygmomanometer wrapped around participant-1's upper arm (Fig. 6(a)). We then reconstructed the 4D dynamics of blood flow through the thenar vasculature in response to cuffing, along with 3D cutaneous features using US mode. The co-registered dual-mode 3D images (Figs. 6(b)-(d)) revealed that blood vessels exhibited a decrease in PA amplitude following cuffing due to blood flow occlusion. When the cuff was released, blood flow resumed, and the PA amplitude rapidly recovered (Fig. 6(c). Additional information and participant-2's images can be found in Figs. 9(a)-(b), Movie 3, and Appendix B. All *in vivo* images were reconstructed from 50 single-shot averaged signals. By enabling dual-mode imaging that provides spatially resolved maps of blood distribution (via PA mode) co-registered with cutaneous structural features (via US mode), the system offers comprehensive anatomical and physiological co-localization information. This spatially resolved data complements traditional non-imaging physiological monitoring techniques, such as pulse oximetry, which provide systemic or single-point measurements without depth or lateral detail.

IV. CONCLUSION

The proposed system is a non-invasive, label-free, low-cost, and ultrafast imaging technique that enables 4D dual-mode imaging using the 1D signal captured by a single detector. This approach achieves an imaging speed of up to a thousand volumes per second. We have demonstrated the ability of the system to visualize dual-mode 4D hemodynamic responses and cutaneous structural information in the thenar region of the human hand. Using a single universal object-independent calibration, we were able to reconstruct both PA and US contrast images of target objects. PA mode imaging provides detailed blood molecule information, while US mode offers cutaneous features, such as shape, size, and position. With minimal hardware complexity, we have demonstrated a system that combines PA and US modalities in a single platform, harnessing the complementary strengths of both techniques to offer a comprehensive understanding of the co-localization of molecular and structural information. Understanding the co-localization of cutaneous features and blood vessels is essential for skin physiology and has significant clinical relevance in dermatology, vascular medicine, and plastic surgery [32]–[34]. We believe that this system may have a significant impact on diagnosing various pathological conditions. By addressing the current challenges of the proposed system, we expect it to attract broader interest for use in various clinical applications. One of the key challenges is penetration depth. In PA mode, *in vivo* penetration depth is limited to approximately 3.6 mm due to strong attenuation of

This article has been accepted for publication in IEEE Transactions on Medical Imaging. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TMI.2026.3653667

the 532 nm light by endogenous chromophores in biological tissue. In US mode, to minimize impedance mismatch and enable direct transmission of the 10 MHz US source signal, we adopted an orthogonal insonification mode. Due to the limited depth of penetration and high scattering of US, we found that positioning the target at a 45-degree angle in the FOV relative to the US source achieves optimal insonification and signal detection. In this orthogonal scheme, penetration is limited to approximately 1.5 mm. Another limitation of the system is its small 8 mm × 8 mm FOV, which is constrained by the diameter of the delay glass rod. The dimensions of the prism and delay rod are designed so that the diameter of the rod is significantly smaller than the right-angle edge of the prism to ensure effective scrambling of the acoustic signals. The length of the delay rod is much greater than the edge length of the prism, which extends the duration of the object-independent signal to ensure universal calibration. Finally, the spatial resolution of the system is currently constrained by the limited-view effect caused by the acoustic impedance mismatch between the object and the imaging system. Additionally, the SNR in US mode could be improved by using appropriate wavelengths and US sources with higher energy, which would enhance the imaging of denser and deeper structures. Furthermore, the current reconstruction algorithm involves the use of regularization, which can result in the loss of histological information in US mode.

