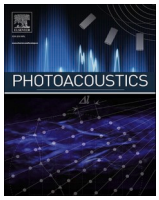
Photoacoustics 42 (2025) 100690

Contents lists available at ScienceDirect

Photoacoustics

journal homepage: www.elsevier.com/locate/pacs



Enhanced clinical photoacoustic vascular imaging through a skin localization network and adaptive weighting

Chuqin Huang [a], Emily Zheng [a], Wenhan Zheng [a], Huijuan Zhang [a], Yanda Cheng [a], Xiaoyu Zhang [b], Varun Shijo [a,b], Robert W. Bing [a], Isabel Komornicki [c], Linda M. Harris [c], Ermelinda Bonaccio [d], Kazuaki Takabe [e], Emma Zhang [a], Wenyao Xu [b], Jun Xia [a,b,*]

[a] Department of Biomedical Engineering, University at Buffalo, The State University of New York, Buffalo, NY 14228, United States
[b] Department of Computer Science and Engineering, University at Buffalo, The State University of New York, Buffalo, NY 14228, United States
[c] Department of Surgery, University at Buffalo, The State University of New York, Buffalo, NY 14228, United States
[d] Department of Breast Imaging, Roswell Park Comprehensive Cancer Center, Buffalo, NY, 14263, United States
[e] Department of Surgery, Roswell Park Comprehensive Cancer Center, Buffalo, NY 14263, United States

ARTICLE INFO

Keywords:
Photoacoustic
Photoacoustic tomography
Deep learning
Image enhancement

ABSTRACT

Photoacoustic tomography (PAT) is an emerging imaging modality with widespread applications in both preclinical and clinical studies. Despite its promising capabilities to provide high-resolution images, the visualization of vessels might be hampered by skin signals and attenuation in tissues. In this study, we have introduced a framework to retrieve deep vessels. It combines a deep learning network to segment skin layers and an adaptive weighting algorithm to compensate for attenuation. Evaluation of enhancement using vessel occupancy metrics and signal-to-noise ratio (SNR) demonstrates that the proposed method significantly recovers deep vessels across various body positions and skin tones. These findings indicate the method's potential to enhance quantitative analysis in preclinical and clinical photoacoustic research.

1. Introduction

Photoacoustic (PA) tomography (PAT) is an emerging imaging technique based on the photoacoustic effect, which maps the optical absorption with acoustic detection [1–3]. This technique demonstrates promising capabilities in generating high-resolution images in deep tissues as the attenuation of acoustic waves is much less than the light in tissue. Such advancements position PAT favorably for diverse applications in both preclinical and clinical imaging fields, including but not limited to breast cancer detection [4–8], foot ulcer imaging for perfusion [9–11], palm biometric extraction [12], and skin cancer detection [13].

Like other optical imaging methods, the performance of PAT is affected by skin color. It is widely observed that the PA signal in dark skin surpasses those in light skin due to higher optical absorbance [14]. Consequently, it impedes the visualization of vessels beneath the skin. In addition, most PAT results were demonstrated in Maximum Amplitude Projection (MAP), where the skin signals might easily suppress the underlying vasculature. Considering that various photoacoustic studies rely on analyzing vascular features [8,15,16], this limitation might lead to compromised results, especially in subjects with dark skin [17,18]. In addition, the optical and acoustic attenuation in tissues will also degrade the PA signals over depth. This is extremely important in peripheral arterial disease, in which African American can disproportionately effected [19]. Without proper consideration of the skin location, a depth-enhanced weighting strategy would amplify both the skin and vessel signals.

Different methodologies have been introduced to mitigate the effects of skin signals in PA imaging. Zhang *et al.* developed an automatic skin profile detection method to locate the skin layers in photoacoustic microscopy (PAM) images [20]. Kim *et al.* revealed deep vessels by removing skin layers in PA images from super-resolution localization PAM [21]. However, clinical PAT systems usually employ low-frequency transducers for a deep imaging depth, and the axial resolution is much poorer compared to PAM systems. Therefore, it is challenging to distinguish the edge of skin layers in PAT. To address this issue, *Lee et al.* and Choi *et al.* utilized ultrasound (US) to identify the skin contour and then used that information to remove the skin signals in PAT[9,22]. However, some PAT systems do not have US imaging capabilities.

* Corresponding author at: Department of Biomedical Engineering, University at Buffalo, The State University of New York, Buffalo, NY 14228, United States.
E-mail address: junxia@buffalo.edu (J. Xia).

https://doi.org/10.1016/j.pacs.2025.100690
Received 31 October 2024; Received in revised form 19 December 2024; Accepted 15 January 2025
Available online 21 January 2025

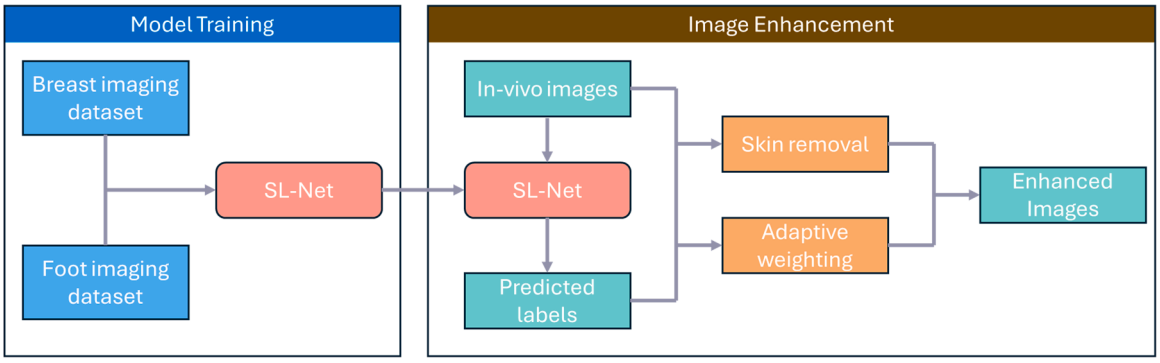


Fig. 1. Framework of the proposed algorithm. The workflow can be divided into two sections: SL-Net training and image enhancement. The SL-Net is trained with manually labeled images, and the trained model can predict skin locations in the in-vivo PA images. Based on the labeled skin location, the image enhancement section utilizes an adaptive weighting algorithm to recover deep vessels.

Deep learning, as a rapidly growing field, has been widely applied to medical imaging in recent years. Numerous deep learning methodologies have been proposed in photoacoustic imaging, targeting to improve the image quality from different aspects [23–26]. As for skin signal identification, Ly *et al.* presented a Sliding-U-Net model for segmenting skin and vessel profiles in PAM [27]. Yuan *et al.* proposed Hy-Net for blood vessel segmentation for PA images [28]. Zhao *et al.* implemented HM-3DCE-Net to segment the OR-PAM images in 3D space [29]. While all these models can detect the vessels in PA images and hence eliminate the effect of skin signals, they were all trained and validated on PAM images, which have much higher resolution due to high-frequency transducers. Therefore, they might not perform well on low-frequency PAT imaging systems, because the skin features are different. Schellenberg *et al.* developed a deep-learning network for semantic segmentation on multispectral PA images [30]. However, it requires multi-wavelength PA images where vessels are enhanced, posing practical limitations in single-wavelength PAT systems [11]. Zheng *et al.* implemented a 3DFD UNet to enhance the vessel in the PA images from the linear-array PAT system and reduce the signal signals [31]. However, this model was trained based on simulation data, which cannot fully mimic the variation in human skin color. Recently, Wang *et al.* introduced a skin-removal algorithm based on a 2.5D-DeepFPN model [32]. While their method worked well in light-skin-colored subjects, the team did not demonstrate its effectiveness in a clinical setting nor in subjects with dark skin color. In addition, subjects were scanned with skin immersed in water, which may not be feasible in clinical practice.

In this study, we proposed a novel framework combining a deep learning network for segmenting the skin layers and an adaptive weighting algorithm to account for signal attenuation beneath the skin. This combined approach aims to recover deep vessels in PA images, facilitating quantitative analysis in translational research.

2. Methods

2.1. Overall workflow

The overall workflow is shown in Fig. 1. We started with the manual labeling of skin pixels in more than 5000 frames extracted from PA images of human breasts and feet. These data were used to train a residual UNet. The trained network is named the Skin Localization Network (SL-Net). Leveraging on the skin locations predicted by SL-Net, we further developed an adaptive weighting method to restore deep vessels, resulting in better recovery of vascular features across different depths. The efficiency of this approach is evaluated by quantitatively comparing the vessel occupancy before and after processing.

2.2. Dataset preparation

2.2.1. Data source

The original photoacoustic imaging data used in this study comes from two PA clinical studies: The breast cancer imaging project and the foot ulcer imaging project [11,31]. We selected the breast and foot as the anatomical sites for our experiments based on their clinical significance. Breast cancer is the most diagnosed cancer worldwide and one of the leading causes for cancer related death in women, early detection is extremely important to improve survival rate [33,34]. Compared to traditional imaging techniques such as X-ray mammogram, MRI and ultrasound, photoacoustic imaging provides a high-resolution, non-radiation and cost-effective imaging solution for breast especially in dense breast [35,36]. Chronic leg ulcers affect approximately 6.5 million Americans and are associated with significant morbidity, reduced quality of life and high treatment cost [37]. Since many chronic ulcers have underlying vascular insufficiency, accurate assessment of tissue perfusion is critical to treatment planning and monitoring. Compared to other existing diagnostic techniques, photoacoustic imaging systems are capable of precisely evaluating perfusion conditions without radiation or contrast agent injection [9–11].

In the breast imaging system, a U-shaped scan is used to enlarge the field of view to ensure that the whole breast of the subject is covered. The light source is provided by a portable laser with a 20 Hz pulse repetition frequency (PRF). Photoacoustic signals are captured by a customized linear transducer array with a central frequency of 2.25 MHz. The received signals are then transmitted to a 40 dB pre-amplifier (Photosound) and acquired by a data acquisition system (Vantage 256, Verasonics) with 14.28 MHz sampling rate. For the foot imaging system, a customized linear transducer array with 2.25 MHz is mounted on the translation stage to perform the imaging of the dorsal side of the foot. A 10 Hz portable laser is employed to provide excitation and a portable DAQ (Photosound) is used to capture the PA signals with 40 MHz sampling rate. More details of the experiment setup can be found in the reference [5,11]. For in-vivo imaging experiments mentioned above, all studies were approved by the Institutional Review Board of the University at Buffalo, under different study protocols for breast imaging and foot imaging. All human subjects provided informed consent after fully understanding the implications of their participation. Subjects were recruited by clinical collaborators.

The acquired PA signals are first filtered by a 2–4 MHz bandpass filter and then reconstructed using the delay-and-sum method [38]. The cross-sectional frames are then stacked to form C-scan images. For breast imaging, we have 850 frames in each C-scan, while for foot imaging, we have 500 frames in each C-scan. Notably, since the same transducers are implemented for both breast and foot imaging, these images are fed into the same network.

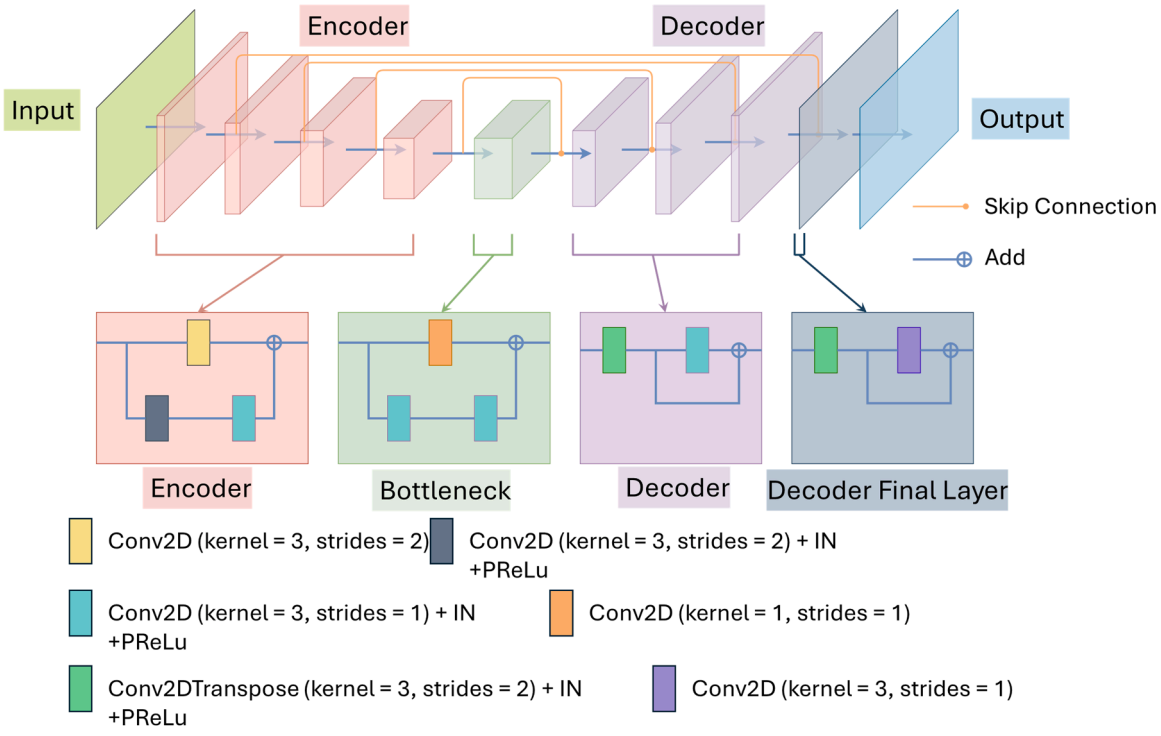


Fig. 2. The architecture of the proposed Skin-Localization Net (SL-Net).

