Research Article Vol. 14, No. 11 / 1 Nov 2023 / Biomedical Optics Express 5817
Biomedical Optics EXPRESS

Deep learning on photoacoustic tomography to remove image distortion due to inaccurate measurement of the scanning radius

SUDEEP MONDAL,[1] SUBHADIP PAUL,[1] NAVJOT SINGH,[2] AND RATAN K SAHA[1,*]

[1]Department of Applied Sciences, Indian Institute of Information Technology Allahabad, Prayagraj, 211015, India
[2]Department of Information Technology, Indian Institute of Information Technology Allahabad, Prayagraj, 211015, India
*ratank.saha@iiita.ac.in

Abstract: Photoacoustic tomography (PAT) is a non-invasive, non-ionizing hybrid imaging modality that holds great potential for various biomedical applications and the incorporation with deep learning (DL) methods has experienced notable advancements in recent times. In a typical 2D PAT setup, a single-element ultrasound detector (USD) is used to collect the PA signals by making a 360° full scan of the imaging region. The traditional backprojection (BP) algorithm has been widely used to reconstruct the PAT images from the acquired signals. Accurate determination of the scanning radius (SR) is required for proper image reconstruction. Even a slight deviation from its nominal value can lead to image distortion compromising the quality of the reconstruction. To address this challenge, two approaches have been developed and examined herein. The first framework includes a modified version of dense U-Net (DUNet) architecture. The second procedure involves a DL-based convolutional neural network (CNN) for image classification followed by a DUNet. The first protocol was trained with heterogeneous simulated images generated from three different phantoms to learn the relationship between the reconstructed and the corresponding ground truth (GT) images. In the case of the second scheme, the first stage was trained with the same heterogeneous dataset to classify the image type and the second stage was trained individually with the appropriate images. The performance of these architectures has been tested on both simulated and experimental images. The first method can sustain SR deviation up to approximately 6% for simulated images and 5% for experimental images and can accurately reproduce the GTs. The proposed DL-approach extends the limits further (approximately 7% and 8% for simulated and experimental images, respectively). Our results suggest that classification-based DL method does not need a precise assessment of SR for accurate PAT image formation.

1. Introduction

Photoacoustic tomography (PAT) is an emerging medical imaging modality with vast potential in biomedical applications such as breast imaging [1,2], small animal brain imaging [3,4], blood vasculature imaging [5,6], sentinel lymph node imaging [7], and so on [8,9]. It combines the benefits of optical and ultrasound imaging, where light gives high contrast, on the other hand ultrasound offers good penetration depth and high spatial resolution [10,11]. Conventionally, laser pulses of nanosecond duration are used to illuminate the biological sample. The laser energy is absorbed by the chromophores present within the sample, leading to local heating. As a result, ultrasound waves are generated due to rapid thermoelastic expansion (i.e., the PA effect). An ultrasound detector (USD) typically revolves around the sample in a circular manner and detects the PA signals.

#501277 https://doi.org/10.1364/BOE.501277
Journal © 2023 Received 24 Jul 2023; revised 17 Sep 2023; accepted 4 Oct 2023; published 17 Oct 2023

Biomedical Optics EXPRESS

Various algorithms have been developed to reconstruct PAT images utilizing the recorded radio frequency (RF) signals. The backprojection (BP) [12,13] and time-reversal algorithms [14,15,16] are analytical methods. The model-based techniques are iterative approaches. The analytical schemes are computationally fast, but they lack to provide accurate quantitative information of the imaging region [17,18]. They perform well when a full-view dataset is available. The iterative approaches are computationally expensive, but they can facilitate accurate quantitative estimation. They can work faithfully even when a full view dataset is not available. Recently, Li et al., carried out small-animal whole-body imaging at high spatiotemporal resolution utilizing a 512-element full-ring ultrasonic transducer array (5 MHz as the center frequency and 90% (approximately) as the fractional bandwidth) [19]. The same setup was also deployed to generate PAT images of human breasts [20]. Ding et al., developed a 3D PAT imaging system with detection elements of arbitrary size and shape [21]. They also presented a model-based reconstruction protocol that takes advantages of graphics processing unit (GPU) for real-time imaging. Such a setup may become useful in clinical setting [22].

Recently, machine learning (ML) or deep learning (DL) methods have been extensively utilized for PAT image reconstruction [23,24,25]. In general, two different approaches have been adopted. In the first procedure, the PA signals at many different angular locations are estimated using the ML/DL methods and then image reconstruction is accomplished. Dense spatial sampling is known to provide convincing images. In the second convention, approximate images are formed using the acquired PA signals and realizing a reconstruction algorithm. After that ML/DL frameworks are implemented to form improved PAT images. For example, DL-based convolutional neural network (CNN) framework involving U-Net architecture has been widely used to achieve this [26]. The U-Net model has made significant developments in the field of PAT, making it an ideal choice for remedying the calibration process for many parameters [26,27,28,29]. A modified U-Net architecture has also been applied in PAT imaging for sparse data problems [30]. Another advanced version of U-Net architecture, named fully dense U-Net (FD U-Net), is proposed for 2D PAT artifact removal [31]. Note that in most of the above mentioned works, images of a single phantom are used to form the training dataset. Very limited studies incorporate heterogeneous dataset during training [32]. However, PA and ultrasonic imaging can work together. The idea is that the ultrasonic imaging modality will help to visualize large objects (e.g., tumor, cyst) whereas the PA imaging will display small structures (e.g., blood vessels). In such a situation, a heterogeneous dataset (ultrasonic images and PA images) can be built and hence, performance of a network may be evaluated. It will be useful for dual modality imaging.

The ultimate goal of these studies is to produce PA images free from artifacts and distortions originating from system-dependent, medium-dependent or algorithm-dependent factors. The common sources are the finite width of the excitation laser pulse, the finite bandwidth of the USD, sparsity of data, inaccurate measurement of the scanning radius (SR) and many more. As an example, the effect of inaccurate measurement of SR is given in Fig. S1. It is evident from Fig. S1(a) and (b) that the ground truth (GT) and the reconstructed image, respectively exhibit close match if the exact value of SR is fed to the algorithm. Image distortion greatly occurs for a small deviation of SR as can be seen from Fig. S1(c) and (d). It may be noted that appropriate evaluation of SR in an experimental setting is very tedious and cumbersome [33]. In practice, an approximate SR is taken first and then an image is formed using a reconstruction algorithm (e.g., the BP algorithm). After that, the SR is tuned so that the best image can be obtained. The decision is taken visually and it may vary from person to person. Naturally, it may not be the best procedure. A DL method has been employed to address this issue [32,34]. The network took an input image of size 128×128 and generated an output image of the same size. The network was trained with 1260 simulated images for two different numerical phantoms. The training time was approximately 4 hours. The results showed that the network was able to form

Biomedical Optics EXPRESS

satisfactory reconstructions as long as there was a variation of SR up to ±5%. Therefore, it may be attempted to design a network which would take less time to train, include a variety of images in the training dataset and can tolerate larger deviation of SR.

The objective of this paper is twofold. The first objective is to investigate how a DL network would perform if it is trained with a heterogeneous dataset. The second objective is to develop a strategy which will be computationally efficient as well as can tolerate larger fluctuation in SR without compromising the image-quality. Two post-processing DL-based approaches have been formulated and assessed rigorously and systematically. The first approach includes a dense U-Net architecture (DUNet) which utilizes the concept of dense blocks to aid improved prediction. The second one involves image classification in combination with the DUNet. The first scheme was trained and optimized with 1200 heterogeneous images generated via simulations for three unique samples namely, two-point, multi-ellipse and vasculature phantoms. In the second methodology, first a three layered U-Net architecture was applied. It was trained with the same heterogeneous dataset to classify the images into three categories. Afterward, the DUNet was implemented to the individual categories for image prediction. The DUNet was optimized solely with a set of 400 images of each phantom. Both the approaches were tested using simulated and experimental datasets. The efficiency of these networks were compared with the universal BP algorithm and a popular CNN model [30]. It has been seen that the first technique can permit SR deviation up to 43.75 ± 2.65 mm and 43.75 ± 2.20 mm for simulated and experimental images, respectively. It can recover the GTs perfectly within these ranges. These limits are further stretched by the second protocol and are found to be 43.75 ± 3.06 mm and 43.75 ± 3.5 mm, respectively. To the best of our knowledge, this approach has never been implemented in the context of PAT imaging. It may find realistic applications and may guide future studies.