For US mode imaging, to minimize interference orthogonal configuration was selected to facilitate integration of the US source with the ER based PA imaging setup. Introducing the US source along the ER's detection axis (e.g., from the bottom or through the sample) would cause acoustic interference with the sensitive single-element receiving detector, strong reflections from ER interfaces, and physical obstruction of the PA laser path or sample access. Orthogonal insonification enabled simpler acoustic coupling without altering the PA detection geometry. This geometry favors detection of side- or forward-scattered (relative to the incident ultrasound beam but still entering the ER) rather than backscattered ultrasound, making the contrast more sensitive to rough surfaces, complex textures, and small heterogeneities. Specular reflectors perpendicular to the beam may yield weaker signals unless they scatter significantly at 90 degrees. Despite being unconventional, this mode offers co-registered structural information alongside PA data, as demonstrated in Fig. 6, revealing cutaneous and subcutaneous features. It is especially useful for superficial imaging applications where multimodal information is prioritized over depth, and aligns with approaches used in scattering-based US imaging. In the reconstructed US images (Fig. 6 and 9), features that are parallel (horizontal features) to the US source appear more prominent due to the directional insonification by an unfocused transducer, the anisotropic scattering properties of biological tissues, and the angular sensitivity of the imaging system. To validate this effect, we performed rotation experiments with phantoms having known linear features, which confirmed that these patterns rotate consistently with the sample orientation. This demonstrates that the patterns reflect the physical alignment of structures relative to the fixed US source and detector geometry, rather than being arbitrary reconstruction artifacts.

With the current hardware configuration setup, in the US mode for *in vivo* human hand imaging (as shown in Figs. 4(b), 6, 9, and 10), several identifiable skin-related structures and layers are observed, visualized based on their acoustic scattering properties. We observed the size, shape, and contour, as well as the elevation/thickness of cutaneous features on the fingertip (Fig. 4(b)) and palm of the hand (Figs. 6 and 9). We also observed the flexion creases on the knuckle, the epidermal layer and its thickness on the wrist of the human hand, and the skeleton of an adult zebrafish, as presented in Fig. 10. While we did not perform concurrent histological validation in this study to precisely correlate every visualized feature with specific microscopic skin layers, the observed structures and their appearance are consistent with those typically identified in cutaneous ultrasonography. In the current implementation, the US mode provides basic structural contrast that aids in localizing the region where PA signals arise. While detailed anatomical layering of the skin is not clearly resolved, the complementary US information helps to contextualize PA signal distribution.

To mitigate the limitations associated with increasing the FOV while preserving ergodicity in ER-based systems, several strategies can be considered. First, the development of advanced ER designs—including the exploration of novel materials with more favorable acoustic properties or alternative geometric configurations such as multi-segment ERs or engineered scattering structures—may enable larger effective areas without compromising ergodicity. Second, optimized calibration schemes, such as multi-spot laser calibration or compressive calibration techniques (as discussed in prior internal evaluations), could significantly reduce the data acquisition and time requirements associated with calibrating larger ERs. Third, computational strategies that incorporate more efficient image reconstruction algorithms, enhanced GPU-based parallel processing, or deep learning-based methods could alleviate the increased computational load imposed by larger FOVs. Lastly, a modular approach involving the integration of multiple smaller ER modules presents another possible avenue, though this would inherently introduce greater system complexity. Together, these strategies offer a pathway toward scalable, high-performance ER imaging systems with extended FOV capabilities.

To enhance imaging depth in both PA and US modes, several strategies can be implemented. In PA mode, wavelength optimization is a key approach, where the use of longer near-infrared (NIR) wavelengths (typically in the 700–900 nm range) helps minimize optical absorption by water and reduces scattering, thereby enabling deeper light penetration into tissue. Additionally, increasing the laser fluence, while staying within established safety limits, can improve signal strength from deeper regions. Advanced light delivery techniques—such as wavefront shaping or employing multiple

This article has been accepted for publication in IEEE Transactions on Medical Imaging. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/TMI.2026.3653667

illumination angles—can further improve photon delivery to subsurface targets. Alongside hardware-based improvements, the application of advanced signal processing algorithms for denoising and signal recovery can enhance the quality and depth of PA signals. In US mode, using lower frequency transducers (e.g., 5 MHz) can reduce acoustic attenuation in tissue, although this typically comes with a trade-off in spatial resolution. Moreover, optimization of the receiving transducer, including enhanced sensitivity and bandwidth tuning for low-frequency scattered signals, can help preserve signal fidelity. Acoustic coupling also plays a critical role; improving the interface between the sample, the ER, and the transducer can significantly enhance transmission efficiency. Finally, adaptive signal processing techniques such as coded excitation, matched filtering, and beamforming strategies tailored for virtual array configurations may further improve the signal-to-noise ratio (SNR) for structures located at greater depths.