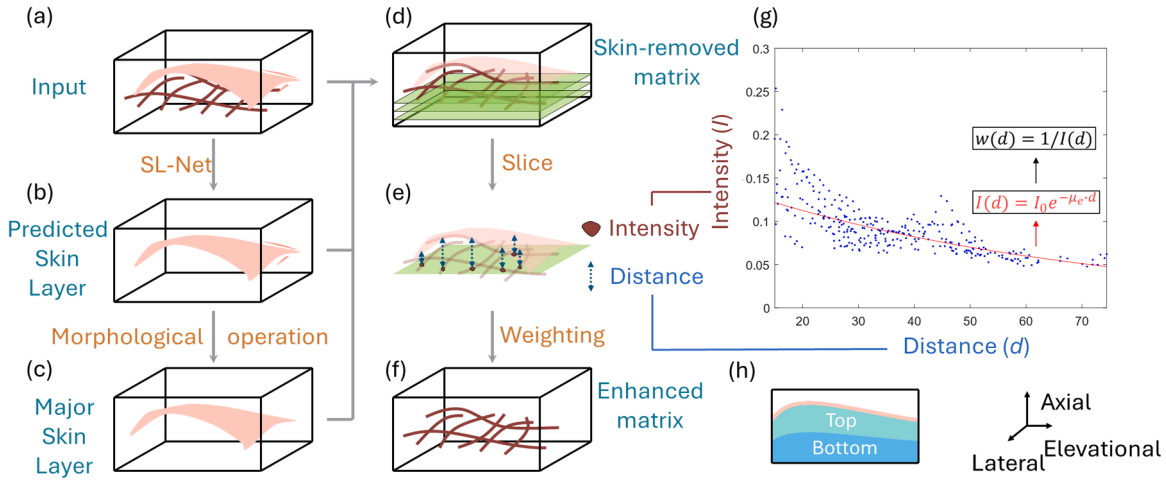


Fig. 3. The workflow of the adaptive weighting calculation.

2.2.2. *Data labelling*

We utilized slices in the axial-lateral direction of the reconstructed data as the network input. Each frame (B-scan) in the dataset underwent normalization by scaling image intensity to a range from 0 to 1. Manual labeling of the skin layer in PA images was executed using the brush tool within the MATLAB 'Volume Segmentor' toolbox. After labeling all the frames, we saved the resulting labels as a logical matrix mirroring the size of the C-scan image. Subsequently, frames lacking labels were omitted from the dataset. The final dataset includes 5313 pairs of frames and labels, with 2438 pairs from the foot imaging dataset and 2,8752875 pairs from the breast imaging dataset. Notably, images included in the training set are captured from multiple subjects with various skin tones to improve the model's generalization ability.

However, it's important to note that the labels might lack precision at the pixel level due to limitations stemming from the relatively poor resolution of the low-frequency transducer used in both human imaging studies. Additionally, the presence of noise in the reconstructed data might further complicate the accurate delineation of the skin layer.

2.3. *Data pre-processing and data augmentation*

As the original PA images are acquired from different subjects at different positions and from different systems, the sizes of the frames are different. After loading the images and labels from the training dataset, the images and corresponding labels that have pixel numbers less than 256 along the axial direction are padded to 256 pixels to fit the input size of the network. After that, the intensity of the image is scaled into a range from 0 to 1.

Owing to the intricate nature of PAT images, the manual labeling of the skin region at the pixel level is challenging and time-intensive. Consequently, the adoption of data augmentation techniques becomes imperative to this study. The augmentation process involves a series of

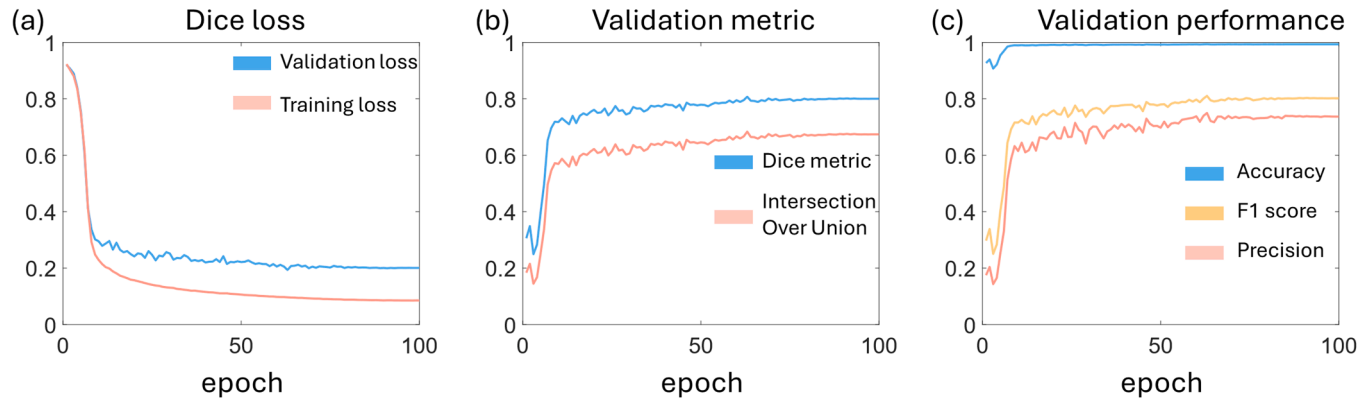


Fig. 4. Model performance during training. (a) The dice loss in the training dataset and validation dataset. (b) The dice metric and intersection over union (IoU) performance in validation. (c) The accuracy, f1 score, and precision in validation.

Table 1
The comparison of the performance of various models. IoU: intersection over union; Acc: Accuracy; Prec: Precision.

Model	Dice score	Dice loss	IoU	Acc	F1 score	Prec	Trainable Params num	Inference time per frame (ms)
SL-Net	**0.795**	**0.205**	**0.669**	**0.993**	**0.804**	**0.739**	6.495 M	0.049 ± 0.006
UNet	0.765	0.235	0.628	0.992	0.763	0.703	2.638 M	**0.024 ± 0.003**
UNet+ +	0.747	0.253	0.607	0.990	0.732	0.632	**0.152 M**	0.088 ± 0.011
VNet	0.794	0.206	0.666	0.992	0.801	0.702	9.358 M	0.056 ± 0.006
SwinUNetR	0.792	0.208	0.655	0.992	0.795	0.723	25.122 M	0.158 ± 0.019

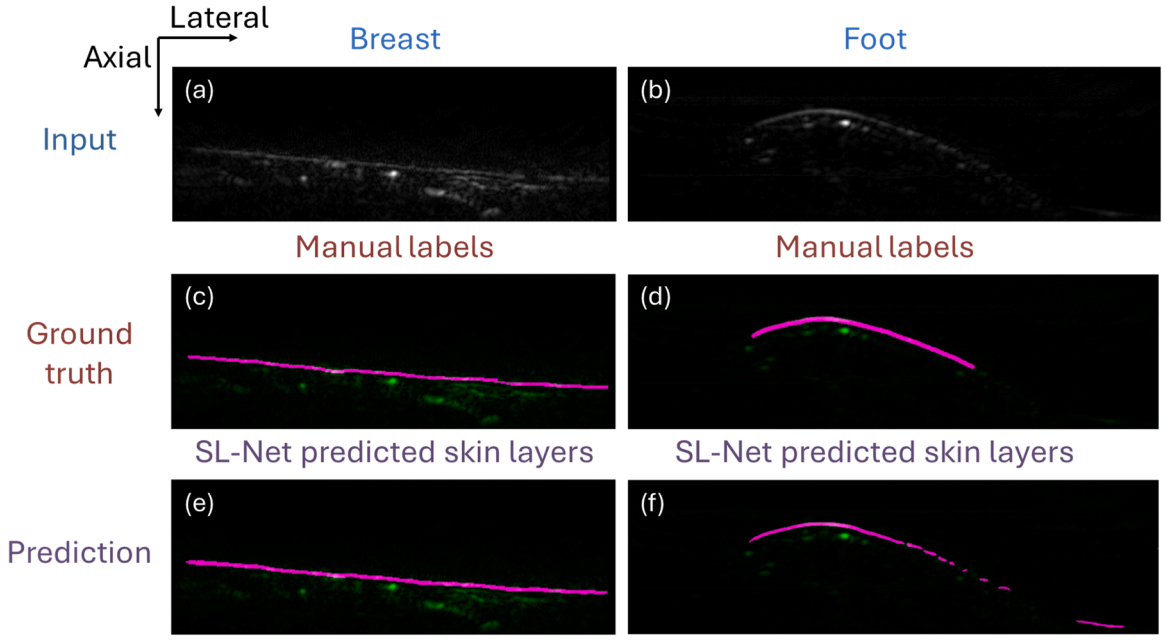


Fig. 5. Performance evaluation on the in-vivo data. (a) An original cross-sectional image of the breast in the test dataset. (b) An original cross-sectional image of the foot in the test dataset. (c) A manual skin label of the breast image in (a). (d) Manual skin label of the foot image in (b). (e) Predicted skin location of the breast image, using the trained SL-Net. (f) Predicted skin location of the foot image, using the trained SL-Net.

operations. Initially, the image undergoes random scaling, with a scaling factor chosen uniformly between 0.8 and 1.25. Subsequently, the scaled images are subjected to random rotation within an angular range of −0.05 to 0.05 radians. Following rotation, the algorithm randomly cropped the image into four sub-images, each with dimensions of 256 by 256 pixels, with the central pixel corresponding to a location within the skin region. Additionally, the augmentation protocol incorporates flipping, executed with 0.5 probability. Meanwhile, the same transformations have been applied to the corresponding labels. These augmentation strategies diversify and enrich the training dataset, thereby enhancing the model's capacity to generalize across various skin features.