2. Theoretical aspect

2.1. PA signal generation

Let us consider that a soft biological tissue containing chromophores is irradiated by a short-pulse laser light (which can be represented as a delta function, δ(t)), that leads to the generation of acoustic waves. The mathematical equation for PA signal production and propagation across an acoustically homogeneous medium can be written as [35,36],

$$\nabla^2 p(\mathbf{r}, t) - \frac{1}{v^2}\frac{\partial^2 p(\mathbf{r}, t)}{\partial t^2} = -\frac{p_0(\mathbf{r})}{v^2}\frac{d\delta(t)}{dt}, \tag{1}$$

here $p(\mathbf{r}, t)$ is the acoustic pressure at position $\mathbf{r}$ and time t; v is the speed of the sound for the medium; $p_0(\mathbf{r}) = \Gamma(\mathbf{r})A(\mathbf{r})$ is the initial pressure buildup due to light absorption; where $\Gamma(\mathbf{r})$ is the Grüneisen parameter and defined as $\Gamma(\mathbf{r}) = v^2\beta/C_p$; β indicates the isobaric volume expansion coefficient and C_p denotes the specific heat for the absorbing region. Further, $A(\mathbf{r})$ is the spatial absorption function. There are several numerical methods available to solve Eq. (1), such as the pseudo-spectral method [37], finite element method [38], Green's function approach etc [39]. The cartoon, Fig. S2(a), displays a 2D situation and demonstrates that a point detector is placed at $\mathbf{r}_{SR}$ and the pressure signal is coming from the point Q. The detected signal is essentially the linear superposition of the tiny signals generated by these point sources (on the arch).

2.2. PAT image reconstruction

The PAT image reconstruction aims to determine the initial pressure rise from a series of collected acoustic signals. The precise reconstruction formulas for planar, cylindrical and spherical detection surfaces are rigorously deduced in [40]. However, these expressions are difficult to execute since they include many integrations or series summations. Therefore, a simplified

Biomedical Optics EXPRESS

temporal domain reconstruction algorithm, the so-called universal BP algorithm, has been derived in [35]. The initial pressure increment $p_0(\mathbf{r}_Q)$ can be obtained using this algorithm as [12,35].

$$p_0(\mathbf{r}_Q) = \frac{1}{\Omega'} \int_{\Omega'} b(\mathbf{r}_{SR}, t) d\Omega' \Big|_{t=\frac{|\mathbf{r}_{SR}-\mathbf{r}_Q|}{v}} \tag{2}$$

where, Ω' is the total solid angle subtended by the whole measurement surface S' (for planer surface $\Omega' = 2\pi$ and for cylindrical, spherical surfaces $\Omega' = 4\pi$). Here, $b(\mathbf{r}_{SR}, t)$ is related to the detector position $\mathbf{r}_{SR}$ and it is known as the BP term. It is expressed as [12,35],

$$b(\mathbf{r}_{SR}, t) = 2p(\mathbf{r}_{SR}, t) - 2t\frac{\partial p(\mathbf{r}_{SR}, t)}{\partial t}. \tag{3}$$

A schematic is included in Fig. S2(b) to illustrate the symbols. The notation $\mathrm{d}\Omega'$ represents the solid angle subtended by the detection element $\mathrm{d}S'$ at a reconstruction point $\mathbf{r}_Q$ and is defined as [12],

$$d\Omega' = \frac{dS'}{|\mathbf{r}_{SR} - \mathbf{r}_Q|^2} \left[\mathbf{n}_0^{S'} \cdot \frac{(\mathbf{r}_{SR} - \mathbf{r}_Q)}{|\mathbf{r}_{SR} - \mathbf{r}_Q|} \right] \tag{4}$$

whereas $\mathbf{n}_0^{S'}$ is the normal to the surface S'.

It is worthy to emphasize here that $\mathbf{r}_{SR}$ needs to be known precisely for faithful implementation of the BP algorithm. It is required for accurate calculations of the BP term as well as the solid angle. The BP term is very very sensitive (phase of the signal is involved) and any small deviation in $\mathbf{r}_{SR}$ can lead to poor image reconstruction. The ML and DL methods can work remarkably well even for approximate $\mathbf{r}_{SR}$.

2.3. U-Net architecture

The U-Net approach was primarily developed for image segmentation tasks [41]. A representative diagram is presented in Fig. S3. The U-Net design is often composed of three parts: a contracting path or encoder, a bridge, and an expanding path or decoder. The addition of skip connections in the U-Net design distinguishes it from traditional CNNs such as VGG-16, VGG-19, autoencoder and so on [42]. It allows the network to propagate gradient information across distant feature maps. The contractive path is made up of a series of convolution operations using the rectified linear unit (ReLU) as the non-linear activation function and max-pooling layers. Convolution layers help to extract high-level features, whereas max-pooling layers aid in the reduction of the spatial dimensions of the input image. So as the contractive path advances, the image dimension decreases and the number of features increases. Next, the bridge part acts as an intermediary between the contractive and expansive paths. It also consists of multiple convolution operations (with ReLU as the activation function). The expansive path seeks to reconstruct an output image by recovering spatial information. It consists of upsampling and convolutional layers (with ReLU as the activation function). The image dimension gradually increases and the count of features decreases. Upsampling is done with transposed convolution, which helps in the recovery of lost spatial information. The orientation of these two paths and their linkage via skip connections in matching layers result in a symmetrical design resembling the letter U.

3. Materials and methods

3.1. Simulation method

Deep learning is data-hungry. A large amount of data is required to train it efficiently vis-a-vis improve its performance. Generally, the training dataset optimizes the model. The ground truth and the corresponding reconstructed images are included within the training dataset for a test case.

Biomedical Optics EXPRESS

Here, the k-Wave toolbox was utilized to generate the simulated PAT dataset [37]. The PAT simulation setup is shown in Fig. S4(a). Three different binary numerical phantoms, namely two-point, multi-ellipse, and vasculature phantoms, were used in this work [see Fig. S4(b)-(d)]. Each image is accompanied by a colorbar, quantifying the gray levels. The computational domain was divided into 1001×1001 grid points with spacing, $dx = dy = 0.1$ mm. A perfectly matched layer was attached to the computational domain from the outside (thickness = $10dx$). The medium was considered to be acoustically homogeneous- speed of the sound, $v = 1500$ m/s and density, $\rho = 1000$ kg/m^3. Ideal point detectors were uniformly placed between 0^o to 360^o to capture the PA signals (central frequency 2.25 MHz and 70% bandwidth). The SR was fixed to be 43.75 mm. The signal-to-noise ratio (SNR) was maintained at 55 dB while running the forward simulation. The sampling interval was 20 ns.

The universal BP algorithm was used to generate the reconstructed images. The SR was randomly varied between 41.55 and 45.95 mm (43.75 mm ± 5%) and accordingly, 400 images for each phantom at 400 different SRs were reconstructed. A total of 1200 heterogeneous reconstructed images for three phantoms contributed to our dataset. For the sake of simplicity, the first term of the Eq. (3) is considered during the implementation of Eq. (2).

3.2. Experimental method

Three experimental phantoms were used to collect the PA signals as well. These phantoms were designed to be analogous to the numerical phantoms and are shown in Fig. 1. Each had a gelatin base (8% gelatin in milk). The first phantom contained two vertically placed pencil leads (with diameter 0.4 mm) at distances of 0 and 6 mm from the center [Fig. 1(a)]. The top surfaces of the pencil leads and the gelatin base were kept at the same level. The second and third samples had multi-ellipse and vasculature structures printed on transparent sheets using a desktop printer (HP LASERJET 30A; ink- black noire, tinta negra). These sheets were then glued onto the gelatin base, as shown in Fig. 1(b) and Fig. 1(c), respectively.

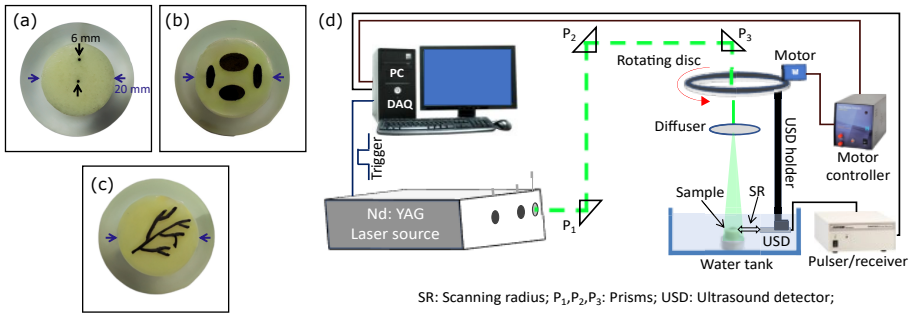

Fig. 1. Schematic representation of the experimental setup.

The experimental setup that was used for collecting forward data is illustrated in Fig. 1(d). A Q-switched Nd:YAG laser (model NT352C-10-SH-H, Ekspla) emitting 532 nm wavelength beam was employed. The light pulses had a duration of 6 ns and a repetition rate of 10 Hz with ≈ 7 mJ/cm^2 fluence per pulse. The fluence was well under the ANSI safety limit, ensuring that the experiments were conducted with utmost care and consideration for human safety [43]. The laser beam was directed through three right-angle uncoated prisms (designated as P1, P2 and P3 in Fig. 1) and one uncoated plano-concave lens (diffuser) in order to reach the sample. The resulting PA signals were captured by a single-element ultrasonic transducer (V325-SU, Panametric) with a center frequency of 2.25 MHz, a fractional bandwidth of 70%, and a diameter of 10 mm. The detected signals were then amplified with 55 dB gain using a pulser/receiver (DPR300, JSR Ultrasonics) and stored via a data acquisition card (PCIe-9852, ADLINK) at a

Biomedical Optics EXPRESS

sampling frequency of 50 MHz. The ultrasonic transducer was set to rotate at a speed of 0.5 degree/s, with a scanning radius of approximately 43.75 mm. The experimental setup utilized a customized scanning system (Holmarc, India). To reconstruct images, the universal BP algorithm was employed. For each phantom, a total of 21 PAT images were generated by assigning 21 different values of SR between 39.35 to 48.15 mm (43.75 mm ± 10%).