APPENDIX

A. Theoretical resolution analysis

In US mode, 10 MHz scattered signals are detected by a single-element ultrasonic transducer and then decoded into signals at an 80 × 80 array of virtual transducers. Based on the spectrum of the signal detected by the integrated PMN-PT transducer (with a 30 MHz central frequency), we can approximate the detection spectrum for each virtual transducer element. From the frequency response (Fig. 7(a)), we calculate the center frequency f_0 and bandwidth f_w as follows:

$$f_0 = \frac{\int_0^{f_m} f|s(f)|^2 df}{\int_0^{f_m} |s(f)|^2 df} \approx 6.50\,\text{MHz} \tag{3}$$

$$f_w = \sqrt{\frac{\int_0^{f_m} (f - f_0)^2 |s(f)|^2 df}{\int_0^{f_m} |s(f)|^2 df}} \approx 3.72\,\text{MHz} \tag{4}$$

From the US imaging signal (Fig. 7(a)), we define the upper frequency as f_m = 35 MHz. The speed of sound in fused silica (C_g = 5.96 mm / μs) is higher than that in water (C_w = 1.48 mm / μs), leading to a critical angle (θ_c) for the refraction of ultrasonic waves as they pass from water into fused silica. This phenomenon results in a limited-view effect in the system (Fig. 7(e) and (f)), as ultrasonic waves with incidence angles greater than the critical angle is completely reflected and cannot be detected by the transducer. The critical angle is calculated as $\sin\theta_c = C_w \,/\, C_g \approx 0.25$, which yields an equivalent numerical aperture (NA) of the virtual transducer 2D array. For a point object P, a larger field of view (FOV 2) tends to have a greater acceptance angle compared to a smaller field of view (FOV 1) due to impedance mismatch as illustrated in Figs. 7(e)–(f). Similarly, for a given FOV, a point object located at P, closer to the interface, tends to have a higher effective numerical aperture (NA) compared to a point object located farther away at P'. The longitudinal wave speed is used here to determine the critical angle, as the single-element transducer is most sensitive to longitudinal waves. For the virtual 2D array, the axial resolution (along the z-axis) is calculated as R_z = 0.88 C_w / f_w = 0.35 mm, while the lateral resolution (in the xy plane) is R_{xy} = 0.71 C_w / NAf_0 = 0.64 mm. Similarly, for PA-mode US signal (Fig. 7(b), we set f_m = 50 MHz and calculate f_0 = 7.60 MHz and f_w = 11.84 MHz. The estimated lateral and axial resolutions are R_z = 0.11 mm and R_{xy} = 0.55 mm, respectively. The theoretically calculated resolution values are very close to the experimentally measured values.