2.4. *The model architecture of Skin-Localization-Net (SL-Net)*

As shown in Fig. 2, the SL-Net is developed based on 2D deep residual UNet architecture [39]. The model facilitates the transfer of spatial information from the encoder to the decoder through skip connections. Furthermore, the inclusion of residual connections addresses the vanishing gradient problem, ensuring effective gradient learning and

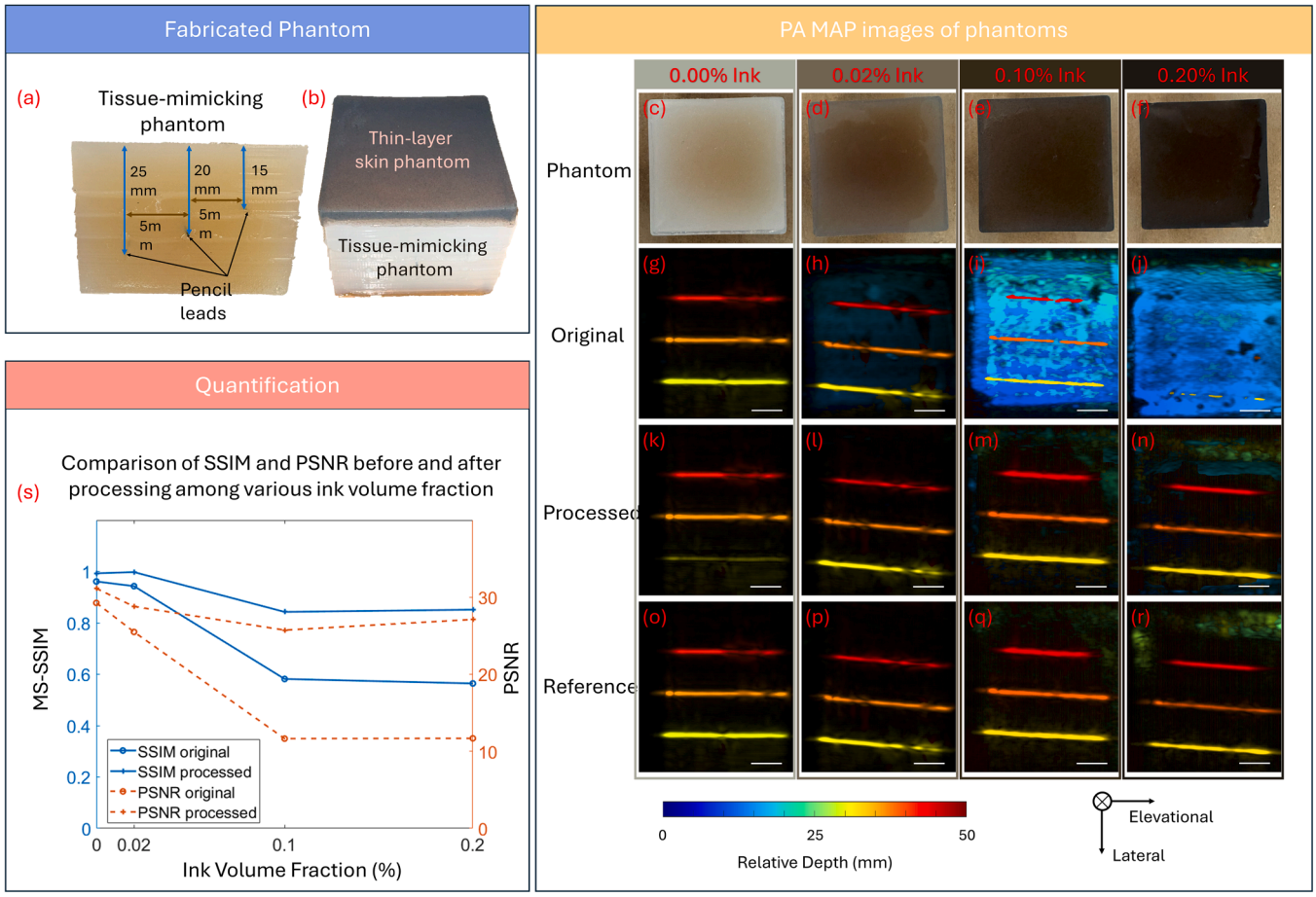


Fig. 6. Phantom experiment results. (a) Side view of the tissue-mimicking phantom, pencil leads are marked by black arrow and their locations are labeled. (b) A photo of a combined phantom. A thin-layer skin phantom (0.1 % volume fraction ink) is stacked on the top of the tissue-mimicking phantom. (c) –(f) Photos of phantoms with 0.00 %, 0.02 %, 0.10 % and 0.20 % volume fraction India Ink added, respectively. (g)-(j) Original PA Maximum amplitude projection (MAP) images of phantoms with various volume fraction ink. (k)-(n) Skin removed PA MAP images of phantoms by the proposed SL-Net. (o)-(r) MAP images of tissue-mimicking phantoms (s) Comparison of SSIM and PSNR before and after processing among various volume fraction of ink. All PA images are depth-encoded. Scale bar: 10 mm.

propagation through the network's deep layers.

The network comprises four residual blocks in both the encoding and decoding pathways, with a bottleneck residual block serving as an intermediary between these two branches, as shown in Fig. 2. The spatial dimension reduction of feature maps is achieved with the convolution with a stride of 2, which replaces the max pooling layer in the basic UNet model, allowing the network to learn flexible spatial transformations. The residual connection is widely applied to all the blocks. Within the contracting path, each block integrates a residual connection and a convolution unit featuring a stride of 2 (gray block). The residual connection encompasses a two-stride convolution (yellow block) and a one-stride convolution (cyan block), both utilizing a kernel size of 3. In the bottleneck structure, the residual connection involves two convolutions with a kernel size of 3 and a stride of 1, while the convolution unit (orange block) deploys a kernel size of 1 and a stride of 1. Within the expansive path, all convolutions employ a kernel size of 3. Each layer consists of a transpose convolution (green) with a stride of 2, a convolution with a stride of 1, and a residual connection. The model uniformly implements instance normalization with the momentum set to 0.1 and an epsilon value of 10^{-5}. Additionally, Parametric Rectified Linear Unit (PReLU) activation functions are utilized across the network for non-linear transformations.

2.5. Deep learning implementation

The model and data augmentation are built based on the Medical Open Network for Artificial Intelligence (MONAI) library, which is a widely used deep learning toolkit for medical image analysis [40]. All the network-related scripts are implemented using Python with Pytorch 1.12.1 plus cuda 11.3 backend. The dataset is randomly split with a ratio of 0.75: 0.2: 0.05 for training, validation, and testing, respectively. The training platform is a workstation with AMD Ryzen 9 3950X CPU, 128 GB RAM, and Nvidia GTX 3090 graphic card.

The loss function implemented in this training is the dice loss with the sigmoid function for prediction [41]. A sigmoid function is used as the activation function combined with a threshold at 0.5 to convert the model prediction into binary output. Considering the different sizes of the frames coming from different systems, we utilized the SlidingWindowInferer function in the MONAI library in validation and test procedures, with an overlap ratio set to 0.5. To evaluate the model performance, we selected several metrics: dice metric, dice loss, mean intersection over union (IoU), accuracy, F1 score, and precision. These metrics are tracked during the validation and calculated on the test dataset after training.

During the model training phase, we have implemented various optimization techniques to enhance performance. The AdamW optimization method is employed [41], with a learning rate set at 0.001 and a weight decay of 10^{-5}. Additionally, we utilized the cosine annealing schedule to further optimize training dynamics [42].

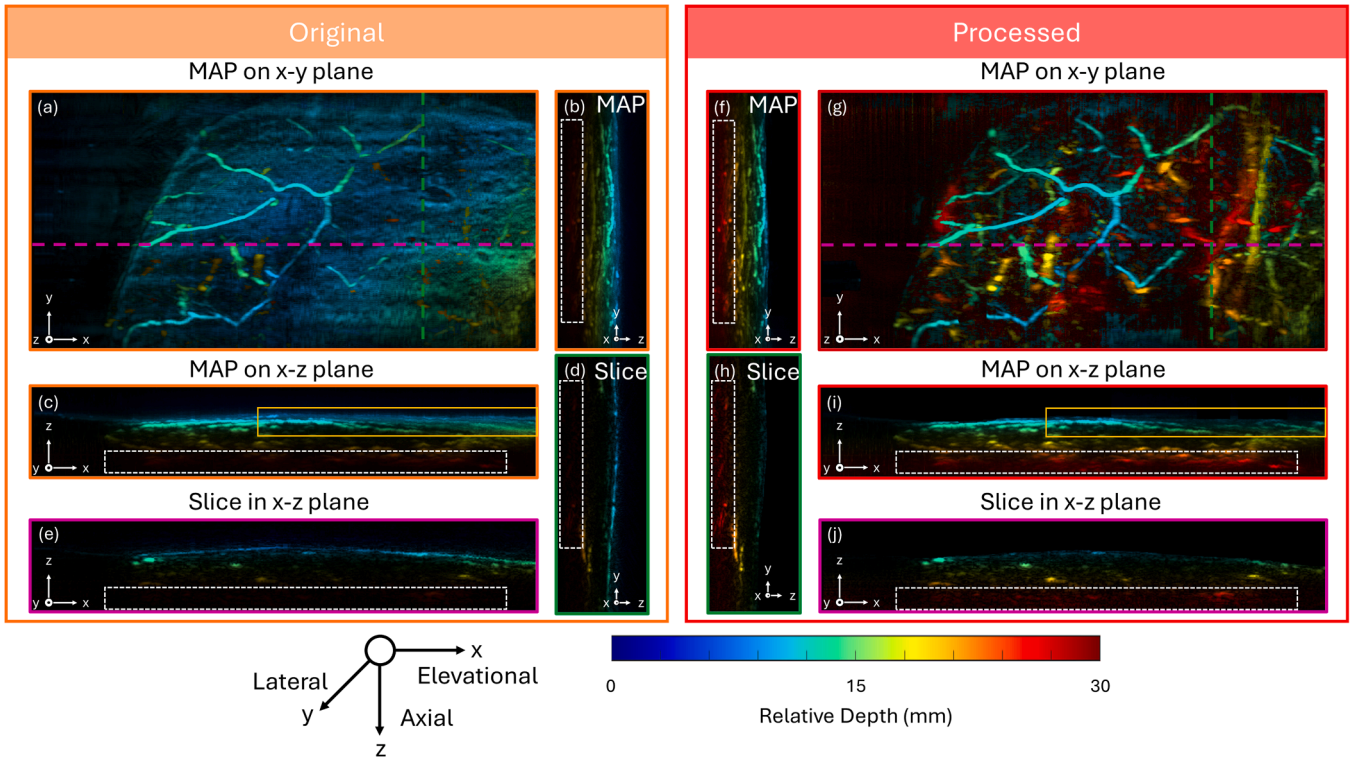


Fig. 7. Comparison of maximum projection amplitude (MAP) images and slices between original and processed data from the same subject. (a) The original MAP image projected on the elevational-lateral plane. (b) The original MAP image projected on the lateral-axial plane. (c) The original MAP image on the elevational-axial plane. (d) An original PA image sliced in the lateral-axial plane (along green dashed line in (a)). (e) An original PA image sliced in the elevational-axial plane (along magenta dashed line in (a)). (f) The processed MAP image projected on the lateral-axial plane. (g) The processed MAP image projected on the elevational-lateral plane. (h) A processed image sliced in the lateral-axial plane (along green dashed line in (g)). (i) The processed MAP image projected on the elevational-axial plane. (j) A processed image sliced in the elevational-axial plane (along magenta dashed line in (a)). Deep vessels recovered by the proposed method are highlighted by white dash rectangle while removed skin layers are marked by yellow box. All PA images are depth encoded. x, y, z axis represents elevational, lateral and axial direction, respectively.

2.6. Adaptive weighting based on skin location

The initial photoacoustic signal can be written as:

$$p_0(\vec{r}) = \Gamma A_e = \Gamma \mu_a F(\vec{r}) \tag{1}$$

Where Γ denotes the Grueneisen parameter, A_e is the absorbed energy density, which is a product of the absorption coefficient μ_a and the local optical fluence $F(\vec{r})$. However, the optical energy density decreases over depth in tissue due to light absorption and scattering. The local optical fluence can be expressed as:

$$F(\vec{r}) = F_0\left(\vec{r}_s\right) e^{-\mu_{eo}\left(\vec{r}_s - \vec{r}\right)} \tag{2}$$

Where $F_0(\vec{r}_s)$ is the optical fluence of the incident light at skin layers, μ_{eo} is the effective optical attenuation coefficient of the tissue, which combines absorption and scattering [43]. Here, we assume that the tissue is optically homogeneous within our imaging depth for both breast and foot. By combining Eq. (1) and Eq. (2), the initial photoacoustic signal can be written as:

$$p_0(\vec{r}) = \Gamma \mu_a F_0\left(\vec{r}_s\right) e^{-\mu_{eo}\left(\vec{r}_s - \vec{r}\right)} \tag{3}$$

On the other hand, the acoustic signals also attenuate during propagation to the detector at a rate of approximately 0.6 dB/MHz/cm [44]. Considering that both light intensity and acoustic signals attenuate exponentially over depth, we can use the exponential function to derive a weighting factor that compensates for the attenuation.

To compensate for the attenuation, generic weighting has been widely implemented in multiple studies [45,46]. It is constructed based on Eq. (3) and can be expressed as an inverted exponential decay function applied to the whole area starting from an assumed flat skin surface [4,47]. However, since the shape of skin layers varies among subjects and imaging anatomical sites, the generic weighting can be improved by utilizing skin locations predicted by SL-Net into consideration. As shown in Fig. 3, we have taken a series of sequential steps to estimate the effective attenuation coefficient. First, as shown in Figs. 3 (a)-3(c), the skin labels predicted by the SL-Net undergo a morphological filtering process wherein the region with the largest area is selected as the major skin label. This process allows us to remove false labelled skins, which typically come from noises and imaging artifacts. Subsequently, a skin-removed matrix presented in Fig. 3(d) is generated by combining the original reconstruction image with the major skin labels. We then initiated an iterative search along the axial direction to determine the relationship between photoacoustic intensity and the axial (depth) distance. To extract the vessel regions in each slice, which are marked by green in Fig. 3(e), we first performed contrast enhancement with MATLAB built-in function 'imadjust', and then binarized the enhanced slice with a threshold set at 1. To account for noise, we only binarize regions with more than 100 pixels. They are then classified as vessels containing regions and their average intensity is calculated. Meanwhile, the distance from the regions to the skin layers, which is indicated by the blue arrow in Fig. 3(e), is estimated. The corresponding average intensity and distance are stored as data pairs. An example of intensity vs depth data is shown in Fig. 3(g).