3.3. Network architecture

3.3.1. Image prediction using dense U-Net

Figure 2 illustrates the dense U-Net (DUNet) architecture that we used to tackle our contextual problem with certain changes compared to [31]. It received an input image with a size of 128 × 128 pixels and produced an output image of the same dimensions. In the first layer of the contracting path, the input image underwent a 2D convolution operation with a 3 × 3 kernel to obtain 16 feature maps from the input image depth of 1. Afterward, a dense block was implemented to operate on the f feature maps as input, which yielded an output of 2f feature maps, with a growth rate of f/4. That is, it extracted 32 features from 16 feature maps. The architecture of the dense bock is shown in Fig. 3. Four layers were present in each dense block, which executed a series of 2D convolution operations using a 1 × 1 kernel and a 2D convolution operation with a 3 × 3 kernel. The final output of the dense block was connected to all previous convolutional layers through the concatenation of their outputs, resulting in a dense connection within the block. Instead of using max-pooling to reduce the dimensionality of the image, we employed a downsample after each dense block. This downsample consisted of a 2D convolution operation with a 1 × 1 kernel (with a stride 1) and another 2D convolution operation with a 3 × 3 kernel (with a stride of 2). This allowed the image dimensionality to be downscaled by half while maintaining the important features. This process was repeated in the encoder until the number of feature maps reaches 1024 and the dimensionality of the image was reduced to 4 × 4.

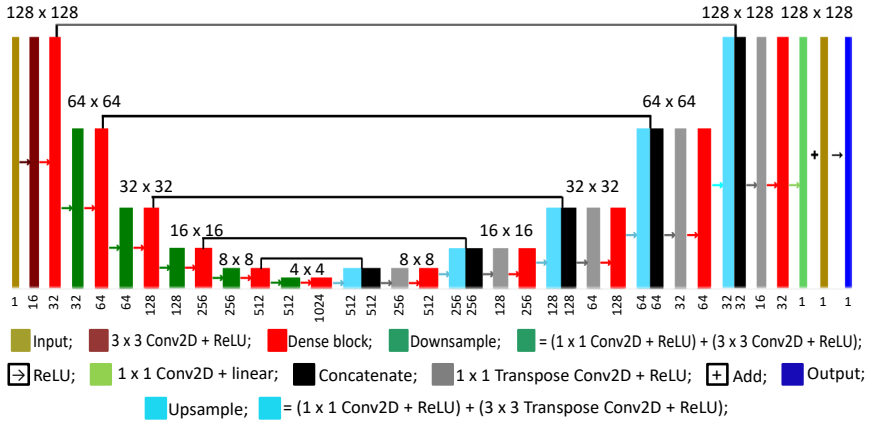

Fig. 2. Demonstration of the dense U-Net (DUNet) architecture.

At the beginning of the decoder section, the 4 × 4 image with 1024 feature maps was upsampled. This upsample operation included a 2D convolution operation with a 1 × 1 kernel (with a stride of 1) and a 2D transposed convolution operation with a 3 × 3 kernel (with a stride of 2), which doubled the image dimensionality while halving the feature size. After the upsampling operation, a skip connection was established between the corresponding output of the dense block in the same layer of the encoder path and the current layer in the decoder part. Then the image underwent a 2D convolution operation with a 1 × 1 kernel and reduced the feature by a factor of 4. After that, it went through a dense block which increased the number of feature maps by a

Biomedical Optics EXPRESS

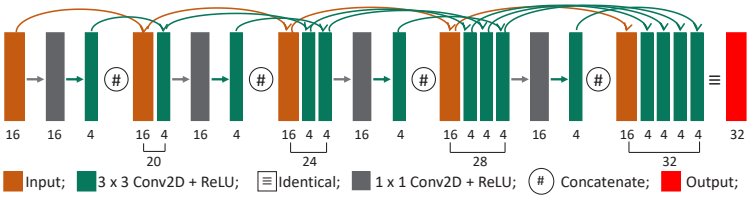

Fig. 3. Graphical elaboration of dense block of the first layer of the DUNet.

factor of 2. And the process continued until the image size increased to 128 × 128 with feature map 32. ReLU was used as an activation function in the convolution operations. In the next step, the network went through another 2D convolution operation with a 1 × 1 kernel and a linear activation function, which produced an output image with dimensions of 128 × 128 and a depth of 1. This output image was then added to the initial input image, and the result was passed through the ReLU activation function to speed up the training process [31,44]. It was seen that the performance of this network was acceptable when trained and tested for the same phantom. But it failed for heterogeneous phantoms in some situations.

3.3.2. Image classification with U-Net and prediction using dense U-Net

In order to address this issue, we utilized two distinct U-Net architectures. First U-Net network named UNet$_C$ (shown in Fig. 4). It took an input image with a size of 128 × 128 pixels and produced class-level output (here class level was 3). Initially, in the contracting path, the input image was processed through a pair of consecutive 2D convolution operations with a 3 × 3 kernel to obtain 32 feature maps from an input image depth of 1. Then, it was downsampled using a 2D max-pooling operation with a stride of 2, resulting in a 50% reduction in image dimensionality while retaining the same number of features (i.e., image size 64 × 64 and 32 features). The process was then continued iteratively, and in the bridge part the image size was reduced to 16 × 16, and the maximum feature maps reached 256. At the beginning of the expansive path, a 2D transposed convolution operation with a 2 × 2 kernel was applied to the 16 × 16 image that had a feature map of 256. This operation resulted in doubling the image dimensions to 32 × 32 while halving the number of features. Then, a skip connection was established between the output of the transposed convolution and the corresponding output of the same layer in the contracting path. Similarly, the process was repeated until the image size reached 128 × 128 pixels with 32 feature maps. The ReLU activation function was used in the convolution operations. Finally, the feature map output of the last convolutional layer was flattened into a 1D array, and this was connected to three fully connected dense layers (with softmax as an activation function) to produce class-level output for image classification.

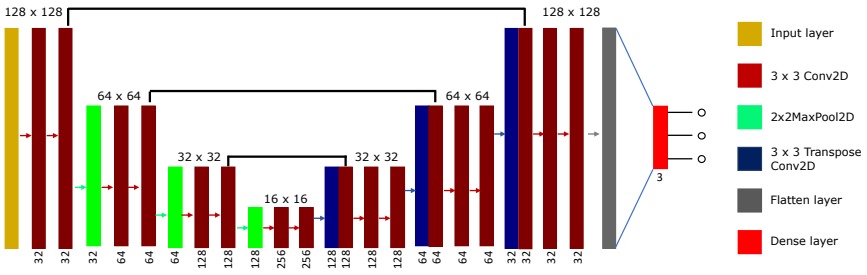

Fig. 4. Illustration of UNet$_C$- image classification by U-Net model.

Biomedical Optics EXPRESS

Now we had another three models whose architecture was the same DUNet that had been discussed in para 3.3.1. However, instead of training them with heterogeneous images, this time we trained them differently. The two-point dense U-Net ($DUNet_{TP}$) was trained solely with two-point simulated reconstructed images along with the ground truth. The multi-ellipse dense U-Net ($DUNet_{ME}$) and vasculature dense U-Net ($DUNet_{VA}$) were trained with corresponding simulated reconstructed images and their respective ground truths. The workflow for this approach is illustrated in Fig. 5), where we first applied $UNet_C$ on the testing dataset and stored them in three different arrays (first, second and third arrays for two-point, multi-ellipse and vasculature phantoms, respectively). Next, we employed models $DUNet_{TP}$, $DUNet_{ME}$, and $DUNet_{VA}$ to these arrays, respectively, for image generation. This approach allowed for a more specialized model to be applied to specific test data, resulting in improved reconstruction quality.

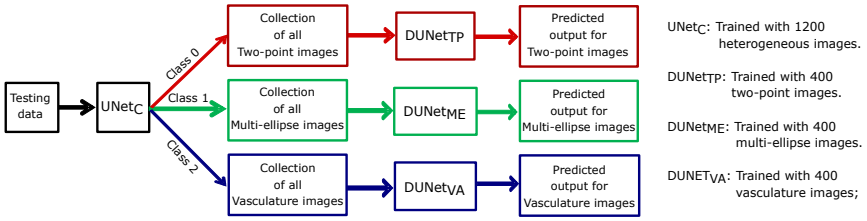

Fig. 5. Block diagram detailing image classification and subsequent image formation.