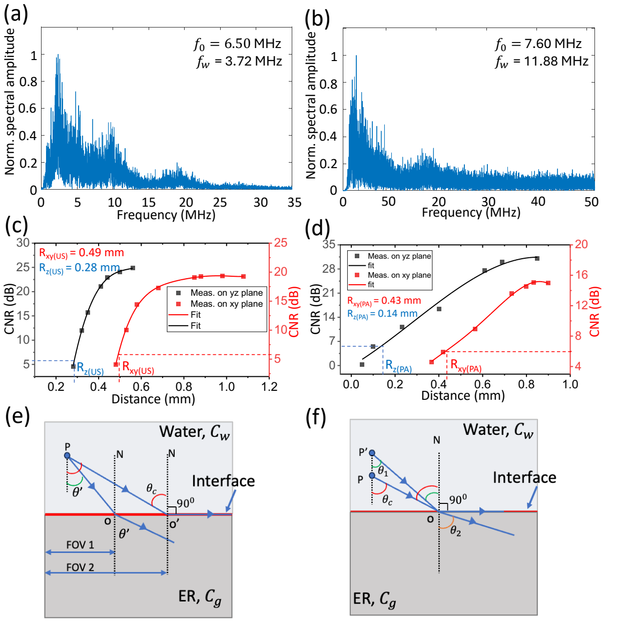


Fig. 7: **Quantification of the center frequency and bandwidth.** Typical frequency spectra of (a) a scattered US signal and (b) a PA signal recorded by a single-element transducer. (c) CNR-based resolution quantification in US mode. The CNR versus the distance between the two lines in the xy and yz planes shows the lateral R_{xy}(US) and axial R_z(US) resolutions, measured as the distance at 6 dB CNR. (d) CNR-based resolution quantification in PA mode. The CNR versus the distance between the two lines in the xy and yz planes shows the lateral R_{xy}(PA) and axial R_z(PA) resolutions, measured as the distance at 6 dB CNR. Illustration showing the effect of (e) FOV and (f) imaging height on NA.

B. Captions for supplementary movies

Movie 1: 3D tracking of a point target in US mode. The movie shows the moving path of a point Phantom 2 reconstructed using single-shot imaging, 5-shots averaging, and 10-shots averaging.

Movie 2: Ink flushing through a tube. The movie shows dual-mode imaging from the Phantom 2 experiment, with ink (red, color-coded for PA mode imaging) flushed through a tube (green, color-coded for US mode imaging), reconstructed using 10-shot, 50-shot, and 100-shot averaging.

Movie 3: *In vivo* dual-mode imaging of human hand. The movie shows dual-mode *in vivo* imaging of the cutaneous features and vessels of the thenar region of the hand. In PA mode, a sphygmomanometer was used to induce blood flow variation (red, color-coded). Data recording began with baseline measurements, followed by cuff inflation and maintenance for several seconds before release. The system was then switched

This article has been accepted for publication in IEEE Transactions on Medical Imaging. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TMI.2026.3653667

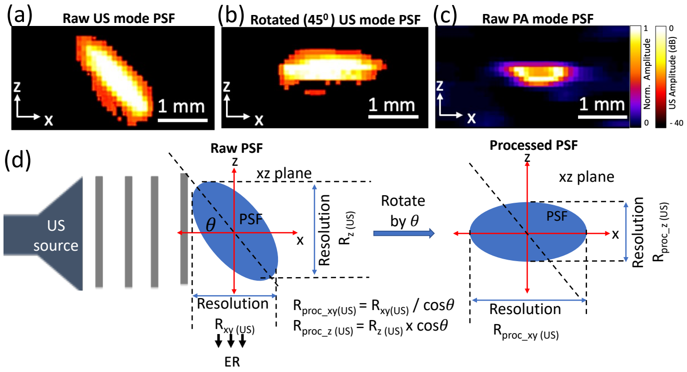


Fig. 8: **Reconstructed PSF.** (a) The US mode raw PSF in the xz plane is inclined at angle of 45° due to orthogonal insonification. (b) For co-alignment with the PA mode PSF, the US PSF is rotated by θ, resulting in scaling up of the z resolution and scaling down of the xy resolution by a factor of $\cos\theta$. (c) 45° rotated US mode PSF in the xz plane. (d) PA mode reconstructed PSF in the xz plane.

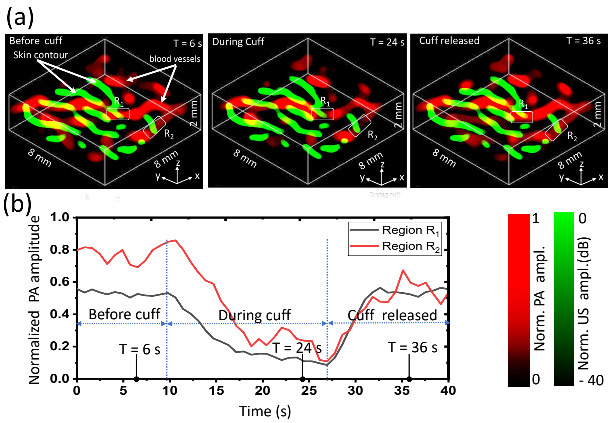


Fig. 9: **Dual-mode *in vivo* imaging of the human hand: participant-2.** (a) Dual-mode images are reconstructed before inflation, during inflation, and after cuff release. Green color-coded structures represent the US images of cutaneous features in the thenar region, while red color-coded structures represent the PA images of blood vessels in the same region. (b) Mean PA amplitude variations observed during the experiment at vascular regions R_1 and R_2. Additional information can be found in Movie 3 and Appendix B.

to US mode to image scattering-based features (green, color-coded) in the same region. Data was acquired at a speed of 50 Hz and reconstructed using 50-shot averaging.

