Deep-E: A Fully-Dense Neural Network for Improving the Elevation Resolution in Linear-Array-Based Photoacoustic Tomography

Huijuan Zhang, *Graduate Student Member, IEEE*, Wei Bo, Depeng Wang, Anthony DiSpirito III, *Graduate Student Member, IEEE*, Chuqin Huang, Nikhila Nyayapathi, Emily Zheng, Tri Vu, Yiyang Gong, Junjie Yao, *Member, IEEE*, Wenyao Xu, *Senior Member, IEEE*, and Jun Xia, *Member, IEEE*

***Abstract*—Linear-array-based photoacoustic tomography has shown broad applications in biomedical research and preclinical imaging. However, the elevational resolution of a linear array is fundamentally limited due to the weak cylindrical focus of the transducer element. While several methods have been proposed to address this issue, they have all handled the problem in a less time-efficient way. In this work, we propose to improve the elevational resolution of a linear array through Deep-E, a fully dense neural network based on U-net. Deep-E exhibits high computational efficiency by converting the three-dimensional problem into a two-dimension problem: it focused on training a model to enhance the resolution along elevational direction by only using the 2D slices in the axial and elevational plane and thereby reducing the computational burden in simulation and training. We demonstrated the efficacy of Deep-E using various datasets, including simulation, phantom, and human subject results. We found that Deep-E could improve elevational resolution by at least four times and recover the object's true size. We envision that Deep-E will have a significant impact in linear-array-based photoacoustic imaging studies by providing high-speed and high-resolution image enhancement.**

***Index Terms*—Breast imaging, convolutional neural network, deep learning, elevation resolution, linear transducer array, photoacoustic tomography, resolution enhancement.**

I. Introduction

PHOTOACOUSTIC (PA) computed tomography (PACT) is a hybrid biomedical imaging modality that combines the merits of high optical absorption contrast and high acoustic resolution [1]. In PACT, a short-pulsed laser provides excitation light absorbed by biomolecules such as hemoglobin, lipid, or melanin, which causes thermoelastic expansion and generates acoustic waves that propagate through the tissue. Ultrasound transducer arrays detect these acoustic waves and form an image of the optical absorber distribution. As a noninvasive imaging technique with deep penetration (>3cm), PACT is sensitive to endogenous contrasts, particularly hemoglobin, which plays an essential role in the functioning of biological tissue [2], [3]. Over the past few years, PACT has been demonstrated in various preclinical and clinical applications, including human breast scanning [4]–[6], functional imaging [7], cardiology imaging [8], biometrics [9], and small animal whole-body imaging [10]. Among different types of transducers, linear transducer arrays are widely used due to their handheld convenience. The piezoelectric elements of the linear transducer arrays are arranged in a line to form a planar field of view. Linear arrays have low manufacturing costs and can be conveniently integrated with light sources for photoacoustic imaging [11].

An intrinsic limitation for the linear array is the poor three-dimensional (3D) imaging performance. As shown in Fig. 1(a), the 3D resolution of a linear-array is defined along with lateral, axial, and elevational directions. The element pitch determines the lateral resolution, which typically equals one acoustic

Manuscript received October 27, 2021; revised December 7, 2021; accepted December 15, 2021. Date of publication December 20, 2021; date of current version May 2, 2022. This work was supported in part by the Susan G. Komen Foundation under Grant CCR17481211 and in part by the National Institute of Health under Grant R01EB029596 and Grant R01EB028978. *(Corresponding author: Jun Xia.)*

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Institution Review Board of University at Buffalo.

Huijuan Zhang, Chuqin Huang, Nikhila Nyayapathi, Emily Zheng, and Jun Xia are with the Department of Biomedical Engineering, University at Buffalo, State University of New York, Buffalo, NY 14260 USA (e-mail: huijuanz@buffalo.edu; chuqinhu@buffalo.edu; nikhilan@buffalo.edu; emilyzhe@buffalo.edu; junxia@buffalo.edu).

Wei Bo and Wenyao Xu are with the Department of Computer Science and Engineering, University at Buffalo, State University of New York, Buffalo, NY 14260 USA (e-mail: weibo@buffalo.edu; wenyaoxu@buffalo.edu).

Depeng Wang was with the Department of Biomedical Engineering, University at Buffalo, State University of New York, Buffalo, NY 14260 USA. He is now with the Department of Biomedical Engineering, Duke University, Durham, NC 27708 USA (e-mail: depeng.wang@duke.edu).

Anthony DiSpirito III, Tri Vu, Yiyang Gong, and Junjie Yao are with the Department of Biomedical Engineering, Duke University, Durham, NC 27708 USA (e-mail: anthony.dispirito@duke.edu; tri.vu@duke.edu; yiyang.gong@duke.edu; junjie.yao@duke.edu).

This article has supplementary downloadable material available at https://doi.org/10.1109/TMI.2021.3137060, provided by the authors.

Digital Object Identifier 10.1109/TMI.2021.3137060

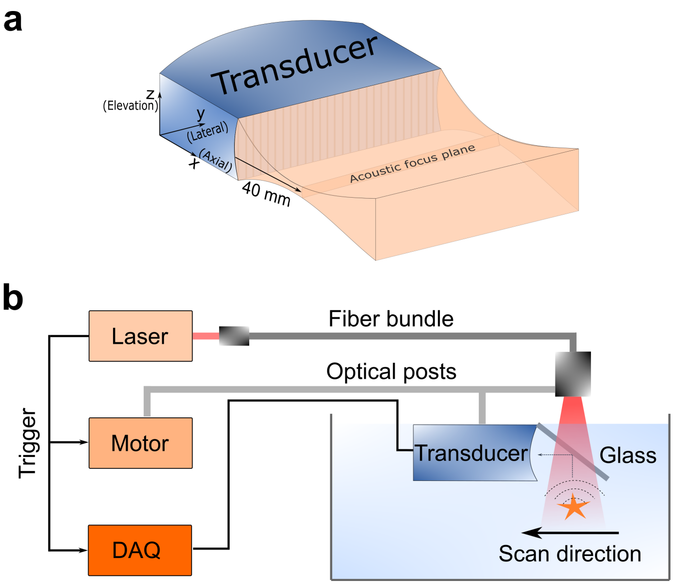

Fig. 1. A schematic of the linear transducer array and experimental setup. (a). Geometry configuration of the linear transducer array. The X-axis represents the direction of imaging depth. The acoustic focus of the transducer is at a 40 mm axial depth. The Y-axis represents the lateral direction, which is along the width of the transducer. The Z-axis represents the elevational direction that is along with the height of the transducer array. (b). Schematic of the experimental PACT imaging setup. The fiber output and linear ultrasound transducer array are immersed in the water tank for light illumination and signal detection, respectively. 3D imaging is achieved through a motorized translation stage. The laser synchronously triggers the motor for scanning and the DAQ for sampling.

wavelength at the central frequency. The axial direction is perpendicular to the transducer element, and its resolution typically equals half the acoustic wavelength at the central frequency. The elevational direction is perpendicular to the lateral and axial plane, and a linear array can be scanned along this direction to form a 3D image. However, the elevational resolution is lower due to the large transducer element height and its fixed cylindrical focus. In conventional 3D image formation, each 2D frame is reconstructed separately, and then these frames are stacked to form a 3D image [12]. Because each 2D reconstruction assumes that all photoacoustic signals come from the same imaging plane, the out-of-plane signals degrade the elevational resolution [13].