As illustrated in Fig. 3(g), the distribution of mean intensity relative to distance adheres to an exponential decay curve. To enhance fitting accuracy, we removed the first 20 % and the last 10 % of distance data.

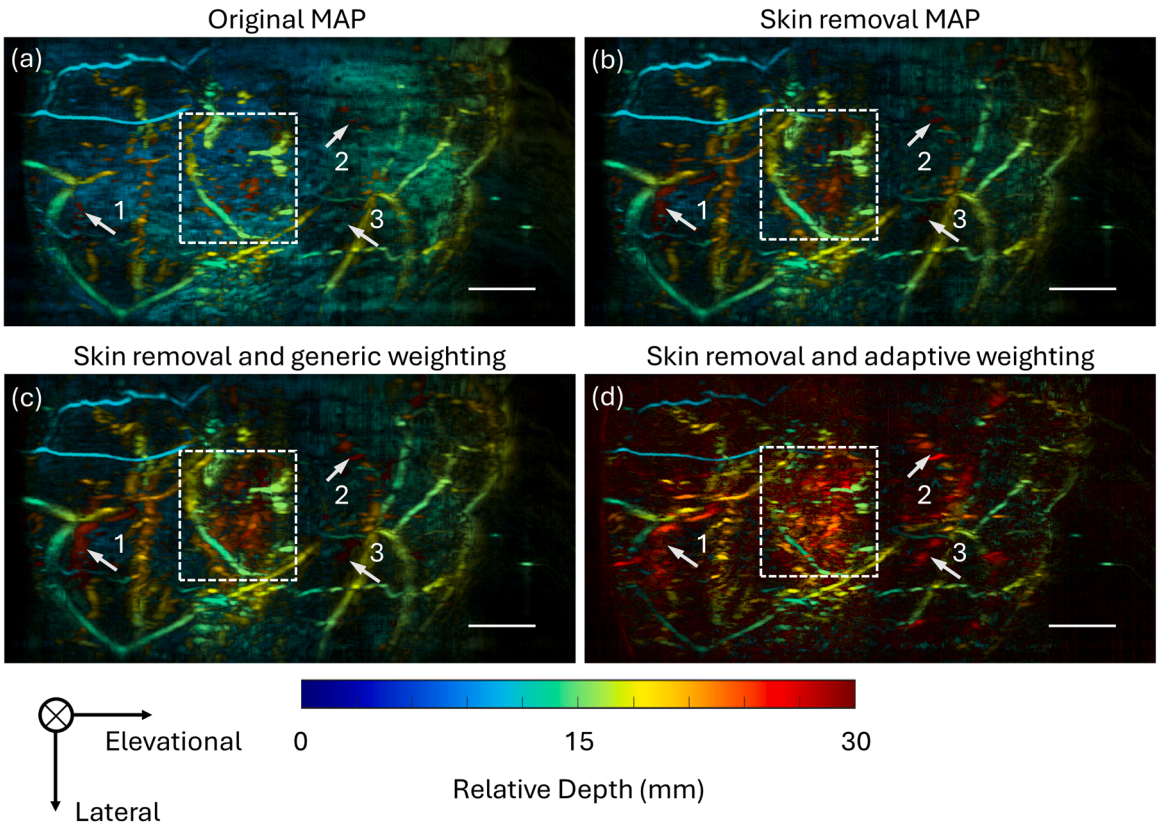


Fig. 8. Demonstration of the skin removal and depth enhancement method on a subject with dark skin color. (a) A MAP image from the original data without any weighting. (b) A MAP image from the skin-removed data without any weighting. (c) A MAP image from the skin-removed data with generic weighting. (d) A MAP image from the skin-removed data with adaptive weighting.

This adjustment is made to account for the relatively low signal-to-noise ratio and sparse distribution of vessels in the deepest region and the potential misclassification of skin labels in the shallow region. Subsequently, a one-degree exponential fitting using the Nonlinear Least Squares method is applied to the curve. The final step involves the calculation of the enhanced matrix shown in Fig. 3(f) by multiplying the weighting obtained from the fitted curve with the skin-removed matrix.

2.7. Vessel occupancy quantification

To quantify the effect of the adaptive weighting method, we utilized the vessel occupancy metric proposed in [11]. Initially, the Frangi filter was applied to the MAP images to enhance vessel contrast [48]. Subsequently, contrast-limited adaptive histogram equalization (CLAHE) and moving average were employed to further improve image quality [49]. Vessel masks were segmented using the ISODATA method [50], while the convex hull was computed to label the region of interest (ROI). Vessel occupancy was then estimated as the ratio of the vessel area to the ROI area. The metric is calculated at both global and local levels. Fig. 3 (h) presents a schematic drawing of a cross-sectional image, where the skin is labeled in pink. The entire area under the skin layers is treated as the 'Global' region. We then divided the global region into two equal subregions: top half and bottom half, marked as tint and blue in Fig. 3 (h), respectively. On the global scale, we projected the whole volume beneath the skin into MAP images and quantified the vessel occupancy. For local quantification, we extracted the pixels belonging to the selected subregions from the global MAP image to form a regional MAP and quantified its local vessel occupancy.

2.8. Signal-to-noise ratio quantification

Given the difficulty in obtaining ground-truth vessel distribution in tissues, the vessel masks segmented during vessel occupancy quantification were used to identify signal regions, while non-vessel areas were labeled as background. To evaluate image quality, the signal-to-noise ratio (SNR) was calculated as the ratio of the mean signal intensity within the vessel regions to the standard deviation of the background noise, following the approach in [51].

3. Results

3.1. Model performance

The loss and the evaluation metrics are tracked during the training and validation processes. The training took approximately 3 hours and 8 minutes to complete a 100-epoch session using the workstation described in method section. Results shown in Fig. 4 demonstrate that the SL-net converged during the training, and the network performed well in both the training and validation datasets.

To evaluate the effectiveness of the proposed model, we conducted a comparison of various segmentation models, including well-established architectures like U-Net and its variance. All models were trained using the same dataset and hypermeters, and the model was constructed based on the MONAI library [40]. Table 1 presents the evaluation metrics for the different models on the test dataset. Among these models, the proposed SL-Net demonstrated the best performance on this task.

Additionally, we evaluated the inference time of each model, considering that faster processing speeds are essential for enhancing clinical PA imaging by enabling rapid diagnosis. To assess this, we calculated the average inference time per frame for each model. The

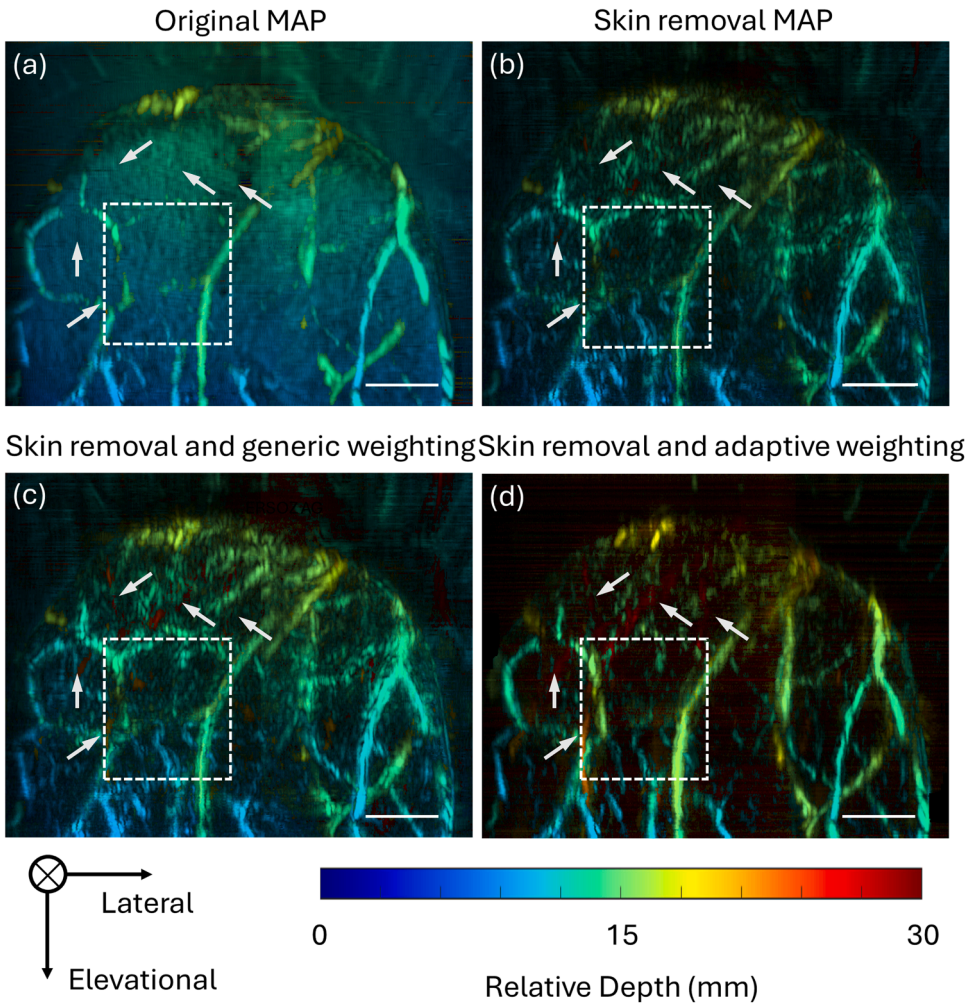


Fig. 9. Comparison between original and enhanced PA images. (a) An original MAP image from the PACT system. (b) A MAP image from the skin-removed data without any weighting. (c) A PA MAP image with skin removal and generic weighting. (d) A PA MAP image with skin removal and adaptive weighting. The suspicious region is labeled by the white dashed rectangle. The scale bar represents 20 mm.

evaluation was conducted using a dataset comprising 10 samples from foot imaging (500 frames per sample) and 10 samples from breast imaging (950 frames per sample). The inference time of each model was recorded for every sample, and the average inference time per frame was calculated. These experiments were performed on a desktop equipped with an AMD Ryzen 5 5600 G processor and an NVIDIA RTX 3060 GPU, which is a commonly used configuration in practical scenarios.

Among the compared models, the proposed SL-Net demonstrated the second fastest inference speed, surpassed only by the original UNet. Although SwinUNetR achieved evaluation metrics comparable to those of SL-Net, its inference time was more than three times longer, limiting its potential in 3D volumetric data processing. We have conducted a comprehensive investigation comparing performance of these models on in-vivo samples from both breast and foot in Appendix A section Model Performance Evaluation.

3.2. Cross-section inspection for skin segmentation

As shown in Fig. 5, we randomly picked two images in the test dataset to examine the performance of the SL-Net. The first row presents the original images, which are inputs of the model. The second row is the ground truth, created by manual labeling. The last row demonstrates the predicted skin location. Comparing the two, we can see that the model performs as well as human labeling in both breast and foot images.

3.3. Phantom experiments

We conducted phantom experiments to validate the performance of the proposed SL-Net. The phantom used in this study comprises two components: a cubic tissue-mimicking section embedded with pencil leads to represent human tissue with vessels, and a series of thin-layer phantoms designed to simulate skin layers of varying skin tones. Detailed information on phantom fabrication was provided in Appendix A Section Methods for Phantom Experiments.