3.3.3. Network optimization and generation of dataset

In order to train the DUNet, $DUNet_{TP}$, $DUNet_{ME}$, and $DUNet_{VA}$ architectures, the Adam optimizer, with its default setting, was used as the optimizer; Glorot normal was taken as the kernel (weight) initializer; mean absolute error (MAE) was considered as the loss function; and the batch size and epoch were set to 8 and 100, respectively. A dropout layer (20%) was included in every dense block, and an early stopping callback with the patience of 5 was used to reduce overfitting during the network training, i.e., the training process would be stopped if the loss on the validation data did not improve for 5 consecutive epochs. On the other hand, the $UNet_C$ architecture underwent three changes: first, sparse categorical crossentropy was used as the loss function. This loss function is commonly used for multi-class classification problems; second, the number of epochs was set to 20; and lastly, no dropout layer had been used.

The DUNet was trained with $400 \times 3 = 1200$ heterogeneous PAT images generated via the BP algorithm utilizing simulated PA signals. The GT for each phantom was also included. After that 90% of the population (randomly chosen) built the training dataset. The remaining images were part of the validation dataset. The training dataset optimized the CNN network and the validation dataset tuned the hyperparameters. The $DUNet_{TP}$ was trained with 400 images of the two-point phantom following the same procedure. Similarly, $DUNet_{ME}$ and $DUNet_{VA}$ were trained. Two types of testing datasets were generated to evaluate the performance of the CNN architectures. For simulated dataset comprising of $21 \times 3 = 63$ images, the numerical values of SNR and SR were randomly varied between 35 to 55 dB and 39.35 to 48.15 mm (43.75 mm ± 10%), respectively while generating the PAT images for three phantoms. None of which were included in the training set. The BP scheme was employed to form the test images using the simulated and experimental signals. The BP images fed to various CNNs as inputs so that the networks may predict distortion free images. Another testing dataset contained 63 experimental images. The computations were performed in a personal computer with a Windows 11, 64-bit architecture, 16 GB of RAM, AMD Ryzen-5 4600H CPU and 4 GB NVIDIA GeForce GTX 1650 GPU with 896 CUDA cores. The CNNs were developed utilizing Python 3.9.13 and TensorFlow v2.10 packages [45].

Biomedical Optics EXPRESS

3.3.4. Performance evaluation of the proposed networks

The performance of the DUNet and UNet$_C$ + DUNet had been evaluated exhaustively and systematically using images generated via simulations and experiments. The performance of the Davoudi's UNet (details can be found in [30]) was also included for comparison. Three CNN architectures were trained on the same training dataset and their performances were examined on the same testing dataset. Reconstructed images through the CNN architectures and BP algorithm were also quantified using the peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM). A PSNR value above 30 dB is typically acceptable, indicating a close resemblance between the ground truth and reconstructed images, while a desirable SSIM value approaches one [46]. The values were evaluated using the default scikit-image library [47].

4. Simulation and experimental results

4.1. Performance on simulated dataset

Figure 6 demonstrates representative images for the two-point phantom. Each reconstructed image is normalized by its highest pixel value. A color bar attached to each image indicates the numerical values of the gray levels. Figure 6(a) contains the GT (target image) for the two-point phantom. The images presented in the first and second rows have been generated at SR=41.11 and 45.95 mm, respectively (i.e., 6% and 5% away from the actual SR=43.75 mm, respectively). The noise level present in the RF signals is about 38.70 dB for the first row whereas the same is approximately 45.75 dB for the second row. The first, second, third and fourth columns contain the recreated images for the BP, Davoudi's UNet, DUNet and UNet$_C$ + DUNet$_{TP}$ techniques, respectively. Note that the BP image acts as the input for the CNN protocols. Ideally the output of any reconstruction algorithm or any DL network should resemble Fig. 6(a). The computed values of PSNR and SSIM are inserted in each image. It can be seen from Fig. 6(b) and (f) that the BP algorithm fails to produce the GT as the SRs are deviated from the original value. They also contain sufficient amount of streak artifacts. Figure 6(c) and (g) illustrate that Davoudi's UNet can perfectly recover the GT. The DUNet produces an overlapping image of multi-ellipse and vasculature phantoms when SR=41.11 mm but it obtains a correct structure at SR=45.75 mm [Fig. 6(d) and (h)]. The UNet$_C$ + DUNet$_{TP}$ method, Fig. 6(e) and (i), can form the light absorbing structures well but they are faintly accompanied by ring artifacts. The average values of the PSNR and SSIM for the proposed method are greater than the other methods (see rows 3 to 5, columns 2 and 5, respectively of Table 1). The predicted results (i.e., right or wrong image formation) for the CNNs are presented in Table 2 for a series of scanning radii with arbitrary noise levels (randomly varying between 35 to 55 dB in the RF lines). The third, fourth and fifth columns of the table detail the outcomes of the networks (rows 4 to 24).

Table 1. Quantitative values of various parameters to compare the performance of different networks for simulated dataset. NA represents not applicable.

Method	Average PSNR			Average SSIM			Epoch	Time (min:sec)
	Two-Point	Multi-ellipse	Vasculature	Two-Point	Multi-ellipse	Vasculature		
Davoudi's UNet	30.48	26.00	23.64	0.88	0.86	0.89	41	16:59
DUNet	28.83	36.17	23.79	0.84	0.94	0.81	65	25:08
UNet$_C$ + DUNet$_{TP}$	33.54	NA	NA	0.95	NA	NA	20 + 33 = 53	08:17
UNet$_C$ + DUNet$_{ME}$	NA	34.53	NA	NA	0.97	NA	20 + 39 = 59	09:29
UNet$_C$ + DUNet$_{VA}$	NA	NA	34.07	NA	NA	0.97	20 + 37 = 57	08:57

Next the proposed method has been evaluated on the simulated dataset for the multi-ellipse phantom and some illustrative images are presented in Fig. 7. The SRs are the same as that of the previous case but the noise levels are 48.91 and 44.59 dB, respectively. As also found in the earlier case, the BP algorithm fails to produce a faithful reconstruction of the GT at both the SRs

Biomedical Optics EXPRESS

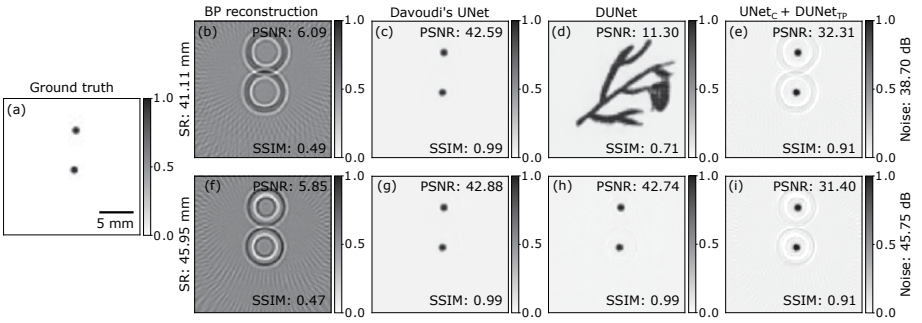

Fig. 6. Reconstruction results using the simulated images for different approaches for the two-point phantom. Each image, from (b) to (i), represents a normalized reconstructed/predicated image and contains a colorbar quantifying the gray shades. The noise level in the RF signals is higher in the top row than that of the bottom row. The SR is 6% lesser compared to its nominal value (43.75 mm) for the top row and 7% more for the bottom row. The calculated values of the image quality parameters (PSNR and SSIM) are embedded on each image.

Table 2. Performance comparison of different models on the testing dataset; where (R,W): (right,wrong) prediction for Davoudi's UNet; (*R,W*): (right,wrong) prediction for DUNet; (r,w):(right,wrong) prediction for UNet$_C$+DUNet.