Various methods, either hardware or software-based, have been developed to improve the elevational resolution in a linear array-based PACT [11]. Before introducing our method, we will provide a brief overview of these approaches. Hardware-based approaches require modification on the imaging geometry or detection scheme. For instance, Gateau *et al.* proposed a scanning geometry that combined translational and rotational scanning to improve the elevational resolution [14]. Schwarz *et al.* investigated a bi-directional scanning geometry such that two linear scans were conducted perpendicular to each other to increase the elevational resolution [15]. These complicated methods increased the scanning time and generated more than twice the data for image reconstruction. Wang *et al.* proposed a detection hardware design based on acoustic diffraction through a thin slit that essentially increased the elevational receiving angle [16]. While this method showed noticeable improvement in elevational resolution, the slit reduced the acquired signal amplitude at each scanning position (though it could be recovered through 3D reconstruction). Software-based approaches use advanced image reconstruction algorithms to account for out-of-plane signals. Here the "plane" refers to the lateral-axial plane. We previously introduced a 3D focal line (3DFL) and a coherent weighted focal line method for 3D image reconstruction [13]. 3DFL reconstruction algorithm improves the elevational resolution by calculating the time of flight in 3D space. At the same time, coherent weighting assigns a weighting factor into the 3DFL reconstructed image to further improve the image contrast and resolution. However, this reconstruction algorithm requires back projection of raw data in 3D, which is time-consuming. Moreover, it is unclear whether the coherent weighting factor preserves the quantitative information of the PA signals.

In recent years, deep learning has been developing rapidly for photoacoustic imaging applications [17]–[21]. Various deep learning networks have been proposed to improve the photoacoustic image quality, degraded due to undersampling, limited-view, or limited bandwidth problems. Depending on the training model, these networks work on either the reconstructed image or raw channel data [22]–[25]. Applications in image segmentation and classification can also be found in PACT [26], [27]. These studies have proved that deep learning techniques are promising to improve photoacoustic imaging quality. To date, most studies focused on improving the lateral resolution, and very few studies reported the use of deep learning to improve the spatial resolution of 3D PACT [28], [29]. In particular, improving the elevational resolution of a linear array and validating the algorithm for human data remains unexplored.

Here, we propose a deep learning-based method to improve the elevational resolution in PACT. In this study, we provide an efficient simulation approach to generate low elevational resolution training data. Since the lateral and axial resolutions are more than twice the elevational direction, we convert the 3D problem into 2D (axial-elevational) and focus the training along the elevational direction. This simplification makes the simulation and training computationally more efficient. Our deep learning model, implemented based on a Fully-dense U-net (FD U-net), is named Deep-E. After validating with simulation, phantom, and human experimental results, we demonstrate that Deep-E provides at least four times improvement in elevational resolution, which is a significant enhancement. Most importantly, instead of simply shrinking the object to a point source, we also verify that Deep-E can recover the object's true size.

II. Methods

A. Network Architecture

In deep learning, we assume an approximate nonlinear relationship between the low-elevational resolution PACT image, X, and high-elevational resolution PACT image, Y, which can be represented in the form of a function F as shown in (1). In this function, θ are the parameters that learn to map X and Y during training. Mean square error (MSE) is used as the model loss function in (2), such that the loss error between the network predicted output, Y_{pred}, and the corresponding ground truth, Y_{true}, is minimized via supervised learning. Both MSE and mean absolute error (MAE) were somewhat comparable in terms of final network performance, though MSE seemed to have a smoother training process. As such, MSE was chosen over MAE as the model loss function.

$$Y = F(X, \theta) \tag{1}$$

$$L_{MSE} = \frac{1}{N}\sum_{i=1}^{N}\left\| Y_{true}^{i} - Y_{pred}^{i}\right\|^{2} \tag{2}$$

U-net is a widely used convolutional neural network in biomedical imaging [30]–[32]. It has encoder-decoder paths with a skip connection in each layer to extract the effective features from the training dataset without losing essential features during the downsampling procedure. The encoder-decoder paths of the U-net have been proven to perform satisfactorily in the segmentation of biomedical images [32]. It has the advantages of computational efficiency and the ability for training with a small dataset. An upgraded U-net, called FD U-net, was chosen as the basis for our model because it has been shown in previous works to yield better image quality outputs than U-net [33], [34]. FD U-net incorporates dense connectivity in both the encoder and decoder paths of the U-net. In each layer, a Dense block is implemented. In these Dense blocks, the block input and each convolutional layer are connected to all subsequent layers through channel-wise concatenation. Thus, the concatenated result of previous convolutional layers in the dense block and the dense block input serves as the input to each subsequent convolutional layer in the dense block convolution chain (right of Fig.2). This architecture, therefore, helps retain all the features learned by previous layers and builds upon the knowledge more explicitly than in normal convolutional blocks. Moreover, dense connectivity in layers allows the neural network to be deeper. It introduces more connections to efficiently propagate the gradient information without a vanishing problem, making the network relatively easier to train. With these many benefits over regular U-net in mind, we chose an FD U-net as the most capable architecture to train our data.

The FD U-net architecture used in this study is shown in Fig. 2. It has six layers. The input matrix size is 256 × 256. The first layer contains a 3 × 3 Convolution block that brings the input image depth of 1 to a depth of 16 channels. Then the first Dense block doubles the number of channels from 16 to 32 convolutional filters using a kernel size 3 × 3. After that, each Dense block is followed by a Down block, which consists of a 1 × 1 Convolution block with a stride of 1 and a 3 × 3 Convolution block with a stride of 2 (which halves the image size). This pattern of Dense block followed by a Down block gradually increases the number of filters to the maximum of 512 and reduces the image size to 8 × 8 at the bottleneck layer (last layer of the downsampling path). Each Convolution block consists of a convolutional layer with the specified kernel size, followed by a ReLU activation function, which is then followed by a batch normalization layer. Although some similar networks have found benefit in switching from ReLU to ELU or other activation functions, we have not in our experience found much difference in overall model performance when changing the activation function. Therefore, we have decided to retain the deep learning community standard of ReLU as the activation function. The second half of the network follows a similar repeated pattern of Up block, followed by shortcut concatenation with filters from earlier in the network (doubling the channel depth), followed by a 1 × 1 Convolution block with the stride of 1 (which reduces the channel depth by a factor of 4) and a 3 × 3 Dense block (which doubles the channel depth again). The Up block is similar to the Down block beginning with a 1 × 1 Convolution block with the stride of 1, but this time the convolution in the typical 3 × 3 Convolution block is replaced by a transposed convolution (or “deconvolution”) with a stride of 2 (which doubles the image size). Then in the final step, a 1 × 1 convolution followed by batch normalization reduces the channel depth from 32 filters to 1 output filter, which is then added to the original network input image to form the final network output.

B. 3D Focal Line Method

For a linear transducer array, the focal line is a line that goes across the element focus and is perpendicular to the x (axial)-z (elevation) plane of the element. Photoacoustic waves emitted from a point on this line will reach the entire surface of the transducer element simultaneously, which will maximize the receiving sensitivity of the transducer. Therefore, for an arbitrary point in 3D space, while the emitted wavefront has multiple paths to reach the transducer element, only the one that interacts with the focal line produces the maximum received amplitude. Inspired by this feature, the 3DFL method used this focal line as an auxiliary line for computing the time of arrival. The reconstruction procedures can be split into three steps. The first step is the 3D reconstruction for a single element. In this step, 3DFL first creates a 3D position matrix and then computes the distance (r) from each matrix point to the transducer surface based on the shortest path that travels across the focal line. Following that, 3DFL calculated the traveling time by dividing the distance (r) by the speed of sound and then back-projected the PA signal to the 3D matrix according to the acquisition time. In the second step, 3DFL repeats the same process for the rest of the 127 elements and sums the 3D matrix for each element. This finishes the reconstruction of one raw data frame acquired at a specific elevation position. In the third step, 3DFL repeats steps one and two for other frames acquired at different elevation positions and then sums all the 3D matrices to generate the final 3D image. The detailed principle of the 3DFL could be found in previous studies from our group [13], [35].