The fabricated tissue-mimicking phantom is shown in Fig. 6(a), where the pencil leads are highlighted by black arrows, and their dimensions are labeled. Fig. 6(b) presents the combined phantom, consisting of the tissue-mimicking phantom and a thin-layer skin phantom with 0.10 % volume fraction ink as an example. Top-view photographs of the fabricated phantoms with Indian Ink volume fractions of 0.00 %, 0.02 %, 0.10 %, and 0.20 % are displayed in Fig. 6(c)-(f), respectively. The corresponding original PA MAP images of these phantoms are presented in Fig. 6(g)-(j). All PA images are depth-encoded to illustrate depth information clearly. As the volume fraction of ink increased in the thin-layer skin phantom, the PA response from the skin phantom became stronger. In the MAP image acquired from the 0.10 % ink phantom (Fig. 6(i)), the PA intensity of the thin-layer skin phantom was comparable to that of the pencil leads. Moreover, in the MAP image of the 0.20 % ink phantom (Fig. 6(j)), it became difficult to identify the pencil leads due to the stronger PA signal from the skin phantom. We applied

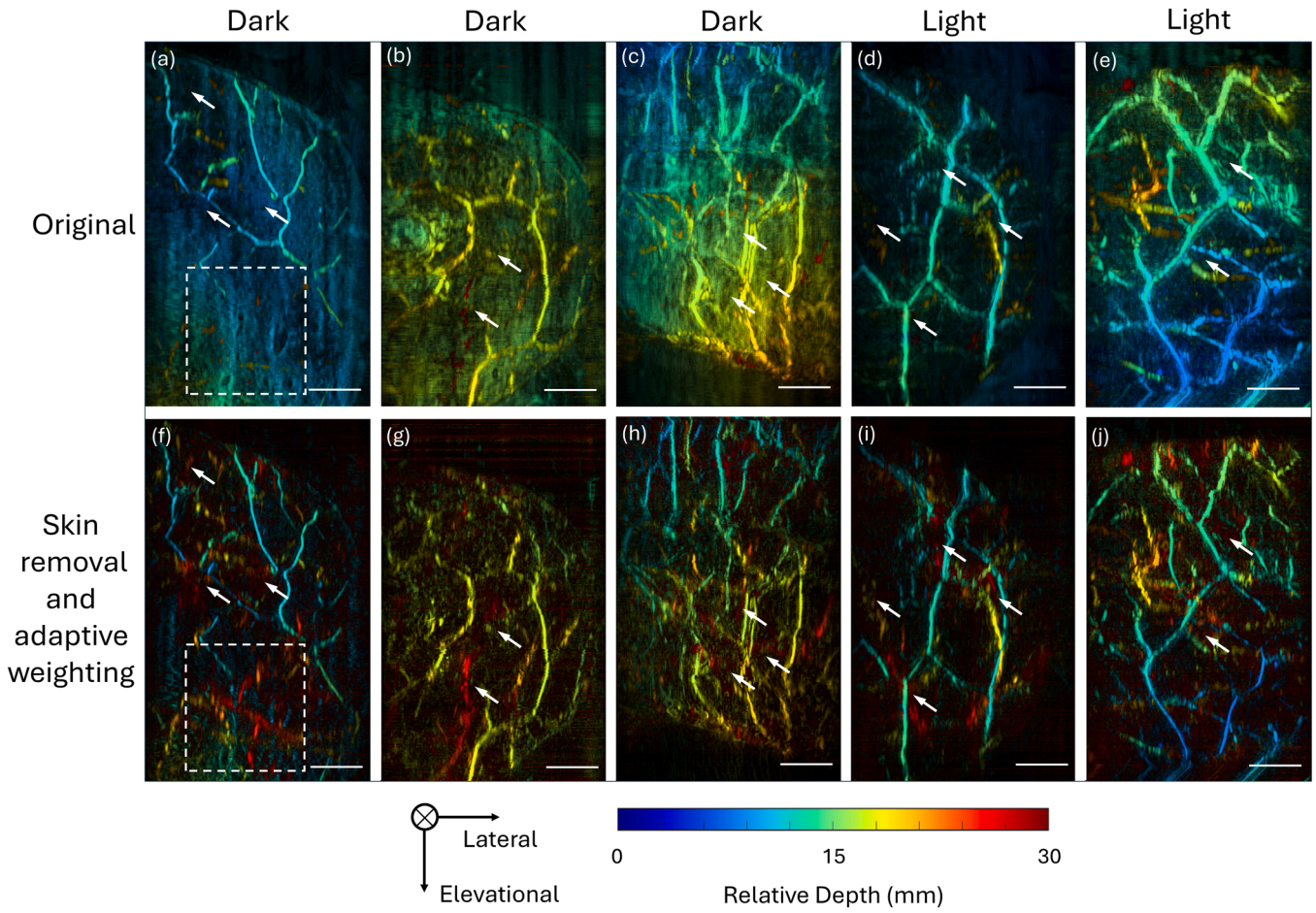


Fig. 10. Comparison of the depth-encoded MAP images before and after skin removal and adaptive weighting from different samples. Each column shows the MAP images from an individual sample: the top row represents MAP images of the original data, while the bottom one exhibits the results of processed data, and the skin tone of the sample is labeled at the top of each image pair. The white rectangle and arrows marked the location of deep vessels that were invisible in the original image. Scale bar:20 mm. The elevation and lateral axes indicate the scanning direction of the ultrasound transducer array.

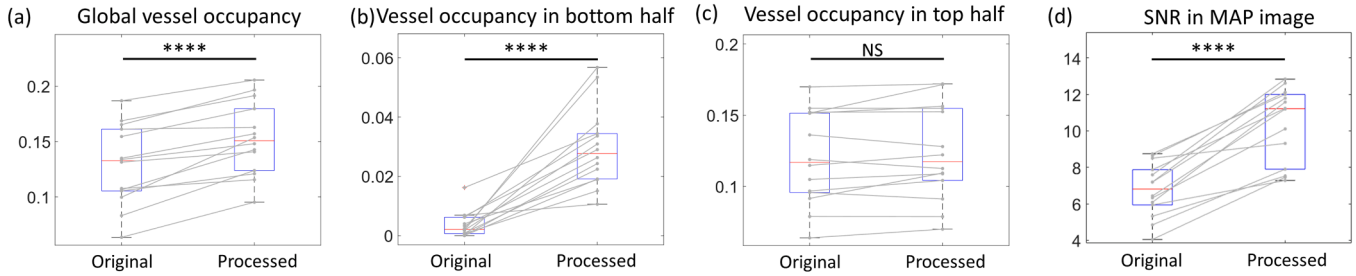


Fig. 11. Boxplots comparing the vessel occupancy across different regions and signal-to-noise ratio (SNR) of Maximum amplitude projection (MAP) images before and after processing. We quantified the significance of the improvement by the two-tailed paired *t*-test. In the context of the statistical significance, the notation is as follows: * ** * denotes $p < 0.0001$, NS denotes not significant ($p \geq 0.05$).

the SL-Net to the acquired PA images to identify the skin layer's location, and the corresponding skin-removed MAP images are displayed in Fig. 6 (k)-(n). After skin removal, signals from thin-layer skin phantoms were effectively suppressed, and the pencil leads were recovered in the MAP images across all phantoms. For reference, MAP images of the tissue-mimicking phantom alone are shown in Fig. 6(o)-(r). By comparing the processed images to the reference images, it is evident that the skin-removed images retain a similar shape and structure to the reference images, demonstrating that the proposed method effectively suppresses signals from the skin phantoms and recovers the pencil leads. Additionally, the processed images showed consistent performance across various ink concentrations, indicating that SL-Net can accurately locate skin layers with different skin tones.

To further evaluate the method, we quantified the Multi-Scale Structural Similarity Index (MS-SSIM) and Peak Signal-to-Noise Ratio (PSNR) in both the original and processed images from phantoms with varying ink concentrations. The quantification results are plotted in Fig. 6(s), with detailed values provided in Appendix A Table S9. MS-SSIM is represented by the blue solid line, while PSNR is shown by the orange dashed line. Circle markers indicate values for the original MAP images, and plus-shaped markers represent values for the processed images. Consistent with the visual observations, MS-SSIM and PSNR

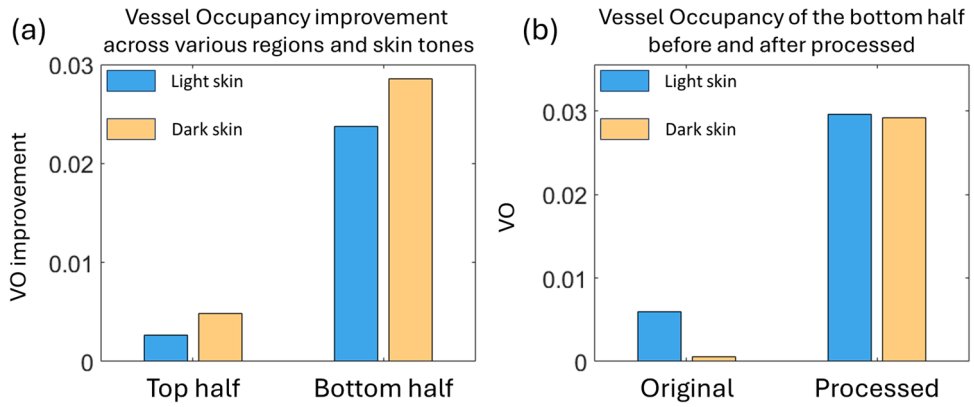


Fig. 12. Comparison of vessel occupancy index between light and dark skin tones. (a) The improvement of vessel occupancy in the top half and bottom half regions after processing with the proposed method. (b) Vessel occupancy in the bottom half region before and after processing.

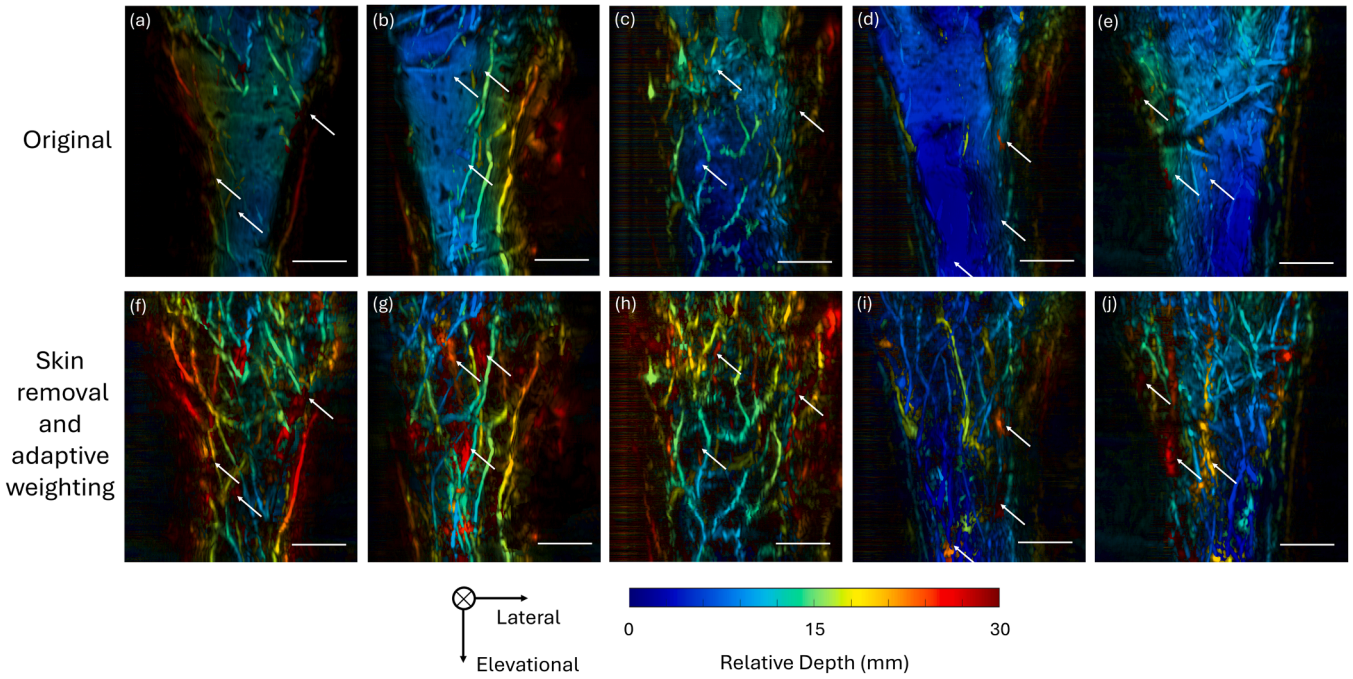


Fig. 13. Comparison of depth-encoded foot MAP images before and after adaptive weighting across five distinct samples with dark skin tone. The top row displays the MAP derived from the original images, while the bottom row shows the MAP after skin removal and adaptive weighting. Scale bar: 20 mm.

values in the original images decreased as the ink concentration increased in original images. However, after processing with SL-Net, both metrics improved significantly. Notably, the MS-SSIM and PSNR values for the 0.20 % ink phantom closely matched those from the 0 % ink phantom, demonstrating the proposed method's ability to mitigate the impact of skin tones in PA imaging.

In summary, the phantom experiments demonstrate that the proposed SL-Net effectively identifies skin layers and suppresses their signals, recovering deep structures such as pencil leads. The method performs robustly across varying skin tone conditions, showing strong potential for implementation in in-vivo experiments.

3.4. Evaluation of in-vivo image enhancement based on SL-Net

3.4.1. Breast imaging

In this study, we collected 14 samples from breast imaging to evaluate the performance of the proposed method. The samples are processed with the proposed skin removal and adaptive weighting algorithm. The selected subjects encompass a diverse range of skin tones, and we classified them into dark (6 samples) and light skin (8 samples) for further analysis. The evaluation starts with investigations on comparing MAP images and slices of one sample from multiple perspectives, following with case studies where outcomes from various processing methods are compared. Subsequently, we quantify the vessel occupancy to assess the efficiency of adaptive weighting.