	Two-Point							Multi-ellipse							Vasculature						
	Simulated				Experimental			Simulated				Experimental			Simulated				Experimental		
SR(mm)	Noise(dB)	Prediction			Prediction			Noise(dB)	Prediction			Prediction			Noise(dB)	Prediction			Prediction		
39.35	38.06	**W**	*W*	**w**	**W**	*W*	**w**	45.07	**W**	*W*	**w**	**W**	*W*	**w**	47.08	**W**	*W*	**w**	**W**	*W*	**w**
39.79	45.06	**W**	*W*	**w**	**W**	*W*	**r**	52.76	**W**	*W*	**w**	**W**	*W*	**r**	42.34	**W**	*W*	**w**	**W**	*W*	**w**
40.23	48.26	**W**	*W*	**w**	**W**	*W*	**r**	37.72	**W**	*R*	**w**	**W**	*W*	**r**	37.24	**W**	*W*	**w**	**W**	*W*	**r**
40.67	54.36	**W**	*W*	**w**	**R**	*R*	**r**	35.74	**W**	*R*	**r**	**W**	*R*	**r**	52.99	**W**	*W*	**r**	**W**	*W*	**r**
41.11	38.70	**R**	*W*	**r**	**R**	*R*	**r**	48.91	**W**	*R*	**r**	**W**	*R*	**r**	48.43	**W**	*W*	**r**	**W**	*R*	**r**
41.55	47.84	**R**	R	**r**	**R**	*R*	**r**	36.38	**R**	*R*	**r**	**W**	*R*	**r**	54.00	**W**	*W*	**r**	**W**	*R*	**r**
41.99	52.48	**R**	*R*	**r**	**R**	*R*	**r**	54.66	**R**	*R*	**r**	**R**	*R*	**r**	39.65	**R**	*R*	**r**	**W**	*R*	**r**
42.43	45.10	**R**	*R*	**r**	**R**	*R*	**r**	52.44	**R**	*R*	**r**	**R**	*R*	**r**	42.00	**R**	*R*	**r**	**W**	*R*	**r**
42.87	52.63	**R**	*R*	**r**	**R**	*R*	**r**	48.11	**R**	*R*	**r**	**R**	*R*	**r**	52.34	**R**	*R*	**r**	**W**	*R*	**r**
43.31	46.36	**R**	*R*	**r**	**R**	*R*	**r**	43.54	**R**	*R*	**r**	**R**	*R*	**r**	48.21	**R**	*R*	**r**	**W**	*R*	**r**
43.75	54.00	**R**	*R*	**r**	**R**	*R*	**r**	53.73	**R**	*R*	**r**	**R**	*R*	**r**	53.22	**R**	*R*	**r**	**W**	*R*	**r**
44.19	49.78	**R**	*R*	**r**	**R**	*R*	**r**	35.41	**R**	*R*	**r**	**R**	*R*	**r**	54.27	**R**	*R*	**r**	**W**	*R*	**r**
44.63	44.85	**R**	*R*	**r**	**R**	*R*	**r**	39.49	**R**	*R*	**r**	**R**	*R*	**r**	39.77	**R**	*R*	**r**	**W**	*R*	**r**
45.07	54.34	**R**	*R*	**r**	**R**	*R*	**r**	49.09	**R**	*R*	**r**	**R**	*R*	**r**	45.76	**R**	*R*	**r**	**W**	*W*	**r**
45.51	45.92	**R**	*R*	**r**	**W**	*R*	**r**	49.11	**R**	*R*	**r**	**R**	*R*	**r**	45.90	**R**	*R*	**r**	**W**	*W*	**r**
45.95	45.75	**R**	*R*	**r**	**W**	*W*	**r**	44.59	**R**	*R*	**r**	**W**	*R*	**r**	52.58	**R**	*R*	**r**	**R**	*W*	**r**
46.39	40.34	**R**	*R*	**r**	**W**	*W*	**w**	51.41	**R**	*R*	**r**	**W**	*R*	**r**	54.21	**W**	*W*	**w**	**W**	*W*	**w**
46.83	39.34	**W**	*W*	**w**	**W**	*W*	**w**	45.75	**W**	*R*	**r**	**W**	*W*	**r**	41.26	**W**	*W*	**w**	**W**	*W*	**w**
47.27	52.29	**W**	*W*	**w**	**W**	*W*	**w**	50.30	**W**	*R*	**r**	**W**	*W*	**r**	47.40	**W**	*W*	**w**	**W**	*W*	**w**
47.71	58.64	**W**	*W*	**w**	**W**	*W*	**w**	54.34	**W**	*W*	**w**	**W**	*W*	**r**	50.55	**W**	*W*	**w**	**W**	*W*	**w**
48.15	51.11	**W**	*W*	**w**	**W**	*W*	**w**	36.35	**W**	*W*	**w**	**W**	*W*	**w**	52.45	**W**	*W*	**w**	**W**	*W*	**w**

Biomedical Optics EXPRESS

[compare Fig. 7(b) and (f) with Fig. 7(a)]. The Davoudi's UNet breakdowns in the former case but works accurately in the later situation [see Fig. 7(c) and (g), respectively]. The DUNet and UNet$_C$ + DUNet$_{ME}$ yield the correct images [see Fig. 7(d) and (h); Fig. 7(e) and (i)]. The overall performance of these methods are comparable (see rows 4 and 5, columns 3 and 6, respectively of Table 1). An exhaustive analysis of the image-formation schemes is provided in Table 2 at different SRs and SNRs (see rows 4 to 24 and columns 10 to 12).

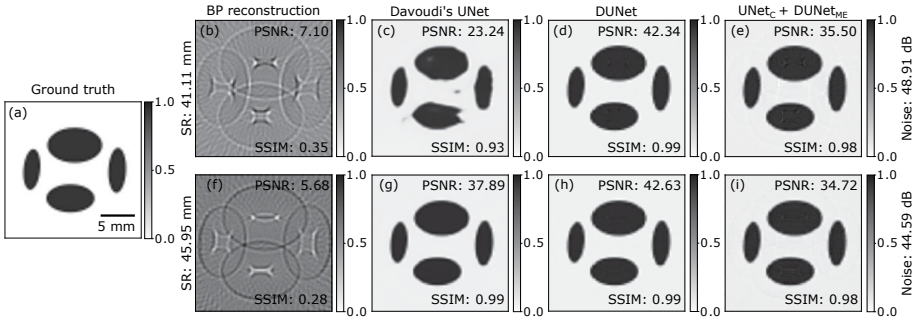

Fig. 7. Performance of various methods on the simulated images for the multi-ellipse phantom.

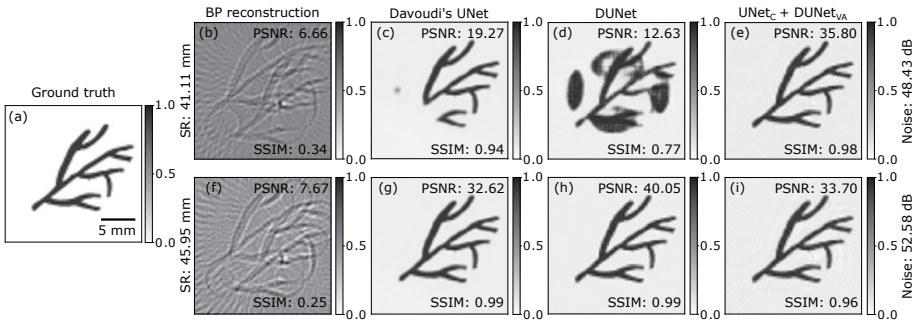

Fig. 8. Performance of various methods on the simulated images for the vasculature phantom.

Finally, the normalized reconstructed images for the vasculature phantom are displayed in Fig. 8. As expected, the BP images are far away from the GT. The Davoudi's network and DUNet are unable to recreate the GT when the SR is set to less than the actual value [see Fig. 8(c) and (d)]. However, the prediction of the third method is correct at the same SR [see Fig. 8(d)]. The performance of the all networks are acceptable as shown in Fig. 8(g) to (i) for this test case (the SR is greater than the true value). In general, the third method outperforms the others as seen from rows 3, 4 and 7, columns 4 and 7 of Table 1. Table 2 details the performance of the CNN protocols considered in this study. The third method is found to be valid for a wide range of SR values varying from 40.67 to 45.95 mm.

4.2. Performance on experimental dataset

Figure 9 exhibits typical images generated by various techniques using the experimental dataset collected for the two-point phantom. The photograph of the sample is inserted in Fig. 9(a) as a ready reference. The BP image consists of multiple concentric circles corresponding to each point source as evident from Fig. 9(b) and (f). Figures 9(c) and (d) reveal that the shapes

Biomedical Optics EXPRESS

are not exactly reproduced. In addition to that Fig. 9(d) retains some unwanted gray shades. Figure 9(e) elucidates that the circular structures are retained but they are surrounded by whitish rings. The first two methods produce wrong results when SR=45.95 mm [see Fig. 9(g) and (h)]. Figure 9(i) looks similar to Fig. 9(e). Table 2 portrays how these CNN frameworks work under various combinations of SR and SNR values (rows 4 to 24, columns 6 to 8). It is clear that the UNet$_C$+DUNet$_{TP}$ has larger validity domain than the other methods.

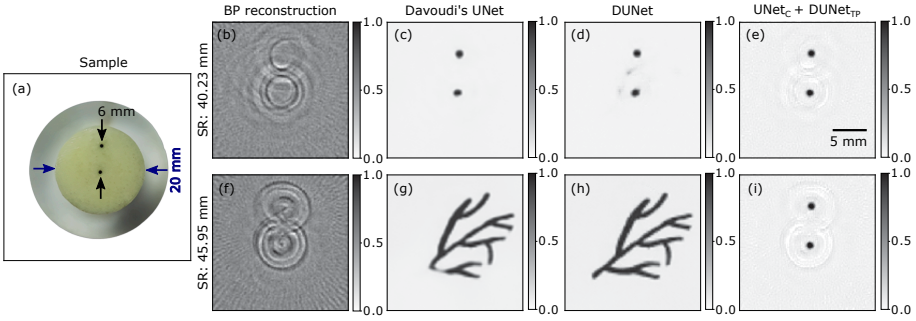

Fig. 9. Reconstruction results on experimental images for different approaches for the two-point sample.