C. Simulation and Training

We proposed a simple method to generate training data for elevational resolution improvement. The simulation images were generated in 2D in the axial-elevational plane (“B-scan” images) instead of 3D. This approach can quickly accumulate a large number of simulated images as input for training. FD U-net is used to learn the features and after training, the network can improve the elevational resolution. This method

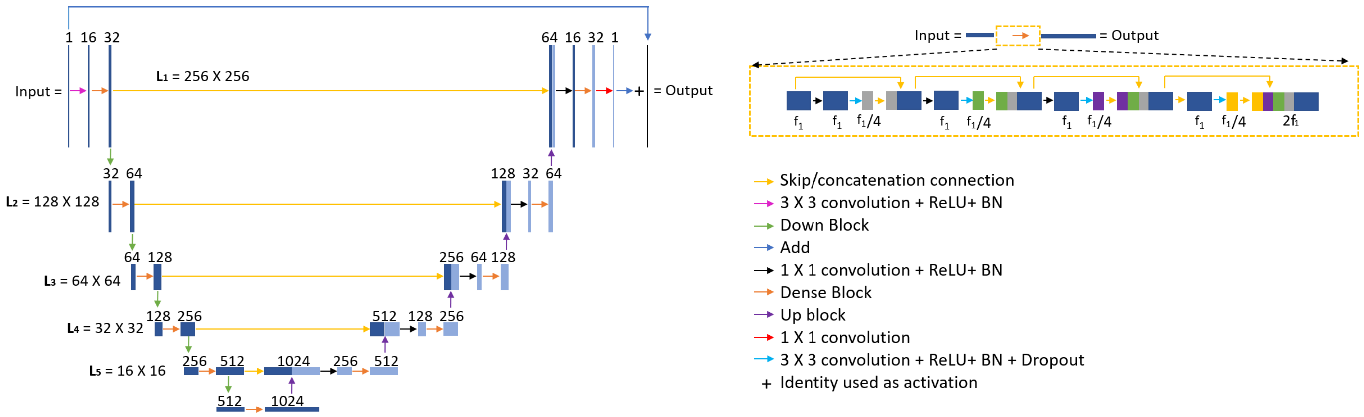

Fig. 2. A schematic of the FD U-net used as the basis for our model. The input matrix size is 256 by 256. The number on the top of each column is the feature maps in each layer. On the right, the yellow block shows the dense connectivity in detail. ReLU: Rectified Linear Unit. BN: Batch normalization. Conv: convolution. The yellow-dashed block shows the dense block at different layers. f1 represents the initial channels of the layer. The growth rate of each layer is four to learn a number of feature maps.

is termed as Deep-E. All the parameters for training a network are shown in Table I.

To generate the input data for training, we use the MATLAB-based photoacoustic simulation toolbox, K-Wave [36]. To prepare the low elevational resolution photoacoustic data in a linear-array-based PACT, a conventional approach would generate 3D data as the input. However, this process is cumbersome and time-consuming, as it needs to prepare a large amount of 3D ground truth images. As such, there is a need for more efficient simulation methods. Considering that the resolution in the lateral dimension is much higher than the elevational dimension, we only generate 2D images in the axial-elevational (AE) plane as training data. This approach is much more convenient and efficient for the preparation of the training dataset. Fig. 3 shows our simulation geometry. An arc-shaped transducer detects the A-line signal along the axial direction. The arc-shaped transducer is assumed to move along the elevational direction to mimic the elevation scanning with a step size of 0.1 mm. Then, the B-scan image is formed in the axial-elevation plane by stacking all the A-lines in sequence. This B-scan image is used as the input data for training. This arc-shaped transducer shares the same parameters as the experimental one, whose details can be found in the next section. It has a length of 15 mm with an acoustic focus at 40 mm axial depth.

The ground truth images are generated using the Insight Segmentation and Registration Toolkit (ITK) [37]. This toolkit can generate 3D vascular structures with multiple vascular branches and different vascular diameters. Sectioning the 3D structure will generate 2D images with similar features as the cross-sectional human breast data. One of the 3D volumes is displayed in Fig. 3. A video in Supplementary 1 (S1) shows the formation of the 3D volume.

To mimic the imaging size of human breast data, we simulate the input data in the AE-plane at the size of 50 × 50 mm with a pixel size of 0.1 mm. The object is placed 30 mm away from the acoustic detector in the axial direction. To achieve 50 mm of elevational scanning distance, we moved a single element 500 steps with 0.1 mm step size along the elevational direction. Eight 3D vessel datasets from ITK are generated for simulation to acquire a large amount of training data. Each 3D vessel dataset is composed of 200 segmented 2D images. It took 20 minutes to simulate 200 2D images in the 3D volume. These datasets are subjected to 4 grades of signal-to-noise ratio (SNR) at 12 dB, 9 dB, 6 dB, and noise-free, respectively. The experimental breast data determine these SNR levels. Supplementary 2 (S2) demonstrates the examples of the training data. In total, 6400 (8 × 200 × 4) images are used as the input. They are trained to achieve a Deep-E framework for elevational resolution improvement. We did not use other augmentation methods such as rotation, shifting, or rescaling [38] as they cannot preserve the acoustic focal position. As the spatial resolution varies at different distances to the acoustic focus, we decided to maintain the acoustic focal position for better training outcomes. Among all the simulated data, 72% were used for training, 18% were used for validation, and 10% were reserved for testing. The setup included an AMD Ryzen 9 3950X CPU, 128 GB RAM, and a NVIDIA GTX 2080Ti.

TABLE I
PARAMETERS FOR THE TRAINING NETWORK

Category/Function	Parameters	Value/Range
Photoacoustic data	Transducer width Transducer acoustic focus Transducer bandwidth Sampling rate Transducer central frequency Ultrasonic wavelength Transducer scanning step size	15 mm 40 mm 65% 9 MHz 2.25 MHz 0.68 mm 0.1 mm
Training image	Image numbers Signal to Noise Ratio Axial samples Elevation samples	6400 6, 9, 12 dB ,and noise free 256 pixels (50mm) 256 pixels (50mm)
Training	Batch size Learning rate Training iterations Training time	8 0.001 41,490 1.7 hours