To validate the performance of the proposed method, we analyzed original and processed PA images from the same subject. In Fig. 7, the original MAP images and slice views from multiple directions are presented in the left orange box, while images processed with skin removal and adaptive weighting are shown in the right red box. All images are depth-encoded along the axial direction (z-axis). Fig. 7(a) displays the original MAP image in elevational-lateral plane (x-y plane), where deep vessels are barely visible. The corresponding MAP projections to the lateral-axial plane (y-z plane) and elevational-axial plane (x-z plane) are shown in Fig. 7(b) and (c), respectively. To further explore the details, we examined slices along the green and magenta dashed lines in Fig. 7 (a), as shown in Fig. 7(d) and (e). After processing by the proposed algorithm, the resulting MAP images in the y-z plane, x-y plane, x-z plane

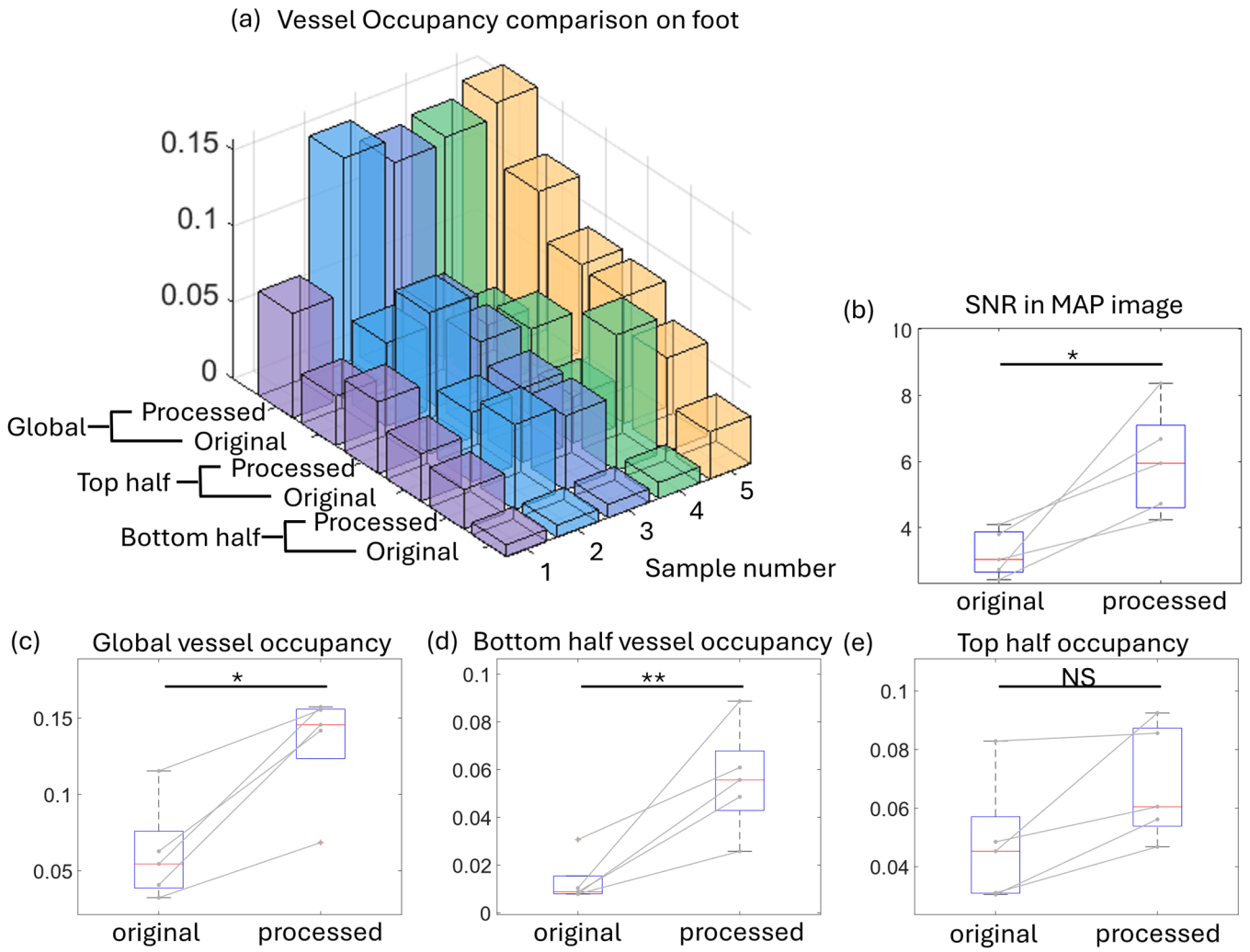


Fig. 14. Comparison of foot vessel occupancy under various conditions. (a) The quantified vessel occupancy before and after processing across three regions in five samples. (b) The quantified signal-to-noise ratio (SNR) in MAP images before and after processing (c)-(e) The vessel occupancy before and after processing in global, top half, and bottom half regions, respectively. The statistical significance is estimated with the two-tailed paired *t*-test and denoted as: * for $p < 0.05$, * * for $p < 0.01$, and NS for not significant ($p \geq 0.05$).

are presented in Fig. 7(f), (g) and (i), respectively. Similarly, slices from the same locations are shown in Fig. 7(h) and (j).

After processing, the MAP images in Fig. 7(g) clearly show that more deep vessels are recovered, while skin signals are significantly suppressed. The processed MAP images across cross-sectional views (Fig. 7 (f) and (i)) further demonstrate that vessels in deep regions become more apparent. Moreover, this observation is validated by slices in Fig. 7 (h) and(j), showing improved visibility of deep vessels compared to the original slices (Fig. 7(d) and (e)). For better illustration, deep vessels recovered by the proposed method are highlighted by white dashed rectangles, and the suppression of skin signals is particularly noticeable in the yellow box area.

Case study 1:

This case study conducts a comparative analysis between MAP images enhanced by the proposed weighting method and the generic weighting method [5]. The image was from an individual with black skin color. Based on the clinical report, the suspected location of the tumors is at a 2 cm distance in the 4 o'clock direction from the nipple. The tumor region is marked by a white dashed rectangle in Fig. 8, with further details available in Appendix A Fig. S1.

The comparison between original and processed images by various weighting images is depicted in Fig. 8. Fig. 8(a) illustrates the original MAP image from the subject, while Fig. 8(b) shows the MAP image from skin-removed data before any weighting. It can be seen that, after skin removal, more vessels can be revealed in Fig. 8(b). However, the residual skin signals still suppressed some vessel signals. After skin removal, generic and adaptive weighting is applied to further enhance the image, and results are shown in Fig. 8(c) and (d), respectively. While both weighting methods enabled the recovery of deep vessels, the adaptive weighting method performs better. For instance, Fig. 8(d) shows the most detailed contour of Vessel 1 among all processed images, and Vessel 2 and Vessel 3 are hard to distinguish in Fig. 8(c) while recovered in Fig. 8(d).

Moreover, the adaptive weighting approach revealed the suspected tumor region (white dashed rectangle, shown as a cluster of vessel signals) within the MAP image, demonstrating its potential for tumor localization. Consequently, the proposed adaptive weighting method based on the SL-Net exhibits distinct advantages in recovering deep vessels in PA images.

Case study 2:

We evaluated the efficacy of both generic weighting and adaptive weighting methods on another subject with black/brown skin tone. According to the clinical report, linear and branching fine pleomorphic microcalcifications spanning about 4 cm were observed in the subject's right breast at the 12 o'clock position from the nipple (the suspicious region is highlighted by a white dashed rectangle in Fig. 9). The imaging

window is set directly above the nipple, excluding the nipple itself, to align with the microcalcification region.

The original MAP image from PACT is illustrated in Fig. 9(a). We employed SL-Net to segment the skin layers and generate skin-removed MAP images shown in Fig. 9(b). Although signals from skin layers have been suppressed after processing, deep vessels close to the microcalcification region are not obvious Therefore, both generic weighting and adaptive weighting techniques were applied to enhance the visibility of deep vessels after skin removal, and the results are presented in Fig. 9(c) and (d), respectively. Upon comparing Figs. 9(c) and 9(d), it is evident that adaptive weighting successfully recovers more deep vessels. Notably, these retrieved vessels are located proximal to the suspicious region, indicating they are potentially related to the health condition of this region and this information might benefit further analysis. Therefore, the adaptive weighting outperforms generic weighting in deep vessel recovery in this case.

Following validation in two case studies with dark skin, we implemented the adaptive weighting to multiple samples with different skin colors. Fig. 10 shows depth-encoded MAP images of several subjects, with the top row displaying original MAP images and the bottom row depicting MAP images with skin removal and adaptive weighting, the skin tone of the sample is marked at the top of each image pair. Within these images, the white arrows highlight the restored deep regions that were invisible in the original image. Comparing the original and processed images, we can see that the skin layers are mostly removed, and vessels are restored, especially vessels in deep regions (colored in red). We also noticed that there are some residual skin signals (marked by a white dashed rectangle), however, the intensity of these signals is significantly suppressed, making little impact on further analysis. Details regarding the samples utilized in this study can be found in Appendix A Table S1.

To evaluate the efficiency of the enhancement, we quantified the vessel occupancy across various regions: global, top half, and bottom half. In addition, we employed the paired *t*-test to assess the significance level. Quantification results are attached as Table S2 and S3 in Appendix A. Fig. 11 illustrates boxplots comparing vessel occupancy between original and processed images, with the gray line tracking individual sample changes in Fig. 11(a)-(c). Fig. 11(a) presents a considerable increase in vessel occupancy at the global level after processing. Fig. 11(b) exhibits a noteworthy increase in vessel occupancy within deep regions, demonstrating the method's efficacy in recovering deep vessels. Conversely, Fig. 11(c) indicates marginal changes in vessel occupancy in the top half region, which is expected due to the minimal impact of attenuation on shallow vessels. The signal-to-noise ratio (SNR) of MAP image was also calculated to quantify the performance of the proposed method. After skin-removal and adaptive weighting, the SNR was improved by 55.20 % on average compared to original images, as shown in Fig. 11(d). Consequently, the proposed method effectively restores vessels in deep regions while preserving vessel features in shallower areas.

We further classified the sample into two categories (dark and light skin tones) to investigate the performance of the proposed method. Considering that dark skin absorbs more optical energy, the PA signal from skin layers is higher, while signals from deep vessels are weaker due to less transmitted optical intensity. Therefore, we compared the improvement of vessel occupancy quantified in the top and bottom half of tissue regions across various skin tones. The results are shown in Fig. 12 (a). Regardless of skin tones, increased vessel occupancy is observed across both shallow and deep regions after processing, especially in the bottom half regions, demonstrating the method's ability to recover deep vessels. In terms of skin tone, adaptive weighting demonstrates more significant improvement on dark skin samples, as the strong skin signals are removed after processing. In addition, we investigated the vessel occupancy index in the bottom half region, and the results are shown in Fig. 12(b). It can be seen that, after processing, the vessel occupancy index in dark skin is comparable to that of light skin. These results further confirm that the proposed method can effectively recover vessels from the deep region.

3.4.2. *Foot imaging*

In addition to breast imaging, we also implemented adaptive weighting to foot imaging to recover deep vessels. The PA signals from deep regions were attenuated and might be blocked by the dark skin, which impedes further feature extraction and analysis. Furthermore, the shape of the dorsal side of the foot is more complex than that of the breast, making it challenging to segment skin layers with manual operation, according to our earlier investigation [11]. In this study, we evaluate the performance of the proposed method on five samples collected from patients. The MAP images from subjects are demonstrated in Fig. 13. The top row shows the depth-encoded MAP from the original images, where the strong skin signals suppress the vessel signals underneath. The MAP images after skin removal and adaptive weighting are presented in the bottom row, where most of the skin signals are suppressed, and the deep vessels are recovered. The comparison of the performance between the generic weighting and proposed adaptive weighting is presented in Appendix A Fig. S2.

The quantitative evaluation involved extracting vessel occupancy from MAP images using the same method as breast imaging. Fig. 14(a) provides a comparative analysis of vessel occupancy improvement with adaptive weighting across five samples in diverse regions: global, top half, and bottom half. The outcomes indicate that there is a significant enhancement in revealing more vessels in deep regions, thereby facilitating quantitative vessel feature extraction and analysis. Meanwhile, we implemented SNR to quantify the improvement. As shown in Fig. 14 (b), SNR increases 92.42 % on average. Additionally, the paired *t*-test's p-values shown in Fig. 14(c)-(e) confirm the method's efficacy in uncovering more vessels. In addition to deep vessel recovery, we also noticed a significant improvement in shallow regions as shown in Fig. 14 (e). Considering vessels on the foot are closer to skin layers compared to the breast, the proposed method might be more significant in recovering vessels in foot imaging.