Representative normalized images along with the GT for the multi-ellipse phantom are shown in Fig. 10. The geometries of the source structures are not visible from the BP images [see Fig. 10(b) and (f)]. The Davodi's UNet cannot restore the GT for both the SR values [see Fig. 10(c) and (g)]. The DUNet marginally fails in the first case but works fine in the second case [see Fig. 10(d) and (h)]. The UNet$_C$+DUNet$_{ME}$ in both the cases accurately predicts the GT. The explicit delineation can be found in Table 2 for various settings (rows be 4 to 24, columns 13 to 15).

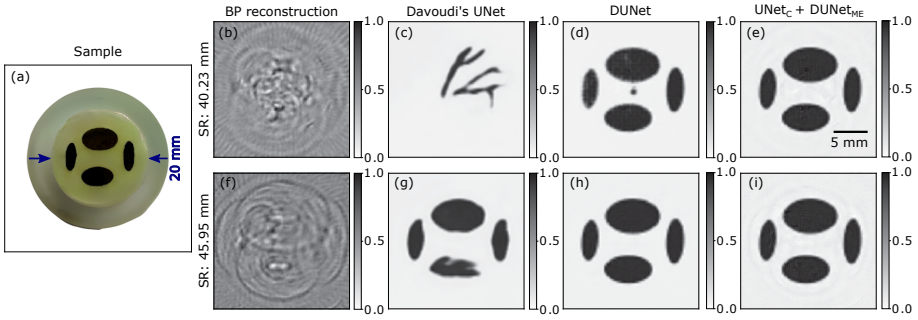

Fig. 10. Reconstruction results on experimental images for different approaches for the multi-ellipse sample.

The findings for the third experimental phantom are visualized in Fig. 11. The GT, normalized images accompanying colorbars are inserted in this figure. The BP technique cannot regenerate the GT [see Fig. 10(b) and (f)]. Table 2 states that Davodi's UNet always fails to reconstruct the GT (except one case- see row 19, column 20). The performance of the DUNet is good in a limited region when SR is varied from 41.11 to 44.63 mm, beyond this range it breaks down. The validity domain of UNet$_C$+DUNet$_{VA}$ is the largest among all the methods. It remains valid in the range of SR=40.23 to 45.95 mm establishing robustness of the method.

Biomedical Optics EXPRESS

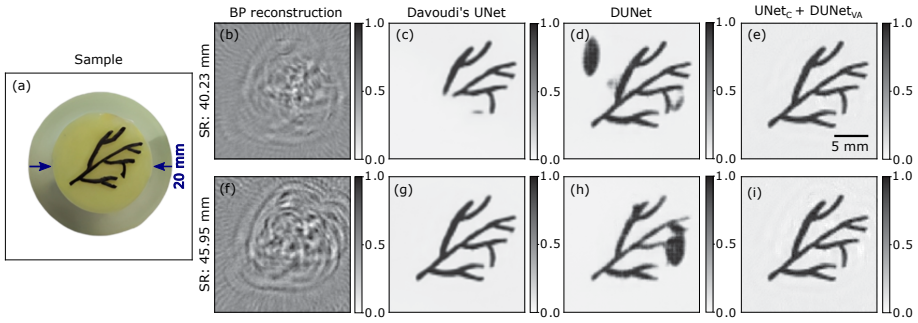

Fig. 11. Reconstruction results on experimental images for different approaches for the vasculature sample.

5. Discussion and conclusions

Accurate image reconstruction is crucial for extracting meaningful information from acquired PA signals. Analytical methods such as the universal BP algorithm and its variants offer simplicity and real-time capabilities. However, these methods have limitations in terms of reconstruction quality, artifact correction, and applicability to complex geometries. On the other hand, a model-based approach uses regularization methods that can suppress noise, enhance image details, and results more interpretable PAT images. But this approach comes with increased computational complexity. Accurate SR is required in both approaches for proper image formation, as discussed earlier. A DL approach involving a pre-processing or post-processing method can be employed. Although the first method is effective, it incurs computational complexity and requires careful parameter tuning [28,48]. Moreover, for a pre-processing approach, the relevant parameter is obtained as a first step, then a reconstruction algorithm is utilized for image creation (second step). The reconstruction algorithm may introduce some errors in the image even for exact SR too because of its limitations. Here, we have implemented a post-processing CNN approach to alleviate the SR issue and improve the quality of the reconstructed images.

In this manuscript, a modified DUNet architecture is applied. The architecture was trained with 1200 heterogeneous simulated images, and its performance was checked with both the simulated and experimental images. Table 2 demonstrates that the DUNet can tolerate SR deviation at least $\approx 43.75 \pm 4.5\%$ considering all simulated phantoms. Similarly, for all experimental images, this range found to be $\approx 43.75 \pm 4\%$. Beyond these regimes, the network fails to perform well. It produces either overlapping-type images or incomplete structures. Previous studies also reported that training and testing with well-matched images yield promising results but struggle when dealing with different-like images [26,31].

Consequently, we pursued the second architecture UNet_C + DUNet. It contained a three-layered UNet architecture trained with 1200 heterogeneous images and classified the testing images into three categories based on their structures. Afterwards, DUNet was individually applied to the three corresponding classified testing images. Prior to this, it was trained separately using two-point, multi-ellipse, and vasculature images. Table 2 illustrates that $\approx 43.75 \pm 5.5\%$ is the minimum acceptance of SR deviation for all the simulated testing images and $\approx 43.75 \pm 7.5\%$ for experimental images. It can be noticed that acceptance of SR deviation for this method is much higher than other CNN approaches. Moreover, Table 1 shows that this network takes very less time (approx. 9 min) to predict the reconstructed images. It may be expected that CNN networks trained with simulated data and tested on experimental or real-time images may not perform satisfactorily [49]. However, we found that the proposed network works better on experimental

Biomedical Optics EXPRESS

images than on simulated images. Further investigation is required to assess the performance of this approach on in vivo dataset.

It can be stated from the graphs presented in Fig. 12 that all the models considered in this study have been trained well by the dataset. This might be due to the fact that the training images were generated only for ±5% variation of SR as well as the noise level was kept constant at 55 dB. Accordingly, it may be assumed that the BP images perhaps retained crucial features of the targets to some extent which resulted in satisfactory training convergence of the networks. However, the testing dataset contained images having SR variation up to ±10% as well as the noise level randomly altered in each testing image between 35 to 55 dB. Therefore, many BP images were severely distorted which acted as the inputs to the CNN models. The first two networks could not overcome these distortions and produced either incomplete or overlapping structures. Nevertheless, it is apparent that the third framework (UNet$_C$+DUNet) mitigates this issue greatly providing acceptable images over a large range. It can be verified from Table 2 as well. It may be emphasized that in this work, image augmentation (i.e., rotation, flipping, zooming, shrinking, shifting etc.) has not been incorporated while training the networks. We anticipate that the proposed model will work faithfully if it is trained with a large number of images and when the network-parameters are suitably tuned. It will be interesting to undertake this project in future.

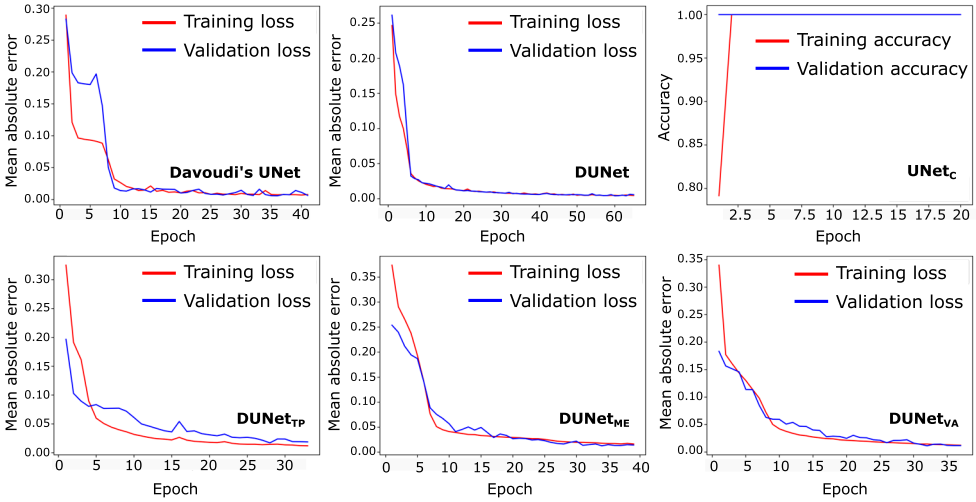

Fig. 12. (a)-(b) Plots of mean absolute error (MAE) for different UNETs during training and validation stages for each epoch. (c) Demonstration of variation of accuracy as a function of epoch. (d)-(f) Same as (a) but for the DUNet$_{TP}$, DUNet$_{ME}$, DUNet$_{VA}$, respectively.