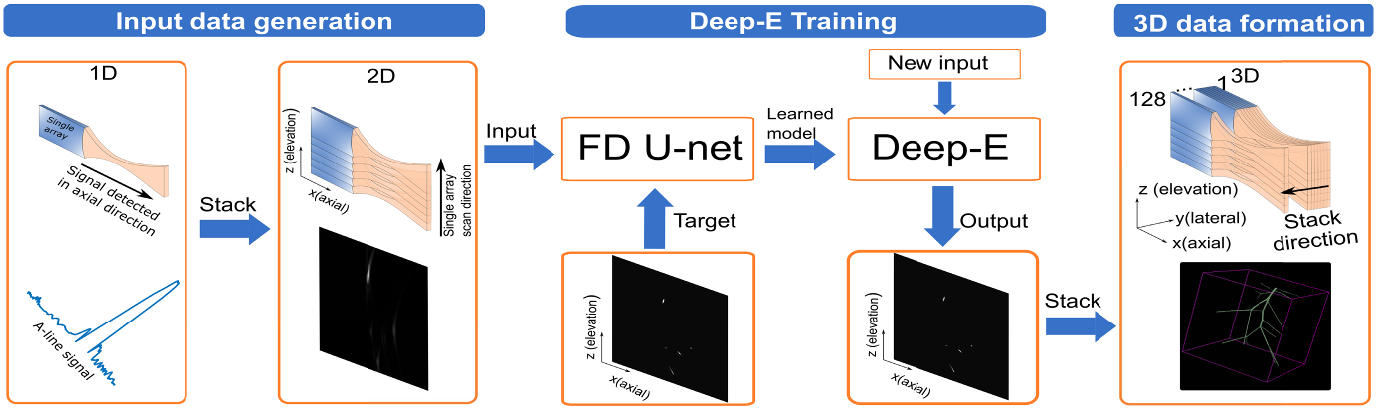

Fig. 3. Workflow of the Deep-E training and validation.

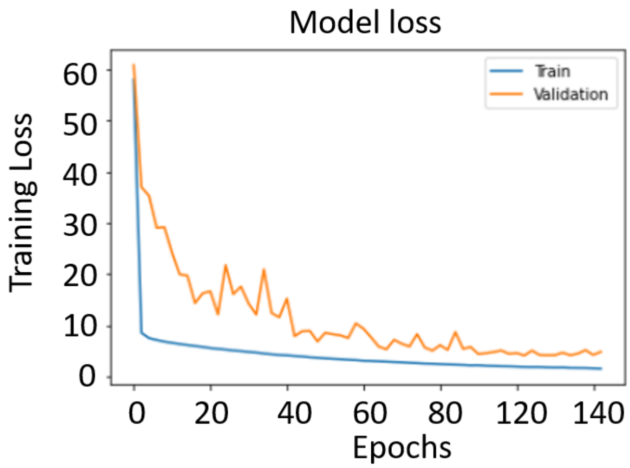

Fig. 4. MSE for training and validation in each epoch.

For the hyperparameters, we used a mini-batch size of 8, a bias initializer of zero, a learning rate of 0.001, and Adam as the optimizer. For the batch normalization, the momentum parameter was set to 0.99, and the epsilon was 0.001. The loss function used in training was the mean squared loss of the intensity between the last layer output and ground truth (2). The networks were implemented using Python 3.7 in Keras with a Tensorflow 2.0 backend in the cloud service of Google Colaboratory [39]. The total training took a total of 1.7 hours with 41,490 training iterations. Fig. 4 shows the training loss variation in MSE for different epochs during training.

D. Experimental Design

The experimental imaging system setup used in this study is presented in Fig. 1(b). A 128-element linear transducer with 2.25 MHz central frequency (Imasonics, Inc.) and 65% bandwidth is used for signal detection. Each element in the array is arc-shaped with a 15 mm elevation length and 40 mm axial focus. Our previous study quantified that the elevation resolution was approximately 1 mm after reconstruction with the 3DFL method [4]. The light source is a 10-Hz pulsed Nd: YAG Continuum Surelite III-10 laser with 1064 nm output wavelength and 10 ns pulse width. The laser output was coupled to a circular input, line-output fiber bundle. The line output and transducer were assembled using a 3D-printed mount. To achieve coaxial light delivery and acoustic detection, a dichroic mirror (cold mirror, Edmund Optics Inc.) was attached at a 45-degree angle to the transducer. The dichroic mirror allows transmission of near-infrared light (~97%) at an incident 45-degree angle. The generated acoustic waves were reflected by the mirror at 90 degrees and detected by the transducer. For linear scanning, a translation stage (McMaster-Carr) was utilized to move the transducer and fiber bundle simultaneously at the speed of 1 mm/s. The synchronization of the light delivery and ultrasound signal detection was achieved with trigger output from the laser. We used the Verasonics Vantage data acquisition (DAQ) system to receive photoacoustic signals and reconstructed raw data with the back-projection algorithm in MATLAB.

To test the performance of Deep-E, we did several experiments for different purposes. First, two pencil lead experiments were conducted to verify that Deep-E could effectively improve the elevational resolution and recover the true pencil lead diameter, even at different imaging depths. Then, we implemented Deep-E on human imaging data to verify whether it works well for *in vivo* data.

1) Pencil Lead Imaging: For the first pencil-lead experiment, we aimed to verify whether Deep-E could improve elevational resolution. As we described above, the object at acoustic focus has the best elevational resolution. Therefore, if an object increases its distance from the acoustic focus, its elevational resolution will gradually degrade. For this experiment, we prepared five pencil leads with a 0.5 mm diameter, which is smaller than the experimentally quantified elevational resolution [4]. These pencil leads were placed on a 3D-printed mount at different depths. One of the pencil leads was at the acoustic focus. The mounted pencil leads were immersed in water mixed with a 2% Intralipid for light scattering [40].

For the second pencil-lead experiment, we aimed to test the accuracy of Deep-E for resolution enhancement. For this purpose, we used pencil leads of various diameters: 0.5, 0.9, and 2 mm. We chose this diameter range based on the

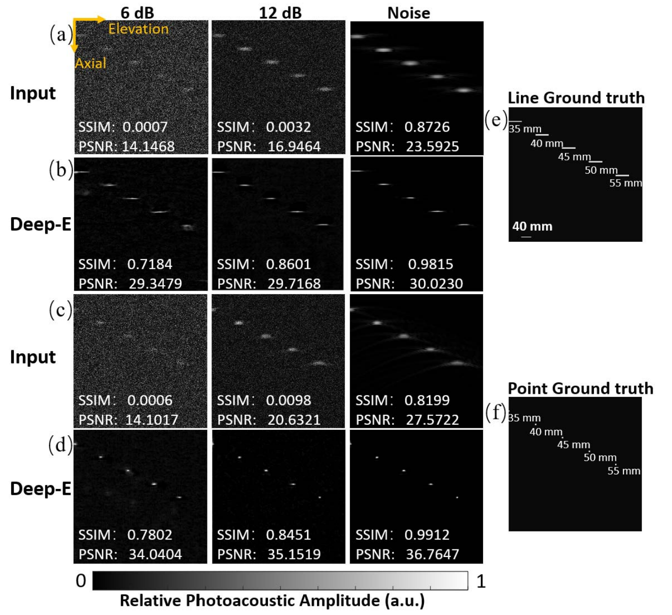

Fig. 5. Numerical data validation. (a): (left to right) input data (line objects) at 6 dB SNR, 12 dB SNR, and noise-free. (b): (left to right) output line data at 6 dB SNR, 12 dB SNR, and noise-free. (c): Ground truth of line object data. Five lines are placed beneath the transducer starting from 35 mm to 55 mm at 5 mm depth increments. From top to bottom, the lengths are 0.5 mm, 1 mm, 2 mm, 3mm, and 4mm, respectively. (d): Input data (point objects) at 6 dB SNR, 12 dB SNR, and noise-free. (e): Output data (point objects) at 6 dB SNR, 12 dB SNR, and noise-free. (f): Ground truth of point data. The diameter of each point is 0.5mm. Five points are placed beneath the transducer starting from 35 mm to 55 mm, at 5 mm depth increments.

breast vessel diameters in women [41]. The phantom is placed beneath the transducer at acoustic focus.