Movie 4: 3D Reconstructed letter phantoms. The movie shows a three-dimensional reconstruction from Phantom 1, displaying the letters “CALTECH,” obtained using the ultrasound (US) mode.

Movie 5: Line phantom moving in FOV. Line Phantom 8 reconstructed using US mode while moving through the imaging field of view.

[1] The data that support the findings of this study are provided within the paper and its Supplementary materials.

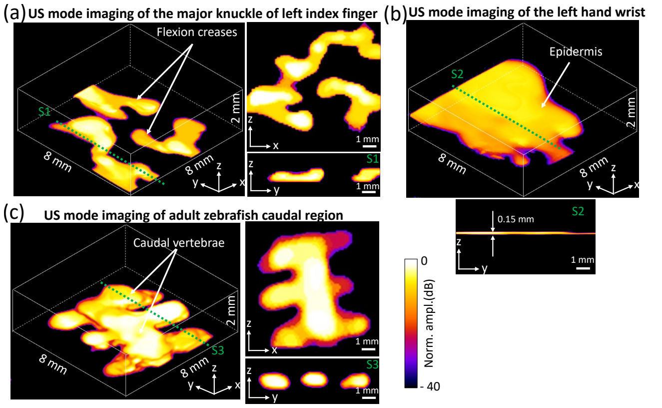


Fig. 10: **US-mode *in vivo* imaging of additional subjects.** Single-shot US imaging of (a) the major knuckle of the index finger and (b) the epidermal layer of the left wrist of an adult human subject. (c) Single-shot imaging of a fixed adult zebrafish skeleton [41].

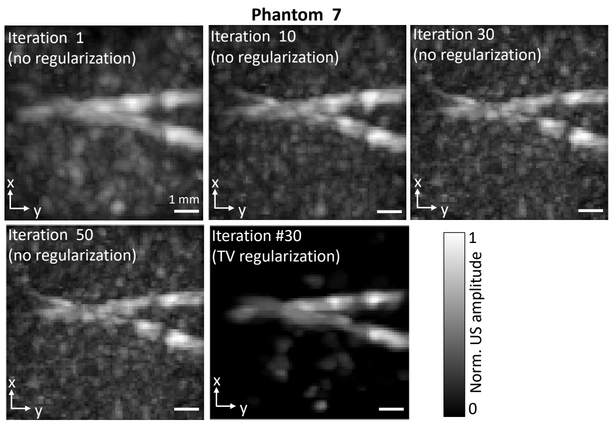


Fig. 11: **Effect of iteration count and TV regularization.** Phantom 7
: US mode reconstructed single strand cross-wire phantom images at different iteration counts, with and without TV regularization.

The reconstruction algorithm and data processing methods can be found in the paper. The reconstruction code is not publicly available because it is proprietary and may be used in licensed technologies.

CONFLICT OF INTEREST

We thank the COILab members for their helpful discussion. L.V.W. has a financial interest in Microphotoacoustics, Inc., CalPACT, LLC, and Union Photoacoustic Technologies, Ltd., which, however, did not support this work. L.V.W is a co-inventor of the patent titled “Single-shot 3D imaging using a single detector” (U.S. Patent No. 18410842), along with Yide Z. This patent is relevant to the subject matter discussed in this manuscript. The rest of the authors declare that they have no competing interests.

---

[1] Data and materials availability

REFERENCES

[1] M. Yang, et al., ”Photoacoustic/ultrasound dual imaging of human thyroid cancers: an initial clinical study,” *Biomed. Opt. express*, vol. 8, no. 7, pp. 3449–3457, 2017.

[2] C. Y. Hao, et al., ”Multimodal photoacoustic/ultrasonic imaging system: a promising imaging method for the evaluation of disease activity in rheumatoid arthritis,” *Eur. Radiol.*, vol. 31, no. 5, p. 3542, 2021.