In conclusion, this work indicates that the DL approaches can mitigate the necessity of accurate estimation of SR for faithful reconstruction of the GT in the context of PAT imaging. We propose a DUNet, which uses the benefits of dense blocks. It can recover the GTs correctly even when the measurement of SR remains uncertain approximately by 5%. This network encounters challenges to produce reliable image reconstruction if SR is varied further. To lessening this issue, we develop UNet$_C$+DUNet, which utilizes the classification method and then employs the DUNet to produce the required image. The simulation and experimental results confirm that it can tolerate SR fluctuation up to approximately 8% and can precisely recover the GTs within this range for various phantoms. Our future work will be to increase the ability to endure more SR deviation. Advanced DL approaches, like transfer learning (training and testing with two different classes

Biomedical Optics EXPRESS

of images), can be utilized for enhancing the performance of classification and reconstruction process.

Funding. Indian Council of Medical Research (56/2/2020-Hae/BMS); Department of Biotechnology, Ministry of Science and Technology, India (BT/PR44547/MED/32/791/2021).

Acknowledgments. The authors thank the members of the Biomedical Imaging Laboratory (BMIL) at IIIT Allahabad for their continuous support.

Disclosures. The authors state no conflicts of interest regarding this article.

Data availability. The dataset used for training and testing purposes can be obtained from the authors upon reasonable request.

Supplemental document. See Supplement 1 for supporting content.

References

1. F. Lucka, M. Pérez-Liva, B. E. Treeby, and B. T. Cox, "High resolution 3d ultrasonic breast imaging by time-domain full waveform inversion," Inverse Probl. **38**(2), 025008 (2022).
2. I. Yamaga, N. Kawaguchi-Sakita, and Y. Asao, *et al.*, "Vascular branching point counts using photoacoustic imaging in the superficial layer of the breast: a potential biomarker for breast cancer," Photoacoustics **11**, 6–13 (2018).
3. X. Yang, A. Maurudis, J. Gamelin, A. Aguirre, Q. Zhu, and L. V. Wang, "Photoacoustic tomography of small animal brain with a curved array transducer," J. Biomed. Opt. **14**(5), 054007 (2009).
4. J. Tang, J. E. Coleman, X. Dai, and H. Jiang, "Wearable 3-d photoacoustic tomography for functional brain imaging in behaving rats," Sci. Rep. **6**(1) 1–10 (2016).
5. K. Jansen, G. van Soest, and A. F. van der Steen, "Intravascular photoacoustic imaging: a new tool for vulnerable plaque identification," Ultrasound Med. & Biol. **40**(6), 1037–1048 (2014).
6. J. Prakash, S. K. Kalva, M. Pramanik, and P. K. Yalavarthy, "Binary photoacoustic tomography for improved vasculature imaging," J. Biomed. Opt. **26**(08), 086004 (2021).
7. W. J. Akers, W. B. Edwards, C. Kim, B. Xu, T. N. Erpelding, L. V. Wang, and S. Achilefu, "Multimodal sentinel lymph node mapping with single-photon emission computed tomography (spect)/computed tomography (ct) and photoacoustic tomography," Transl. Res. **159**(3), 175–181 (2012).
8. X. Yang and L. V. Wang, "Monkey brain cortex imaging by photoacoustic tomography," J. Biomed. Opt. **13**(4), 044009 (2008).
9. L. Nie, X. Cai, K. Maslov, A. Garcia-Uribe, M. A. Anastasio, and L. V. Wang, "Photoacoustic tomography through a whole adult human skull with a photon recycler," J. Biomed. Opt. **17**(11), 110506 (2012).
10. L. V. Wang, *Photoacoustic imaging and spectroscopy* (CRC press, 2017).
11. L. V. Wang and J. Yao, "A practical guide to photoacoustic tomography in the life sciences," Nat. Methods **13**(8), 627–638 (2016).
12. M. Xu and L. V. Wang, "Universal back-projection algorithm for photoacoustic computed tomography," Phys. Rev. E **71**(1), 016706 (2005).
13. P. Burgholzer, J. Bauer-Marschallinger, H. Grün, M. Haltmeier, and G. Paltauf, "Temporal back-projection algorithms for photoacoustic tomography with integrating line detectors," Inverse Probl. **23**(6), S65–S80 (2007).
14. B. E. Treeby, E. Z. Zhang, and B. T. Cox, "Photoacoustic tomography in absorbing acoustic media using time reversal," Inverse Probl. **26**(11), 115003 (2010).
15. M.-L. Li, Y.-C. Tseng, and C.-C. Cheng, "Model-based correction of finite aperture effect in photoacoustic tomography," Opt. Express **18**(25), 26285–26292 (2010).
16. Y. Hristova, P. Kuchment, and L. Nguyen, "Reconstruction and time reversal in thermoacoustic tomography in acoustically homogeneous and inhomogeneous media," Inverse Probl. **24**(5), 055006 (2008).
17. P. Warbal and R. K. Saha, "In silico evaluation of the effect of sensor directivity on photoacoustic tomography imaging," Optik **252**, 168305 (2022).
18. M.-L. Li and C.-C. Cheng, "Model-based reconstruction for photoacoustic tomography with finite aperture detectors," in *2009 IEEE International Ultrasonics Symposium*, (IEEE, 2009), pp. 2359–2362.
19. L. Lin, P. Hu, J. Shi, C. M. Appleton, K. Maslov, L. Li, R. Zhang, and L. V. Wang, "Single-breath-hold photoacoustic computed tomography of the breast," Nat. Commun. **9**(1), 2352 (2018).
20. L. Li, L. Zhu, and C. Ma, *et al.*, "Single-impulse panoramic photoacoustic computed tomography of small-animal whole-body dynamics at high spatiotemporal resolution," Nat. Biomed. Eng. **1**(5), 0071 (2017).
21. L. Ding, X. L. Dean-Ben, and D. Razansky, "Efficient 3-d model-based reconstruction scheme for arbitrary optoacoustic acquisition geometries," IEEE Trans. Med. Imaging **36**(9), 1858–1867 (2017).
22. L. Ding, D. Razansky, and X. L. Dean-Ben, "Model-based reconstruction of large three-dimensional optoacoustic datasets," IEEE Trans. Med. Imaging **39**(9), 2931–2940 (2020).
23. H. Deng, X. Wang, C. Cai, J. Luo, and C. Ma, "Machine-learning enhanced photoacoustic computed tomography in a limited view configuration," in *Advanced Optical Imaging Technologies II*, vol. 11186 (SPIE, 2019), pp. 52–59.'

Biomedical Optics EXPRESS

24. C. Yang, H. Lan, and F. Gao, "Accelerated photoacoustic tomography reconstruction via recurrent inference machines," in *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, (IEEE, 2019), pp. 6371–6374.
25. L. Li and L. V. Wang, "Recent advances in photoacoustic tomography," BME Frontiers (2021).
26. S. Antholzer, M. Haltmeier, and J. Schwab, "Deep learning for photoacoustic tomography from sparse data," Inverse Probl. Sci. Eng. **27**(7), 987–1005 (2019).
27. 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., Ferroelect., Freq. Contr. **67**(12), 2660–2673 (2020).
28. P. Rajendran and M. Pramanik, "Deep learning approach to improve tangential resolution in photoacoustic tomography," Biomed. Opt. Express **11**(12), 7311–7323 (2020).
29. P. Rajendran and M. Pramanik, "High frame rate (3 hz) circular photoacoustic tomography using single-element ultrasound transducer aided with deep learning," J. Biomed. Opt. **27**(06), 066005 (2022).
30. N. Davoudi, X. L. Deán-Ben, and D. Razansky, "Deep learning optoacoustic tomography with sparse data," Nat. Mach. Intell. **1**(10), 453–460 (2019).
31. 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 Inform. **24**(2), 568–576 (2020).
32. P. Rajendran and M. Pramanik, "Deep-learning-based multi-transducer photoacoustic tomography imaging without radius calibration," Opt. Lett. **46**(18), 4510–4513 (2021).
33. S. K. Kalva, Z. Z. Hui, and M. Pramanik, "Calibrating reconstruction radius in a multi single-element ultrasound-transducer-based photoacoustic computed tomography system," J. Opt. Soc. Am. A **35**(5), 764–771 (2018).
34. P. Rajendran and M. Pramanik, "Multi-transducer photoacoustic tomography imaging without radius calibration using deep learning," in *Photons Plus Ultrasound: Imaging and Sensing 2022*, vol. 11960 (SPIE, 2022), pp. 39–45.
35. P. Warbal and R. K. Saha, "Performance comparison of commonly used photoacoustic tomography reconstruction algorithms under various blurring conditions," J. Mod. Opt. **69**(9), 487–501 (2022).
36. R. Prakash, D. Badal, A. Paul, D. Sonker, and R. K. Saha, "Photoacoustic signal simulation using discrete particle approach and its application in tomography," IEEE Trans. Ultrason., Ferroelect., Freq. Contr. **68**(3), 707–717 (2021).
37. B. E. Treeby and B. T. Cox, "k-wave: Matlab toolbox for the simulation and reconstruction of photoacoustic wave fields," J. Biomed. Opt. **15**(2), 021314 (2010).
38. S. Okawa, T. Hirasawa, T. Kushibiki, and M. Ishihara, "Numerical evaluation of linearized image reconstruction based on finite element method for biomedical photoacoustic imaging," Opt. Rev. **20**(5), 442–451 (2013).
39. Y.-L. Sheu and P.-C. Li, "Simulations of photoacoustic wave propagation using a finite-difference time-domain method with berenger's perfectly matched layers," J. Acoust. Soc. Am. **124**(6), 3471–3480 (2008).
40. M. Xu and L. V. Wang, "Analytic explanation of spatial resolution related to bandwidth and detector aperture size in thermoacoustic or photoacoustic reconstruction," Phys. Rev. E **67**(5), 056605 (2003).
41. O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, (Springer, 2015), pp. 234–241.
42. M. Drozdzal, E. Vorontsov, G. Chartrand, S. Kadoury, and C. Pal, "The importance of skip connections in biomedical image segmentation," in *International Workshop on Deep Learning in Medical Image Analysis, International Workshop on Large-Scale Annotation of Biomedical Data and Expert Label Synthesis*, (Springer, 2016), pp. 179–187.
43. A. N. S. Institute and L. I. O. America, "American national standard for safe use of lasers," Laser Inst of America (2000).
44. J. Kim, J. K. Lee, and K. M. Lee, "Accurate image super-resolution using very deep convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, (2016), pp. 1646–1654.
45. M. Abadi, A. Agarwal, and P. Barham, "Tensorflow: Large-scale machine learning on heterogeneous distributed systems," (2015).
46. D. R. I. M. Setiadi, "Psnr vs ssim: imperceptibility quality assessment for image steganography," Multimed. Tools Appl. **80**(6), 8423–8444 (2021).
47. S. Van der Walt, J. L. Schönberger, J. Nunez-Iglesias, F. Boulogne, J. D. Warner, N. Yager, E. Gouillart, and T. Yu, "scikit-image: image processing in python," PeerJ **2**, e453 (2014).
48. C. Yang, H. Lan, F. Gao, and F. Gao, "Deep learning for photoacoustic imaging: a survey," arXiv arXiv:2008.04221 (2020).
49. J. Feng, J. Deng, Z. Li, Z. Sun, H. Dou, and K. Jia, "End-to-end Res-Unet based reconstruction algorithm for photoacoustic imaging," Biomed. Opt. Express **11**(9), 5321 (2020).