2) Human Breast Imaging: The Deep-E framework is also tested for human breast data. To acquire the experimental breast data, the study protocol is approved by the institution review board of the University at Buffalo. The breast imaging system is called Dual scan mammoscope (DSM), whose details were discussed in the papers [4], [42]. In brief, while standing upright, the volunteer placed her breast on the plastic film of the bottom water tank. The plastic film of the top water tank moved down to compress the top skin surface of the breast mildly. The combined transducer and fiber bundle mounts were placed into the water tanks to scan the breast at a step size of 1 mm/s. The transducer was 40 mm away from the skin surface because of a special design to achieve co-planar light delivery and acoustic detection [43]. The breast size determined the elevation scanning length (typically 5 cm). As for imaging depth, the transducer could image 3-4 cm underneath the skin surface. Therefore, we chose a 50 × 50 mm matrix size for the simulation data, which ensures that the simulation range fully covers the experimental condition.

III. Results

This section presents the simulation, phantom, and human imaging results. For better illustration, the 3D images are presented using maximum amplitude projection (MAP) along the axial direction. The MAP images are color encoded by depth along the axial direction. The colors blue to red in MATLAB's jet colormap represent the imaging depth from shallow to deep.

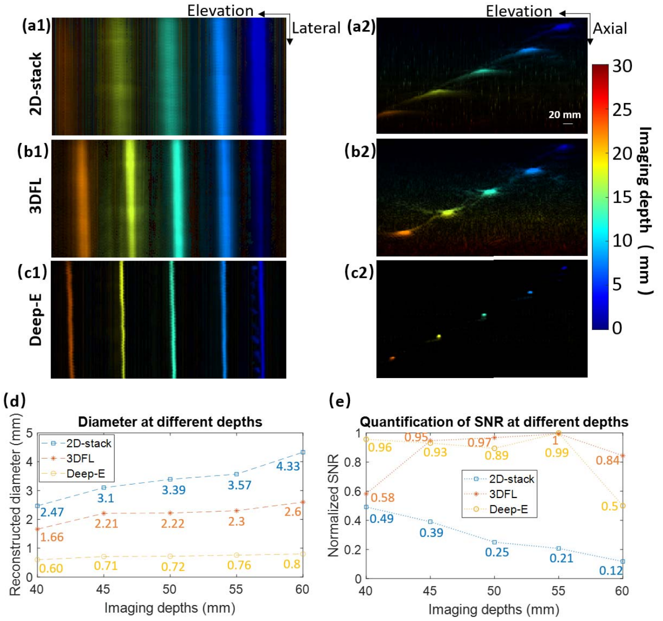

Fig. 6. Validation of Deep-E using 0.5 mm pencil leads placed at different depths. (a1, b1, c1) are the MAP images reconstructed by 2D-stack, 3DFL, and Deep-E, respectively. (a2, b2, c2) are the cross-sectional images reconstructed by 2D-stack, 3DFL, and Deep-E, respectively. (d). Quantification of pencil lead diameter through FWHM. (e). Quantification of 0.5-mm pencil leads SNR at different depths.

A. Validation With Numerical Simulation

We first examined the elevational resolution improvement on the simulation data. Two types of simulation data were used in the validation. As shown in Fig. 5, the first type consists of line objects with different lengths placed at different depths. The line length (0.5mm, 1 mm, 2 mm, 3 mm, and 4 mm) is increased from top to bottom. Its input and output results are presented in Figs. 5(a) and 5(b) under three SNR groups: 6 dB, 12 dB, and noise-free. The corresponding ground truth image is shown in Fig. 5(c). After deep learning, we can see that the Deep-E outputs show five clear line structures with a clean background. In addition, the length of the input is similar to the ground truth. The second type of testing dataset consists of 0.5 mm-diameter point data at different depths. As the imaging depth increases, the elevational resolution becomes poorer, as shown in Fig. 5(d). As expected, the resolution is improved in Fig. 5(e), and the point source is recovered to its original size. Overall, the results show that Deep-E works well with numerical data.

B. Validation With Experimental Pencil Lead Data

Deep-E was further validated through experimental studies. The pencil lead widths recovered by different reconstruction methods were quantified using a full-width at half maximum (FWHM) algorithm. The test objects were five pencil leads of 0.5 mm diameter placed 40 mm to 60 mm away from

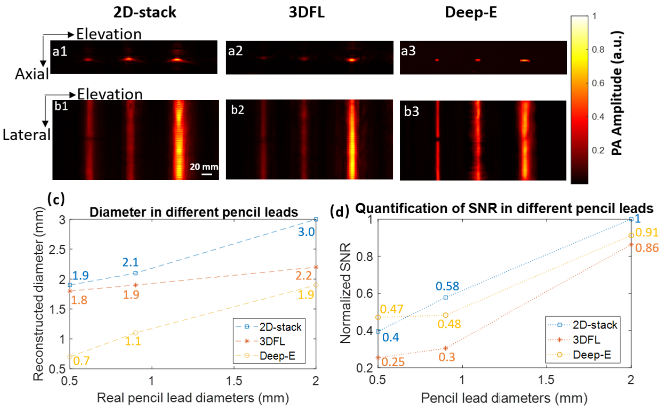

Fig. 7. Validation of Deep-E on pencil leads with different diameters. All pencil leads were placed at the acoustic focus. (a1, a2, a3): Cross-sectional images reconstructed by 2D-stack, 3DFL, and Deep-E, respectively. (b1, b2, b3): MAP images are reconstructed by 2D-stack, 3DFL, and Deep-E, respectively. (c). Quantification of the reconstructed pencil lead diameter through FWHM. (d). Quantification of the SNR of different diameter pencil leads.

the transducer at 5 mm depth increments. The depth-encoded MAP images are shown in the left column of Fig. 6. In the 2D reconstructed result of Fig. 6(a1), it can be seen that the dark blue pencil lead has the sharpest edge because it is placed at the acoustic focus. In contrast, the orange pencil lead shows the blurriest edge because it is farthest from the acoustic focus. In the second column, Figs. 6(a2), 6(b2), and 6(c2) are the cross-sectional images in the AE-plane, which can better demonstrate the effective resolution in the three different reconstruction methods. 3DFL improves the elevational resolution, as shown in Figs. 6(b1) and 6(b2). However, the FWHM of the pencil leads is still higher than the true diameter. In comparison, the Deep-E results shown in Figs. 6(c1) and 6(c2) are much closer to the original diameter of the pencil lead. The FWHM quantification and normalized SNR are summarized in Figs. 6(d) and 6(e). Compared to 2D-stack, Deep-E provided at least four times improvement in the elevational resolution. Deep-E gave the best estimation of true pencil lead diameter and the higher overall SNR among the three reconstructed methods.

Next, we tested pencil leads with three different diameters (0.5, 0.9, and 2mm). They were all placed at the acoustic focus, where optimal the elevational resolution (Fig. 7). Again, the Deep-E output in Fig. 7(a3) continues to recover the pencil lead's true diameter. Compared to 2D-stack and 3DFL, the cross-sectional images in Fig. 7(a3) provided the best diameter estimation. The FWHM quantification in Fig. 7(c) further validated our assumption.