[3] Q. Zhu, ”A review of co-registered transvaginal photoacoustic and ultrasound imaging for ovarian cancer diagnosis,” *Curr. Opin. Biomed. Eng.*, vol. 22, p. 100381, 2022.

[4] N. Nyayapathi, E. Zheng, Q. Zhou, M. Doyley, and J. Xia, ”Dual-modal photoacoustic and ultrasound imaging: from preclinical to clinical applications,” *Front. Photonics*, vol. 5, p. 1359784, 2024.

[5] M. Xu and L. V. Wang, ”Photoacoustic imaging in biomedicine,” *Rev. Sci. Instrum.*, vol. 77, no. 4, p. 041101, 2006.

[6] L. V. Wang and S. Hu, ”Photoacoustic tomography: In vivo imaging from organelles to organs,” *Science*, vol. 335, no. 6075, pp. 1458–1462, 2012.

[7] A. Fenster, D. B. Downey, and H. N. Cardinal, ”Three-dimensional ultrasound imaging,” *Phys. Med. Biol.*, vol. 46, p. R67, 2001.

[8] E. Merz, ”25 Years of 3D ultrasound in prenatal diagnosis (1989–2014),” *Ultraschall. Med.*, vol. 36, pp. 3–8, 2015.

[9] Gottschalk, S., Degtyaruk, O., Mc Larney, B., Rebling, J., Hutter, M. A., Deán-Ben, X. L., Shoham, S., & Razansky, D. (2019). Rapid volumetric optoacoustic imaging of neural dynamics across the mouse brain. *Nature Biomedical Engineering, 3*(5), 392–401.

[10] Deán-Ben, X. L., Sela, G., Ntziachristos, V., & Razansky, D. (2016). Advanced optoacoustic methods for diagnostic and therapeutic applications. *Optica, 3*(9), 999-1007.

[11] Weidenfeld, I., Zakian, C., Duewell, P., Neu, S. J., Overvelt, E., Scherer, F., ... & Stritzker, J. (2019). Homogentisic acid-derived pigment as a biocompatible label for optoacoustic imaging of macrophages. *Nature Communications, 10*(1), 5056.

[12] Pleitez, M. A., Khan, A. A., Soldà, A., Chmyrov, A., Reber, J., Gasparin, F., ... & Scheideler, M. (2020). Label-free metabolic imaging by mid-infrared optoacoustic microscopy in living cells. *Nature Biotechnology, 38*(3), 293–296.

[13] L. Li, et al., ”Single-impulse panoramic photoacoustic computed tomography of small-animal whole-body dynamics at high spatiotemporal resolution,” *Nat. Biomed. Eng.*, vol. 1, p. 0071, 2017.

[14] L. Lin, et al., ”Single-breath-hold photoacoustic computed tomography of the breast,” *Nat. Commun.*, vol. 9, p. 2352, 2018.

[15] S. Na, et al., ”Massively parallel functional photoacoustic computed tomography of the human brain,” *Nat. Biomed. Eng.*, vol. 6, pp. 584–592, 2022.

[16] M. F. Duarte, et al., ”Single-pixel imaging via compressive sampling,” *IEEE Signal Process. Mag.*, vol. 25, pp. 83–91, 2008.

[17] B. Sun, et al., ”3D computational imaging with single-pixel detectors,” *Science*, vol. 340, pp. 844–847, 2013.

[18] M.-J. Sun, et al., ”Single-pixel three-dimensional imaging with time-based depth resolution,” *Nat. Commun.*, vol. 7, p. 12010, 2016.

[19] D. Stellinga, et al., ”Time-of-flight 3D imaging through multimode optical fibers,” *Science*, vol. 374, pp. 1395–1399, 2021.

[20] P. Kruizinga, et al., ”Compressive 3D ultrasound imaging using a single sensor,” *Sci. Adv.*, vol. 3, p. e1701423, 2017.

[21] X. L. Dean-Ben and D. Razansky, ”Localization optoacoustic tomography,” *Light Sci. Appl.*, vol. 7, p. 18004, 2018.

[22] X. L. Deán-Ben, A. Özbek, H. López-Schier, and D. Razansky, ”Acoustic scattering mediated single detector optoacoustic tomography,” *Phys. Rev. Lett.*, vol. 123, p. 174301, 2019.