Supplemental Document

Biomedical Optics EXPRESS

Deep learning on photoacoustic tomography to remove image distortion due to inaccurate measurement of the scanning radius: supplement

SUDEEP MONDAL,[1] SUBHADIP PAUL,[1] NAVJOT SINGH,[2] AND RATAN K SAHA[1,*]

[1]*Department of Applied Sciences, Indian Institute of Information Technology Allahabad, Prayagraj, 211015, India*
[2]*Department of Information Technology, Indian Institute of Information Technology Allahabad, Prayagraj, 211015, India*
[*]*ratank.saha@iiita.ac.in*

Supplement DOI: https://doi.org/10.6084/m9.figshare.24270952

Parent Article DOI: https://doi.org/10.1364/BOE.501277

Deep learning on photoacoustic tomography to remove image distortion due to inaccurate measurement of the scanning radius

SUDEEP MONDAL,[1] SUBHADIP PAUL,[1] NAVJOT SINGH,[2] AND RATAN K SAHA[1,*]

[1]*Department of Applied Sciences, Indian Institute of Information Technology Allahabad, Prayagraj, 211015, India*

[2]*Department of Information Technology, Indian Institute of Information Technology Allahabad, Prayagraj, 211015, India*

[*]*ratank.saha@iiita.ac.in*

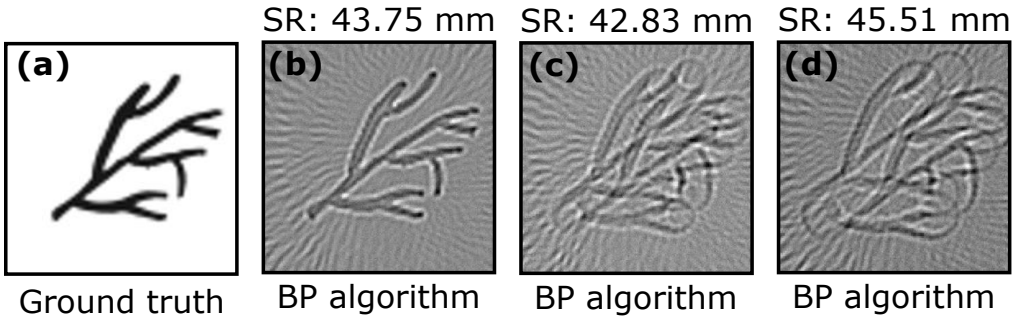

Fig. S1. (a) Ground truth, (b) image reconstruction using the BP algorithm at the original SR. (c) and (d) BP images with inaccurate SRs.

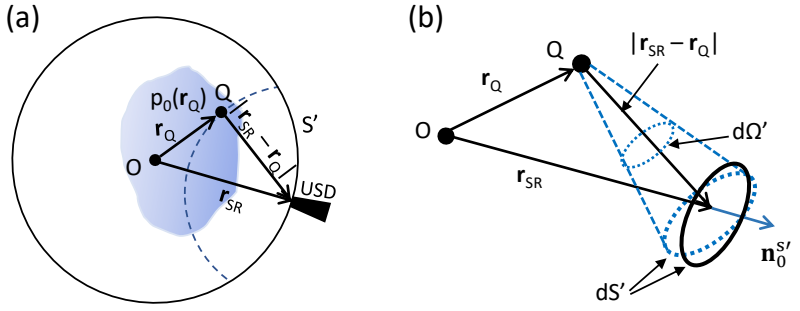

Fig. S2. (a) Measurement of the acoustic signal, coming from a source at Q, by a ultrasound detector (USD) placed on a surface S' at $\mathbf{r}_{SR}$. (b) A diagram showing formation of the solid angle $d\Omega'$ by the detection element dS' at a point Q. Similar figures are available in elsewhere.

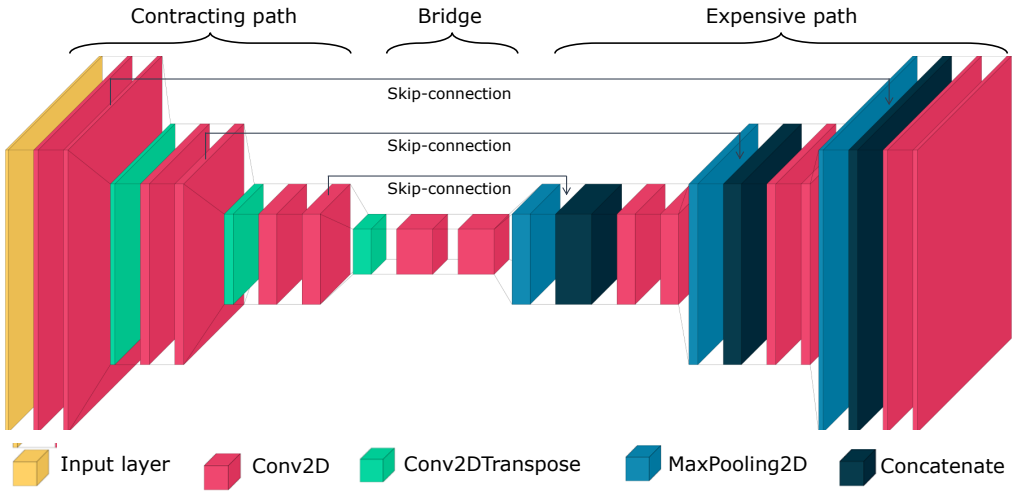

Fig. S3. Block diagram of the U-Net architecture.

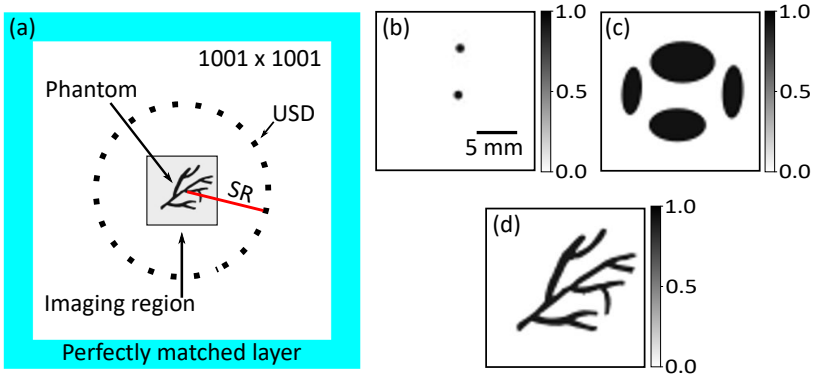

Fig. S4. (a) PAT setup for simulation. Demonstration of the two-point phantom, multi-ellipse phantom and vasculature phantom, respectively in (b), (c) and (d) used in the numerical study. Colorbar represents strength of initial pressure rise.