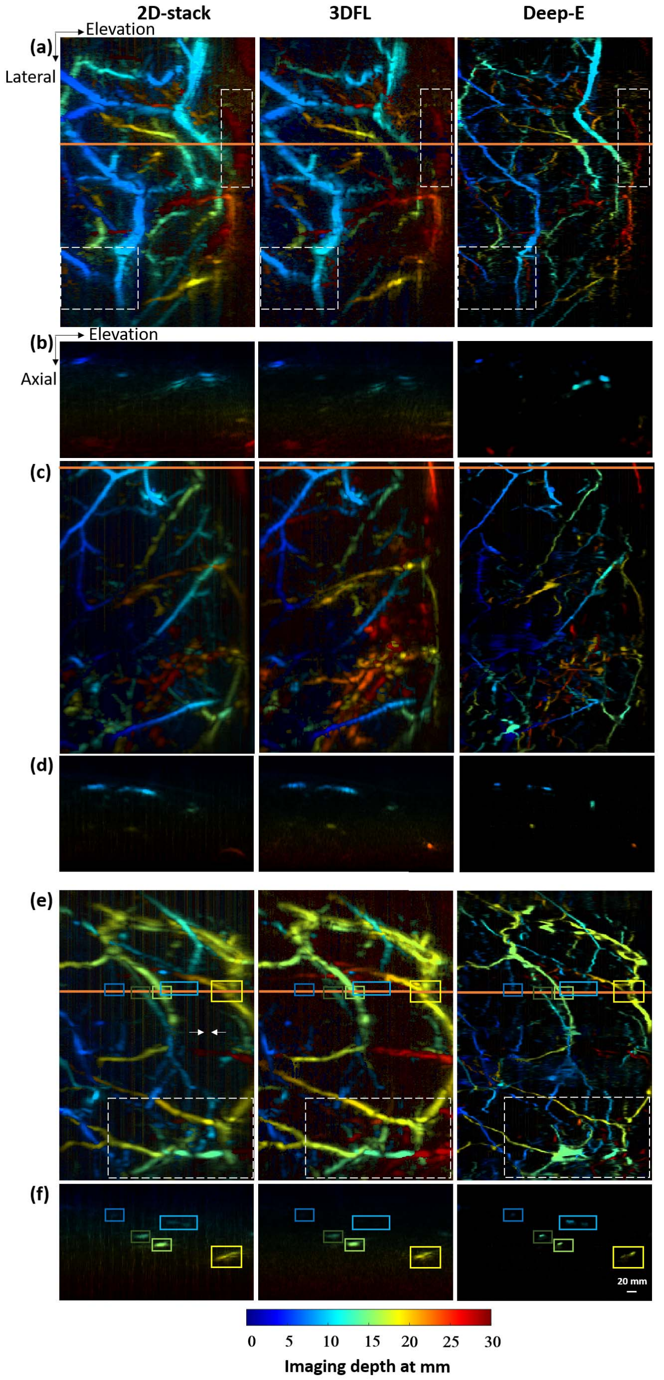

Fig. 8. Validation in human breast data from three human different subjects. From the left to right are three reconstruction methods: 2D-stack, 3DFL, and Deep-E. (a), (c), (e): the reconstructed MAP breast images from three volunteers, respectively. (b), (d), (f): the cross-sectional images from three volunteers, respectively. The orange lines in a, c, e mark the position for the cross-sectional image in b, d, f.

C. Validation With Human Breast Data

To validate Deep-E for *in vivo* applications, we applied the network to human breast imaging results. Human breast data from three different subjects are used. The MAP images in Fig. 8 show the performance of the Deep-E network in comparison to 2D-stack and 3DFL. Figs. 8(a), 8(c), and 8(e) are the conventional 2D reconstruction results. All these data are shown over a 30 mm reconstruction depth with the starting depth at 40 mm away from the transducer. The large starting depth is caused by our co-planar reflector design. As expected, the 2D reconstruction results exhibited poorer elevation resolution, especially for vessels far away from the transducer focus. For example, in the first breast data, the red color vessel on the right edge (marked with white-dashed block) in Fig. 8(a) is blurred into the background. 3DFL recovers the vessels and shows a clean red vessel shape. Compared to these two images, Deep-E gives shaper vascular structures with a clean background. More importantly, Deep-E can extract vascular structures in deep tissue, colored in orange and red, which are hard to notice in 2D-stack and 3DFL images, such as the orange vessel (marked with white-dashed block) at the left corner of Fig. 8(a). For better illustration, we also plotted the cross-sectional images taken across the orange-marked line in the MAP images in Figs. 8(b), 8(d), and 8(f). We specially marked the vessels in Fig. 8(e) with solid blocks. From left to right are five main vessel points at different depths, where they are colored in blue, green, light green, blue and yellow, respectively. Their structures match well with the vessel point on the orange line, as marked with the rectangular block in Fig. 8(f). We can notice that 2D-stack images have the poorest elevational resolution while Deep-E can refine the breast vessels into sharper features. Moreover, we notice that the Deep-E network removed the stripe artifacts induced by the DAQ in Fig. 8(e), marked with white arrows. The photoacoustic overlay on the ultrasound images are present in supplementary 3 (S3).

TABLE II
COMPUTATION TIME FOR 3D DATA RECONSTRUCTION

Computer properties		NVIDIA GeForce RTX 2080Ti, AMD Ryzen 9 3950X CPU	NVIDIA GeForce RTX 2070 super, Intel Core i7-9700K CPU
3DFL		~12 minutes	~20 minutes
Deep-E	Pre	~ 30 seconds	~ 30 seconds
	Post	~ 1 second	~ 1 second

The matrix size for computation is 250*250*430, based on the image size of 50*50*86 mm.
Pre: using 2D stack method to prepare the reconstructed data.
Post: processing the reconstructed data in the trained model.

IV. DISCUSSION

Our work applies deep learning techniques to linear-array-based PACT. We propose a unique deep learning network, Deep-E, which utilizes 2D training data to solve a 3D problem. The novelty of our simulation method is to generate a 2D matrix in the axial-elevational plane using an arc-shaped transducer element instead of generating a 3D matrix using the linear transducer arrays. Deep-E has several advantages: (1) The generation of 2D training data is much faster than the conventional 3D simulation. Because we use 2D images as training data, the simulation and training times are orders of magnitudes shorter than 3D. (2) It provides a high-speed reconstruction of 3D images. Table II provides the computation time for 3D reconstruction in Deep-E and 3DFL. Processing the experimental data in the trained Deep-E model took less than a second. The total computation time for loading and arranging the experimental data, processing in Deep-E, and remapping the Deep-E results in lateral-elevation dimensions is less than 30 seconds in an automated master code. In comparison, 3DFL reconstruction took 12 or 20 minutes (depending on the computation power). (3) The simulation method is not limited by the number of elements. For example, while we demonstrated the results in a 128-element array, the trained network can also be used in 64 or 256 element arrays. Because the experimental data was processed element by element independently in the axial-elevation plane, the number of elements in the array will not affect the final result. (4) The method can also be utilized in other transducer geometries. By adjusting how we combine data from different elements, Deep-E can also be utilized in circular or arc-shaped arrays to improve the elevation resolution. Moreover, since photoacoustic imaging shares the same detection and reconstruction principles with ultrasound, the Deep-E method can also be applied in ultrasonography, allowing for better 3D imaging of anatomical structures [44]. (5) Deep-E also removes noise and artifacts. Our training data was crafted at multiple noise levels while the ground truth images did not contain any noises or artifacts. When the Deep-E model was trained, it removed artifacts and noises that did not look like vascular structures. These advances allow Deep-E to be utilized in a wide range of PACT systems and applications.