[23] G. Montaldo, D. Palacio, M. Tanter, and M. Fink, ”Time reversal kaleidoscope: a smart transducer for three-dimensional ultrasonic imaging,” *Appl. Phys. Lett.*, vol. 84, pp. 3879–3881, 2004.

[24] G. Montaldo, D. Palacio, M. Tanter, and M. Fink, ”Building three-dimensional images using a time-reversal chaotic cavity,” *IEEE Trans. Ultrason. Ferroelectr. Freq. Control*, vol. 52, pp. 1489–1497, 2005.

[25] B. T. Cox and P. C. Beard, ”Photoacoustic tomography with a single detector in a reverberant cavity,” *J. Acoust. Soc. Am.*, vol. 125, pp. 1426–1436, 2009.

[26] M. D. Brown, et al., ”Reverberant cavity photoacoustic imaging,” *Optica*, vol. 6, pp. 821–822, 2019.

[27] Y. Li, et al., ”Snapshot photoacoustic topography through an ergodic relay for high-throughput imaging of optical absorption,” *Nat. Photonics*, vol. 14, pp. 164–170, 2020.

[28] Y. Li, et al., ”Multifocal photoacoustic microscopy using a single-element ultrasonic transducer through an ergodic relay,” *Light Sci. Appl.*, vol. 9, p. 135, 2020.

[29] L. Li, Y. Li, Y. Zhang, and L. V. Wang, ”Snapshot photoacoustic topography through an ergodic relay of optical absorption in vivo,” *Nat. Protoc.*, vol. 14, pp. 164–170, 2021.

[30] Y. Zhao and L. V. Wang, ”Single-shot photoacoustic imaging with single-element transducer through a spatiotemporal encoder,” *J. Biomed. Opt.*, vol. 28, p. 046004, 2023.

[31] Y. Zhang, P. Hu, L. Li, et al., ”Ultrafast longitudinal imaging of haemodynamics via single-shot volumetric photoacoustic tomography with a single-element detector,” *Nature Biomed. Eng.*, vol. 8, pp. 712–725, 2024.

[32] R. M. Pearl and D. Johnson, ”The vascular supply to the skin: An anatomical and physiological reappraisal—Part I,” *Ann. Plast. Surg.*, vol. 11, no. 2, pp. 99–105, 1983.

[33] M. A. Farage, K. W. Miller, E. Berardesca, and H. I. Maibach, ”Clinical implications of aging skin: cutaneous disorders in the elderly,” *Am. J. Clin. Dermatol.*, vol. 10, no. 2, pp. 73–86, 2009.

[34] J. F. Kreusch, ”Vascular patterns in skin tumors,” *Clin. Dermatol.*, vol. 20, no. 3, pp. 248–254, 2002.

[35] S. R. Sternberg, ”Biomedical image processing,” *IEEE Comput.*, vol. 16, pp. 22–34, 1983.

[36] A. Beck and M. Teboulle, ”Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems,” *IEEE Trans. Image Process.*, vol. 18, no. 11, pp. 2419–2434, 2009.

[37] N. Parikh and S. Boyd, ”Proximal algorithms,” *Found. Trends Optim.*, vol. 1, no. 3, pp. 127–239, 2014.

[38] A. Beck and M. Teboulle, ”A fast iterative shrinkage-thresholding algorithm for linear inverse problems,” *SIAM J. Imaging Sci.*, vol. 2, no. 1, pp. 183–202, 2009.

[39] Q. Kong, T. Siauw, and A. Bayen, *Python Programming and Numerical Methods: A Guide for Engineers and Scientists*, Academic Press, 2020.

[40] K. Firouzi, B. T. Cox, B. E. Treeby, and N. Saffari, ”A first-order k-space model for elastic wave propagation in heterogeneous media,” *J. Acoust. Soc. Am.*, vol. 132, no. 3, pp. 1271–1283, 2012.

[41] Nathan C. Bird and Paula M. Mabee, ”Developmental Morphology of the Axial Skeleton of the Zebrafish, Danio rerio”, *Developmental Dynamics* 228:337–357, 2003