4. Discussion

In this study, we proposed a new approach to improve vascular features under the skin. Our method starts with building and training a deep learning model named SL-Net to locate skin layers in cross-section, and then an adaptive weighting method was implemented based on skin locations predicted by SL-Net. The efficacy of the workflow was validated on the human image results from the breast and foot. We quantified the enhancement by vessel occupancy metrics, and the results showed that the algorithm significantly recovers deep vessels across various body positions and skin tones.

Since there is no widely adopted open-source clinical PA imaging dataset [23], we constructed a customized in-vivo dataset with over 5000 frames featuring pixel-level skin labels obtained through manual labeling for training purposes. All data used in this research originated from human imaging experiments from breast and foot imaging. The dataset was acquired using a 2.25 MHz linear-array transducer, offering enhanced penetration depth and a wider field of view. Utilizing clinical PA images circumvents potential discrepancies between simulation data and in-vivo clinical system data, thus enhancing the efficiency of model training. A 2D residual UNet [39] was then trained and tested for skin segmentation. The model uses 2D frames as input, making it lightweight and easy to train. The residual UNet combines the strengths of ResNet and UNet, leveraging the feature learning capabilities of ResNet and the excellent localization abilities of UNet [39]. This architecture is well-suited for skin segmentation tasks where both local features and global context information are essential. In addition, the residual connections in the model can help in addressing class imbalance issues by allowing gradients to bypass problematic layers during back-propagation. In addition, the skip connections assist in gradient flow,

making the training more stable, especially when dealing with limited annotated data, which is common in medical imaging. The developed adaptive weighting mechanism in this study incorporates optical and acoustic attenuations, and both can be approximated as exponential delays over imaging depth. An automatic algorithm is developed to estimate the weighting, eliminating human intervention. Compared to generic weighting, the adaptive weighting factor is only applied to tissues underneath the skin. This is particularly useful for cases with dark and curved skin layers as demonstrated in Fig. 8 and Fig. 13.

For evaluation, the proposed weighting was tested on breast and foot imaging data. Images presented in the results section demonstrate substantial enhancement of deep vessels. We selected vessel occupancy as a quantifiable feature to assess the impact of weighting, representing the ratio of vessel area to the regions of interest. Significant enhancements were observed across all samples and deep regions, notably in subjects with darker skin tones. Results from the *t*-test prove the adaptive weighting's efficacy in retrieving deep vessels. In addition, the recovery of deep vessels by adaptive weighting facilitates a more precise quantification of vessels, since these features are heavily dependent on the precise presentation of vessels.

However, there are some limitations to the proposed methodology. For training the SL-Net, the quantity and quality of ground truth datasets are limited by the time-consuming nature of manual labeling. For instance, despite striving for high-quality pixel-level annotations, limitations in resolution and inherent imaging system noise impede precise labeling. Consequently, the model's predicted labels might not encompass the entire skin region, leading to residual skin signals in MAP images, as shown in Fig. 10. The datasets can be enlarged with PA images from multiple body origins and cross-annotated by multiple readers in the future work. In addition, the adaptive weighting method relies on two assumptions: uniform optical fluence at the skin surface and exponential light attenuation in tissue. While these assumptions facilitate attenuation modeling, their clinical applicability may vary. For instance, complex tissue structures might lead to non-exponential light attenuation [52], which might explain why the performance of the weighting method is less effective in breast images. In the future, more accurate acoustic and optical models can be developed to accommodate such variations.

5. Conclusion

In this study, we trained a deep learning network SL-Net for skin localization and used the predicted labels to guide an adaptive weighting method to enhance the deep vessels. The improvement of the proposed framework has been tested and evaluated on both human breast and foot imaging results. We also quantitatively assessed the method's performance by calculating the vessel occupancy indexes before and after processing. The result demonstrated that adaptive weighting can significantly recover the deep vessels, indicating great potential in clinical and preclinical studies.

CRediT authorship contribution statement

Huang Chuqin: Writing – original draft, Methodology, Investigation, Conceptualization. **Zheng Emily**: Investigation, Data curation. **Zheng Wenhan**: Methodology, Conceptualization. **Zhang Huijuan**: Investigation. **Cheng Yanda**: Investigation, Data curation. **Zhang Xiaoyu**: Software. **Shijo Varun**: Visualization, Software. **Bing Robert W.**: Visualization. **Komornicki Isabel**: Resources. **Harris Linda M.**: Resources. **Bonaccio Ermelinda**: Resources. **Takabe Kazuaki**: Resources. **Zhang Emma**: Data curation. **Xu Wenyao**: Supervision, Funding acquisition. **Xia Jun**: Writing – review & editing, Supervision, Project administration, Funding acquisition, Conceptualization.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the National Institute of Health [Grants R01EB029596, R01EB028978, and R01EB035188].

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.pacs.2025.100690.

Data availability

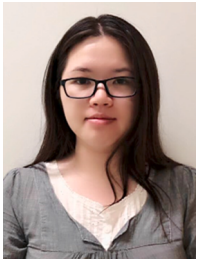
Data will be made available on request.

References

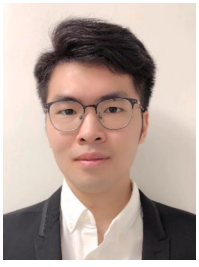
[1] J. Xia, J. Yao, L.V. Wang, Photoacoustic tomography: principles and advances, Electro Waves 147 (2014) 1–22.
[2] L.V. Wang, S. Hu, Photoacoustic tomography: in vivo imaging from organelles to organs, Science 335 (6075) (2012) 1458–1462.
[3] L.V. Wang, H.-i Wu, Biomedical optics: principles and imaging, John Wiley & Sons, 2012.
[4] E. Zheng, H. Zhang, W. Hu, M.M. Doyley, J. Xia, Volumetric tri-modal imaging with combined photoacoustic, ultrasound, and shear wave elastography, J. Appl. Phys. 132 (3) (2022).
[5] E. Zheng, H. Zhang, S. Goswami, I.E. Kabir, M.M. Doyley, J. Xia, Second-generation dual scan mammoscope with photoacoustic, ultrasound, and elastographic imaging capabilities, Front. Oncol. 11 (2021) 779071.
[6] N. Nyayapathi, R. Lim, H. Zhang, W. Zheng, Y. Wang, M. Tiao, K.W. Oh, X.C. Fan, E. Bonaccio, K. Takabe, Dual scan mammoscope (DSM)—a new portable photoacoustic breast imaging system with scanning in craniocaudal plane, IEEE Trans. Biomed. Eng. 67 (5) (2019) 1321–1327.
[7] L. Lin, P. Hu, J. Shi, C.M. Appleton, K. Maslov, L. Li, R. Zhang, L.V. Wang, Single-breath-hold photoacoustic computed tomography of the breast, Nat. Commun. 9 (1) (2018) 2352.
[8] L. Lin, X. Tong, P. Hu, M. Invernizzi, L. Lai, L.V. Wang, Photoacoustic computed tomography of breast cancer in response to neoadjuvant chemotherapy, Adv. Sci. 8 (7) (2021) 2003396.
[9] W. Choi, E.-Y. Park, S. Jeon, Y. Yang, B. Park, J. Ahn, S. Cho, C. Lee, D.-K. Seo, J.-H. Cho, Three-dimensional multistructural quantitative photoacoustic and US imaging of human feet in vivo, Radiology 303 (2) (2022) 467–473.
[10] Y. Wang, Y. Zhan, L.M. Harris, S. Khan, J. Xia, A portable three-dimensional photoacoustic tomography system for imaging of chronic foot ulcers, Quant. Imaging Med. Surg. 9 (5) (2019) 799.
[11] C. Huang, Y. Cheng, W. Zheng, R.W. Bing, H. Zhang, I. Komornicki, L.M. Harris, P. R. Arany, S. Chakraborty, Q. Zhou, Dual-scan photoacoustic tomography for the imaging of vascular structure on foot, IEEE Trans. Ultrason., Ferroelectr., Freq. Control (2023).
[12] W. Zheng, D. Lee, J. Xia, Photoacoustic tomography of fingerprint and underlying vasculature for improved biometric identification, Sci. Rep. 11 (1) (2021) 17536.
[13] Z. Chen, E. Rank, K.M. Meiburger, C. Sinz, A. Hodul, E. Zhang, E. Hoover, M. Minneman, J. Ensher, P.C. Beard, Non-invasive multimodal optical coherence and photoacoustic tomography for human skin imaging, Sci. Rep. 7 (1) (2017) 17975.
[14] G.S.P. Fernandes, J.H. Uliana, L. Bachmann, A.A. Carneiro, M.A.L. Bell, T.Z. Pavan, Impact of skin pigmentation on photoacoustic imaging using linear array transducer: a pilot in vivo study, 2022 IEEE International Ultrasonics Symposium (IUS), IEEE, 2022, pp. 1-4.
[15] M. Sun, C. Li, N. Chen, H. Zhao, L. Ma, C. Liu, Y. Shen, R. Lin, X. Gong, Full three-dimensional segmentation and quantification of tumor vessels for photoacoustic images, Photoacoustics 20 (2020) 100212.
[16] H.-C. Zhou, N. Chen, H. Zhao, T. Yin, J. Zhang, W. Zheng, L. Song, C. Liu, R. Zheng, Optical-resolution photoacoustic microscopy for monitoring vascular normalization during anti-angiogenic therapy, Photoacoustics 15 (2019) 100143.
[17] Y. Mantri, J.V. Jokerst, Impact of skin tone on photoacoustic oximetry and tools to minimize bias, Biomed. Opt. Express 13 (2) (2022) 875–887.
[18] G.S. Fernandes, J.H. Uliana, L. Bachmann, A.A. Carneiro, M.A.L. Bell, T.Z. Pavan, Mitigating skin tone bias in linear array in vivo photoacoustic imaging with short-lag spatial coherence beamforming, Photoacoustics 33 (2023) 100555.
[19] T.A. Miller, J.H. Campbell, N. Bloom, S.R. Wurdeman, Racial disparities in health care with timing to amputation following diabetic foot ulcer, Diabetes Care 45 (10) (2022) 2336–2341.