After testing on simulated vascular data, our Deep-E model can improve the elevational resolution in a wide range of simulation and experimental results. Based on the quantification of FWHM, it shows that Deep-E can improve elevational resolution by at least four times. We first tested the Deep-E using numerical simulation data. We used a variety of loss metrics to compare its performance with input, as shown in Fig. 5. Deep-E has the highest value in peak signal-to-noise (PSNR) and structural similarity index (SSIM) [45], which means that the output has very little noise, and the features look very close to the ground truth. Next, we tested Deep-E on experimental pencil leads data. While both 3DFL and Deep-E can improve the elevational resolution. Deep-E is superior to 3DFL because the reconstructed pencil leads' widths are closer to the real size. This confirms that Deep-E does not simply reduce vessel diameters but rather restore the vessel diameter to its true size and refine the elevational resolution at different imaging depths. We also tested the Deep-E in human breast data. The result in Fig. 8 shows that Deep-E can recover vascular structures much better than 2D-stack and 3DFL. It not only refines the vessel resolution but also reveals deeper vessels in the majority. This is because our training data has a deep imaging depth of 50 mm at different noise levels, which makes Deep-E effective in recovering deeper photoacoustic signals. The stripe artifacts are also removed because Deep-E can differentiate the real photoacoustic signal and background noise. Moreover, although the network focuses on elevational resolution, there also appears to be an improvement in axial and lateral resolution. The axial resolution of pencil lead in Figs. 6 and 7 appear less in Deep-E images compared to other images. Similarly in Fig. 8, there is an improvement in the resolution of blood vessels present in all directions. Because our ground truth data was presented in the axial-elevation plane, resolution along these two directions was improved. The lateral resolution improvement was a secondary effect. Because vessels are orientated in 3D, improvements in the other two directions make an apparent improvement in the lateral direction.

Our results indicate that Deep-E can be successfully applied to *in vivo* experimental data, even if the vascular structures are dense and complicated. To the best of our knowledge, this is the first study that uses deep learning to improve 3D human breast data in PACT.

Although Deep-E has significant performance in enhancing the *in vivo* images' elevational resolution, some limitations can be improved in future studies. First, some vessels cannot be extracted efficiently in Deep-E. For example, the blue vessel in the center of Fig. 8a is discontinuous. This issue is mainly caused by the inconsistency of signal intensity along the vessel. In terms of recovering deep vessels, while Deep-E is superior to 2D reconstruction, it is not as good as 3DFL in certain cases. For example, in Fig. 8e, the red vessels marked with a white-dashed block are more apparent in 3DFL. This is because 3DFL considers receiving angles along the elevation direction. Deeper vessels can be seen at multiple elevation detection positions and thus can be recovered better in 3DFL. In future studies, we plan to use information from adjacent frames to improve vessel continuity and deep vessel recovery. Second, the quantitative information might be lost after Deep-E processing. For instance, in Fig. 7, after Deep-E processing, the intensity of the 0.5 mm pencil lead has been improved while the intensity of the other two pencil leads did not change. The quantitative accuracy could be improved with specially designed training data at various signal intensities. Lastly, the current Deep-E model focuses only on enhancing the elevation resolution. It did not consider improvement in lateral resolution nor other issues in linear array detection, such as the limited-view problem [46]. With the help of high-speed 3D simulation methods [47], a more comprehensive model could be developed to address these issues.

V. Conclusion

This work aims to apply deep learning techniques to enhance the elevational resolution with high speed in 3D PACT. We propose a simple method to generate input data that mimics the poor elevation resolution in a linear array. The pencil leads data demonstrated that Deep-E could improve the elevation resolution and restore the object's true size instead of just converting everything into point sources. Deep-E also exhibited significant resolution improvement on the *in vivo* human breast data. In addition, we were able to restore deeper vascular structures and remove the noise artifact. Moreover, Deep-E is not limited by the number of elements of the transducer or the transducer geometry, which means it can enhance elevational resolution on various imaging systems if the data are appropriately combined. We envision that Deep-E will significantly impact linear-array-based photoacoustic imaging studies by providing high-speed and high-resolution image enhancement.

References

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

[2] X. Xu, H. Liu, and L. V. Wang, "Time-reversed ultrasonically encoded optical focusing into scattering media," *Nature Photon.*, vol. 5, no. 3, p. 154, 2011.

[3] J. Xia, J. Yao, and L. V. Wang, "Photoacoustic tomography: Principles and advances," *Electromagn. Waves*, vol. 147, p. 1, 2014.

[4] N. Nyayapathi *et al.*, "Dual scan mammoscope (DSM)—A new portable photoacoustic breast imaging system with scanning in craniocaudal plane," *IEEE Trans. Biomed. Eng.*, vol. 67, no. 5, pp. 1321–1327, May 2020.

[5] N. Nyayapathi and J. Xia, "Photoacoustic imaging of breast cancer: A mini review of system design and image features," *J. Biomed. Opt.*, vol. 24, no. 12, 2019, Art. no. 121911.

[6] Y. Bao *et al.*, "Development of a digital breast phantom for photoacoustic computed tomography," *Biomed. Opt. Exp.*, vol. 12, no. 3, pp. 1391–1406, 2021.

[7] J. Yao *et al.*, "Noninvasive photoacoustic computed tomography of mouse brain metabolism *in vivo*," *NeuroImage*, vol. 64, pp. 257–266, Jan. 2013.

[8] M. Holotta *et al.*, "Photoacoustic tomography of *ex vivo* mouse hearts with myocardial infarction," *J. Biomed. Opt.*, vol. 16, no. 3, 2011, Art. no. 036007.

[9] Y. Wang *et al.*, "A robust and secure palm vessel biometric sensing system based on photoacoustics," *IEEE Sensors J.*, vol. 18, no. 14, pp. 5993–6000, Jul. 2018.

[10] J. Xia *et al.*, *Whole-Body Ring-Shaped Confocal Photoacoustic Computed Tomography of Small Animals* in vivo. Bellingham, WA, USA: SPIE, 2012.

[11] Y. Wang, Y. Zhan, M. Tiao, and J. Xia, "Review of methods to improve the performance of linear array-based photoacoustic tomography," *J. Innov. Opt. Health Sci.*, vol. 13, no. 2, Mar. 2020, Art. no. 2030003.

[12] J. Xia *et al.*, "Three-dimensional photoacoustic tomography based on the focal-line concept," *J. Biomed. Opt.*, vol. 16, no. 9, 2011, Art. no. 090505.

[13] D. Wang *et al.*, "Three-dimensional photoacoustic tomography through coherent-weighted focal-line-based image reconstruction," in *Proc. Spie*, 2017, Art. no. 100643G.

[14] J. Gateau *et al.*, "Single-side access, isotropic resolution, and multispectral three-dimensional photoacoustic imaging with rotate-translate scanning of ultrasonic detector array," *J. Biomed. Opt.*, vol. 20, no. 5, 2015, Art. no. 056004.

[15] M. Schwarz, A. Buehler, and V. Ntziachristos, "Isotropic high resolution optoacoustic imaging with linear detector arrays in bi-directional scanning," *J. Biophotonics*, vol. 8, nos. 1–2, pp. 60–70, Jan. 2015.

[16] Y. Wang *et al.*, "Slit-enabled linear-array photoacoustic tomography with near isotropic spatial resolution in three dimensions," *Opt. Lett.*, vol. 41, no. 1, pp. 127–130, 2016.

[17] A. Hauptmann and B. T. Cox, "Deep learning in photoacoustic tomography: Current approaches and future directions," *J. Biomed. Opt.*, vol. 25, no. 11, 2020, Art. no. 112903.

[18] H. Deng, H. Qiao, Q. Dai, and C. Ma, "Deep learning in photoacoustic imaging: A review," *J. Biomed. Opt.*, vol. 26, no. 4, Apr. 2021, Art. no. 040901.

[19] C. Yang, H. Lan, F. Gao, and F. Gao, "Review of deep learning for photoacoustic imaging," *Photoacoustics*, vol. 21, Mar. 2021, Art. no. 100215.

[20] K.-T. Hsu, S. Guan, and P. V. Chitnis, "Comparing deep learning frameworks for photoacoustic tomography image reconstruction," *Photoacoustics*, vol. 23, Sep. 2021, Art. no. 100271.

[21] G. Godefroy, B. Arnal, and E. Bossy, "Compensating for visibility artefacts in photoacoustic imaging with a deep learning approach providing prediction uncertainties," *Photoacoustics*, vol. 21, Mar. 2021, Art. no. 100218.

[22] D. Allman, A. Reiter, and M. A. L. Bell, "Photoacoustic source detection and reflection artifact removal enabled by deep learning," *IEEE Trans. Med. Imag.*, vol. 37, no. 6, pp. 1464–1477, Jun. 2018.

[23] N. Awasthi, G. Jain, S. K. Kalva, M. Pramanik, and P. K. Yalavarthy, "Deep neural network-based sinogram super-resolution and bandwidth enhancement for limited-data photoacoustic tomography," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 67, no. 12, pp. 2660–2673, Dec. 2020.

[24] A. Hauptmann *et al.*, "Model-based learning for accelerated, limited-view 3-D photoacoustic tomography," *IEEE Trans. Med. Imag.*, vol. 37, no. 6, pp. 1382–1393, Jun. 2018.

[25] H. Zhang *et al.*, "A new deep learning network for mitigating limited-view and under-sampling artifacts in ring-shaped photoacoustic tomography," *Computerized Med. Imag. Graph.*, vol. 84, Sep. 2020, Art. no. 101720.

[26] A. Y. Yuan *et al.*, "Hybrid deep learning network for vascular segmentation in photoacoustic imaging," *Biomed. Opt. Exp.*, vol. 11, no. 11, pp. 6445–6457, 2020.

[27] J. Zhang, B. Chen, M. Zhou, H. Lan, and F. Gao, "Photoacoustic image classification and segmentation of breast cancer: A feasibility study," *IEEE Access*, vol. 7, pp. 5457–5466, 2018.

[28] A. Sharma and M. Pramanik, "Convolutional neural network for resolution enhancement and noise reduction in acoustic resolution photoacoustic microscopy," *Biomed. Opt. Exp.*, vol. 11, no. 12, pp. 6826–6839, 2020.

[29] P. Rajendran and M. Pramanik, "Deep learning approach to improve tangential resolution in photoacoustic tomography," *Biomed. Opt. Exp.*, vol. 11, no. 12, pp. 7311–7323, 2020.

[30] S. Antholzer *et al.*, "Deep learning for photoacoustic tomography from sparse data," *Inverse Problems Sci. Eng.*, vol. 27, no. 7, pp. 1–19, 2018.

[31] M. Kim, G.-S. Jeng, I. Pelivanov, and M. O'Donnell, "Deep-learning image reconstruction for real-time photoacoustic system," *IEEE Trans. Med. Imag.*, vol. 39, no. 11, pp. 3379–3390, Nov. 2020.

[32] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, 2015, pp. 234–241.

[33] S. Guan, A. A. Khan, S. Sikdar, and P. V. Chitnis, "Fully dense UNet for 2-D sparse photoacoustic tomography artifact removal," *IEEE J. Biomed. Health Informat.*, vol. 24, no. 2, pp. 568–576, Feb. 2020.

[34] A. DiSpirito *et al.*, "Reconstructing undersampled photoacoustic microscopy images using deep learning," *IEEE Trans. Med. Imag.*, vol. 40, no. 2, pp. 562–570, Feb. 2020.

[35] J. Xia *et al.*, "Three-dimensional photoacoustic tomography based on the focal-line concept," *J. Biomed. Opt.*, vol. 16, no. 9, 2011, Art. no. 090505.

[36] B. E. Treeby and B. T. Cox, "k-wave: MATLAB toolbox for the simulation and reconstruction of photoacoustic wave fields," *J. Biomed. Opt.*, vol. 15, no. 2, 2010, Art. no. 021314.

[37] G. Hamarneh and P. Jassi, "VascuSynth: Simulating vascular trees for generating volumetric image data with ground-truth segmentation and tree analysis," *Computerized Med. Imag. Graph.*, vol. 34, no. 8, pp. 605–616, Dec. 2010.

[38] T. Vu, M. Li, H. Humayun, Y. Zhou, and J. Yao, "A generative adversarial network for artifact removal in photoacoustic computed tomography with a linear-array transducer," *Experim. Biol. Med.*, vol. 245, no. 7, pp. 597–605, Apr. 2020.

[39] T. Carneiro, R. V. M. Da Nóbrega, T. Nepomuceno, G.-B. Bian, V. H. C. De Albuquerque, and P. P. Reboucas Filho, "Performance analysis of Google colaboratory as a tool for accelerating deep learning applications," *IEEE Access*, vol. 6, pp. 61677–61685, 2018.

[40] Y. Yamaoka, M. Nambu, and T. Takamatsu, "Frequency-selective multiphoton-excitation-induced photoacoustic microscopy (MEPAM) to visualize the cross sections of dense objects," *Photons Plus Ultrasound, Imag. Sens.*, vol. 7564, Feb. 2010, Art. no. 75642O.

[41] A. Grubstein, M. Yepes, and R. Kiszonas, "Magnetic resonance imaging of breast vascularity in medial versus lateral breast cancer," *Eur. J. Radiol.*, vol. 75, no. 2, pp. e9–e11, Aug. 2010.

[42] N. Nyayapathi *et al.*, "Photoacoustic dual-scan mammoscope: Results from 38 patients," *Biomed. Opt. Exp.*, vol. 12, no. 4, pp. 2054–2063, 2021.

[43] Y. Wang, R. S. A. Lim, H. Zhang, N. Nyayapathi, K. W. Oh, and J. Xia, "Optimizing the light delivery of linear-array-based photoacoustic systems by double acoustic reflectors," *Sci. Rep.*, vol. 8, no. 1, pp. 1–7, Dec. 2018.

[44] T. Lucas, I. Quidu, S. L. Bridal, and J. Gateau, "High-contrast and-resolution 3-D ultrasonography with a clinical linear transducer array scanned in a rotate-translate geometry," *Appl. Sci.*, vol. 11, no. 2, p. 493, Jan. 2021.

[45] A. Hore and D. Ziou, "Image quality metrics: PSNR vs. SSIM," in *Proc. 20th Int. Conf. Pattern Recognit.*, 2010, pp. 2366–2369.

[46] Y. Xu, L. V. Wang, G. Ambartsoumian, and P. Kuchment, "Reconstructions in limited-view thermoacoustic tomography," *Med. Phys.*, vol. 31, no. 4, pp. 724–733, Mar. 2004.

[47] J. Zalev and M. C. Kolios, "Fast 3-D opto-acoustic simulation for linear array with rectangular elements," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 68, no. 5, pp. 1885–1906, May 2020.