[20] H.F. Zhang, K. Maslov, L.V. Wang, Automatic algorithm for skin profile detection in photoacoustic microscopy, J. Biomed. Opt. 14 (2) (2009), 024050-024050-6.
[21] J. Kim, J.Y. Kim, S. Jeon, J.W. Baik, S.H. Cho, C. Kim, Super-resolution localization photoacoustic microscopy using intrinsic red blood cells as contrast absorbers, Light.: Sci. Appl. 8 (1) (2019) 103.
[22] C. Lee, S. Cho, D. Lee, J. Lee, J.-I. Park, H.-J. Kim, S.H. Park, W. Choi, U. Kim, C. Kim, Panoramic volumetric clinical handheld photoacoustic and ultrasound imaging, Photoacoustics 31 (2023) 100512.
[23] H. Zhang, W. Bo, D. Wang, A. DiSpirito, C. Huang, N. Nyayapathi, E. Zheng, T. Vu, Y. Gong, J. Yao, Deep-E: A fully-dense neural network for improving the elevation resolution in linear-array-based photoacoustic tomography, IEEE Trans. Med. Imaging 41 (5) (2021) 1279–1288.
[24] C. Yang, H. Lan, F. Gao, F. Gao, Review of deep learning for photoacoustic imaging, Photoacoustics 21 (2021) 100215.
[25] J. Gröhl, M. Schellenberg, K. Dreher, L. Maier-Hein, Deep learning for biomedical photoacoustic imaging: A review, Photoacoustics 22 (2021) 100241.
[26] Y. Cheng, W. Zheng, R. Bing, H. Zhang, C. Huang, P. Huang, L. Ying, J. Xia, Unsupervised denoising of photoacoustic images based on the Noise2Noise network, Biomed. Opt. Express 15 (8) (2024) 4390–4405.
[27] C.D. Ly, T.H. Vo, S. Mondal, S. Park, J. Choi, T.T.H. Vu, C.-S. Kim, J. Oh, Full-view in vivo skin and blood vessels profile segmentation in photoacoustic imaging based on deep learning, Photoacoustics 25 (2022) 100310.
[28] A.Y. Yuan, Y. Gao, L. Peng, L. Zhou, J. Liu, S. Zhu, W. Song, Hybrid deep learning network for vascular segmentation in photoacoustic imaging, Biomed. Opt. Express 11 (11) (2020) 6445–6457.
[29] H. Zhao, J. Huang, Q. Zhou, N. Chen, L. Liu, X. Wang, T. Wang, L. Chen, C. Liu, C. Zheng, Deep learning-based optical-resolution photoacoustic microscopy for in vivo 3d microvasculature imaging and segmentation, Adv. Intell. Syst. 4 (9) (2022) 2200004.
[30] M. Schellenberg, K.K. Dreher, N. Holzwarth, F. Isensee, A. Reinke, N. Schreck, A. Seitel, M.D. Tizabi, L. Maier-Hein, J. Gröhl, Semantic segmentation of multispectral photoacoustic images using deep learning, Photoacoustics 26 (2022) 100341.
[31] W. Zheng, H. Zhang, C. Huang, V. Shijo, C. Xu, W. Xu, J. Xia, Deep learning enhanced volumetric photoacoustic imaging of vasculature in human, Adv. Sci. (2023) 2301277.
[32] N. Wang, T. Chen, C. Liu, J. Meng, Intelligent skin-removal photoacoustic computed tomography for human based on deep learning, Journal of Biophotonics e202400197.
[33] A. Jemal, F. Bray, M.M. Center, J. Ferlay, E. Ward, D. Forman, Global cancer statistics, CA: a Cancer J. Clin. 61 (2) (2011) 69–90.
[34] H. Sung, J. Ferlay, R.L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, F. Bray, Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries, CA: a Cancer J. Clin. 71 (3) (2021) 209–249.
[35] N. Nyayapathi, J. Xia, Photoacoustic imaging of breast cancer: a mini review of system design and image features, J. Biomed. Opt. 24 (12) (2019), 121911-121911.
[36] L. Lin, L.V. Wang, The emerging role of photoacoustic imaging in clinical oncology, Nat. Rev. Clin. Oncol. 19 (6) (2022) 365–384.
[37] E.J. Boyko, M. Monteiro-Soares, S.G. Wheeler, Peripheral arterial disease, foot ulcers, lower extremity amputations, and diabetes, Diabetes in America. 3rd edition (2018).
[38] M. Xu, L.V. Wang, Universal back-projection algorithm for photoacoustic computed tomography, Phys. Rev. E—Stat., Nonlinear, Soft Matter Phys. 71 (1) (2005) 016706.
[39] Z. Zhang, Q. Liu, Y. Wang, Road extraction by deep residual u-net, IEEE Geosci. Remote Sens. Lett. 15 (5) (2018) 749–753.
[40] M.J. Cardoso, W. Li, R. Brown, N. Ma, E. Kerfoot, Y. Wang, B. Murrey, A. Myronenko, C. Zhao, D. Yang, Monai: An open-source framework for deep learning in healthcare, arXiv preprint arXiv:2211.02701 (2022).
[41] F. Milletari, N. Navab, S.-A. Ahmadi, V-net: Fully convolutional neural networks for volumetric medical image segmentation, 2016 fourth international conference on 3D vision (3DV), Ieee, 2016, pp. 565-571.
[42] I. Loshchilov, F. Hutter, Sgdr: Stochastic gradient descent with warm restarts, arXiv preprint arXiv:1608.03983 (2016).
[43] L. Kocsis, P. Herman, A. Eke, The modified Beer–Lambert law revisited, Phys. Med. Biol. 51 (5) (2006) N91.
[44] M. Xu, L.V. Wang, Photoacoustic imaging in biomedicine, Rev. Sci. Instrum. 77 (4) (2006).
[45] M.A. Mastanduno, S.S. Gambhir, Quantitative photoacoustic image reconstruction improves accuracy in deep tissue structures, Biomed. Opt. Express 7 (10) (2016) 3811–3825.
[46] S. Park, A.A. Oraevsky, R. Su, M.A. Anastasio, Compensation for non-uniform illumination and optical fluence attenuation in three-dimensional optoacoustic tomography of the breast. Photons Plus Ultrasound: Imaging and Sensing 2019, SPIE, 2019, pp. 388–395.
[47] N. Nyayapathi, H. Zhang, E. Zheng, S. Nagarajan, E. Bonaccio, K. Takabe, X.C. Fan, J. Xia, Photoacoustic dual-scan mammoscope: results from 38 patients, Biomed. Opt. Express 12 (4) (2021) 2054–2063.
[48] A.F. Frangi, W.J. Niessen, K.L. Vincken, M.A. Viergever, Multiscale vessel enhancement filtering. International conference on medical image computing and computer-assisted intervention, Springer, 1998, pp. 130–137.
[49] K. Zuiderveld, Contrast limited adaptive histogram equalization, Graph. gems (1994) 474–485.
[50] T. Ridler, S. Calvard, Picture thresholding using an iterative selection method, IEEE Trans. Syst. Man Cyber 8 (8) (1978) 630–632.
[51] K.M. Kempski, M.T. Graham, M.R. Gubbi, T. Palmer, M.A. Lediju Bell, Application of the generalized contrast-to-noise ratio to assess photoacoustic image quality, Biomed. Opt. Express 11 (7) (2020) 3684–3698.
[52] A.B.E. Attia, G. Balasundaram, M. Moothanchery, U. Dinish, R. Bi, V. Ntziachristos, M. Olivo, A review of clinical photoacoustic imaging: Current and future trends, Photoacoustics 16 (2019) 100144.

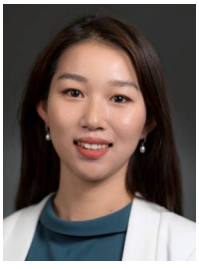
Chuqin Huang is a Ph.D. student in the biomedical Engineering department at University at Buffalo, the State University of New York. He received the B.E. degree in Mechanical Engineering from Huazhong University of Science and Technology in 2018, and the M.S. degree in Mechanical Engineering from University of Washington in 2020. His research includes photoacoustic tomography and its application on foot imaging.



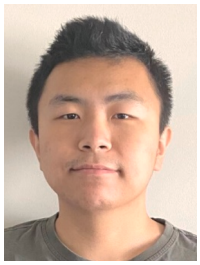
Emily Zheng is a Ph.D. student at the University at Buffalo, the State University of New York. She completed her bachelor's degree in Biomedical Engineering before pursuing her doctoral studies in the same department. Her research primarily focuses on advancing biomedical imaging techniques including photoacoustic and ultrasonic modalities and their clinical applications. Her research involves developing a novel imaging system for breast cancer detection and employing deep-learning models for tumor localization. Additionally, she explores the adaptation of acoustic-based imaging technologies such as ultrasonic elastography and MV-doppler imaging.



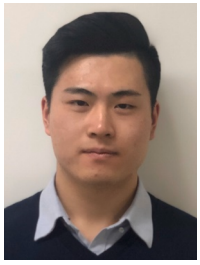
Wenhan Zheng is a Ph.D. student in the Biomedical Engineering Department at the University at Buffalo, the State University of New York. Mr. Zheng has over ten publications in peer-reviewed journals, specifically focusing on the field of photoacoustics. Furthermore, he has acquired substantial industry experience by working with leading global companies in his field.



Dr. Huijuan Zhang, Ph.D., is currently a postdoctoral fellow at Wellman Center for Photomedicine, Massachusetts General Hospital, and Harvard Medical School. She received her doctorate in Biomedical Engineering from University at Buffalo in 2023. Her research interests include photoacoustic imaging, deep learning, and optical coherence imaging. She has published more than twenty papers in peer-reviewed journals on medical imaging research.



Yanda Cheng is a Ph.D. student in the Biomedical Engineering Department at the University at Buffalo, the State University of New York. He received his Bachelor of Science degree in Electrical Engineering from the University of Kentucky in 2020 and received his Master of Science degree in Biomedical Engineering from Cornell University in 2021. Yanda's research focuses on photoacoustic imaging, deep learning, machine learning, and image processing.



Xiaoyu Zhang received the engineering degree from Hefei University of Technology, China, in 2017, the Master degree in engineering from the University of Science and Technology of China, China, in 2020, and now he is pursuing the Ph.D. degree in computer science from the State University of New York at Buffalo, United States. His current research interests include Wireless Sensing, Internet of Things (IoT) and Smart Health.

Varun Shijo is a 3rd-year PhD candidate in the Computer Science department at the University at Buffalo. His research focuses on applying Computer Vision and Deep Learning to medical imaging. Varun has previously worked in the industry applying, integrating, and maintaining DL and ML solutions for software products. He earned his MS in CSE from the University at Buffalo in 2019 and his BE in Information Technology from the University of Mumbai in 2017.

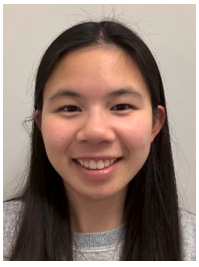
Robert Bing graduated in 2022 with a B.S. in Biomedical Engineering with a concentration in Nanotechnology from Boston University. He is currently working on his Ph.D. in Biomedical Engineering at University at Buffalo, the State University of New York. His research areas of interest include photoacoustic and ultrasonic sensing.

Isabel Komornicki, MSN, NP, Nurse Practitioner with the University at Buffalo Surgeons for the past 16 years in the department of surgery, vascular department. She received her Bachelors and Masters of science in Nursing at SUNY at Buffalo. She is primarily active in the outpatient clinical settings; treating vascular patients in all phases of disease. She has also participated in numerous research projects throughout her career, involving medical devices, pharmaceuticals and novel immunotherapies related vascular and connective tissue disorders. Currently, she is a research coordinator with the SUNY at Buffalo Research Foundation, in addition to clinical practice.

Dr. Linda M Harris is a tenured Professor of Surgery in the Department of Surgery, Division of Vascular Surgery at the Jacobs School of Medicine and Biomedical Sciences, University at Buffalo. She is previous Chief of the Division of Vascular Surgery, and is currently the program director for the vascular training programs. She has been practicing in Buffalo for over 25 years since completing her general surgery and vascular surgery training at the University of Buffalo. Dr Harris has coedited 2 books, written over 20 book chapters, and published well over 100 peer reviewed publications on a wide variety of vascular diseases. She has served as President of the Eastern Vascular Society and the Association of Program Directors in Vascular Surgery. Her research interests include Peripheral arterial disease, including diagnostic assessment.

Dr. Ermelinda Bonaccio is the Chair of Diagnostic and Interventional Radiology and the Director or Breast Imaging at Roswell Park Comprehensive Cancer Center. She received her BA and MA from Harvard University and her MD from Harvard Medical School. She completed her residency in Diagnostic Radiology at NY Hospital/Cornell Medical Center in 1996 and a breast imaging fellowship at Memorial Sloan Kettering Cancer Center in 1997. She has been practicing Breast Imaging since then and has been at Roswell Park since 2004.

Kazuaki Takabe is the Professor, Alfiero Foundation Chair and Clinical Chief of Breast Surgery at Roswell Park Comprehensive Cancer Center at Buffalo, NY. He received his MD from Niigata University in 1992, PhD from Yokohama City University in 1999, and Postdoctoral training at the Salk Institute. He underwent General Surgery Residency in Niigata University Hospital, Japan, as well as at University of California San Diego. He is a US Board certified in General Surgeon since 2007. He further underwent Surgical Oncology Fellowship at Virginia Commonwealth University, where he was recruited as a faculty and raised his rank to full professor. He has joined Roswell Park on 2016 and his current research interests include translational research that has clinical impact.



Emma Zhang is a senior high school student at Williamsville North High School and a summer research assistant at the Department of Biomedical Engineering at University at Buffalo.

Wenyao Xu received a Ph.D. degree from the University of California, Los Angeles, CA, USA, in 2013. He received both an M.S. degree in 2008 and a B.S. degree in 2006 from Zhejiang University, China (both with honors). Currently, Dr. Xu is a Professor at the Computer Science and Engineering Department, the State University of New York (SUNY) at Buffalo, New York, NY, USA. He published over 230 technical papers, co-authored 2 books, and is named inventor on nine International and U.S. patents. His recent research foci include the Internet of Things, Smart Health, and Cyber-Security, and his work has received 11 Best Paper Awards in related research fields. Dr. Xu has served on the technical program committee of numerous conferences in the field of Smart Health, Mobile Computing and the Internet of Things, and has been a TPC co-chair of IEEE BSN 2018/2023 and IEEE/ACM CHASE 2022.

Dr. Jun Xia, Ph.D., is a Professor in the Biomedical Engineering Department at University at Buffalo, the State University of New York. He received his doctorate in Mechanical Engineering from the University of Toronto and postdoctoral training in Biomedical Engineering at Washington University in St. Louis. Dr. Xia has published three book chapters and more than one hundred papers in peer-reviewed journals on photoacoustic and optical research.