Deep neural network-based sinogram super-resolution and bandwidth enhancement for limited-data photoacoustic tomography

Awasthi, Navchetan; Jain, Gaurav; Kalva, Sandeep Kumar; Pramanik, Manojit; Yalavarthy, Phaneendra K.

2020

Awasthi, N., Jain, G., Kalva, S. K., Pramanik, M., & Yalavarthy, P. K. (2020). Deep neural network-based sinogram super-resolution and bandwidth enhancement for limited-data photoacoustic tomography. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 67(12), 2660-2673. doi:10.1109/TUFFC.2020.2977210

<https://hdl.handle.net/10356/146551>

<https://doi.org/10.1109/TUFFC.2020.2977210>

© 2020 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works. The published version is available at:
<https://doi.org/10.1109/TUFFC.2020.2977210>

Downloaded on 26 Apr 2023 13:23:36 SGT

Deep Neural Network Based Sinogram Super-resolution and Bandwidth Enhancement for Limited-data Photoacoustic Tomography

Navchetan Awasthi*, Gaurav Jain*, *Student Member, IEEE*, Sandeep Kumar Kalva, Manojit Pramanik, and Phaneendra K. Yalavarthy, *Senior Member, IEEE*

Abstract—Photoacoustic tomography (PAT) is a non-invasive imaging modality combining the benefits of optical contrast at ultrasonic resolution. Analytical reconstruction algorithms for photoacoustic signals require large number of data points for accurate image reconstruction. However, in practical scenarios, data is collected using limited number of transducers along with data being often corrupted with noise resulting in only qualitative images. Further, the collected boundary data is band-limited due to limited bandwidth of the transducer making the photoacoustic imaging with limited data being qualitative. In this work, a deep neural network based model with loss function being scaled root-mean-squared-error was proposed for super-resolution, denoising as well as bandwidth enhancement of the photoacoustic signals collected at the boundary of the domain. The proposed network has been compared with traditional as well as other popular deep learning methods in numerical as well as experimental cases and is shown to improve the collected boundary data in turn providing superior quality reconstructed photoacoustic image. The improvement obtained in the Pearson Correlation, Structural Similarity Index Metric and Root Mean Square Error was as high as 35.62%, 33.81% and 41.07% respectively for phantom cases and Signal to Noise Ratio improvement in the reconstructed photoacoustic images was as high as 11.65 dB for *in-vivo* cases as compared to reconstructed image obtained using original limited bandwidth data. Code is available at <https://sites.google.com/site/sercmig/home/dnnpat>.

Index Terms—Super-resolution, Image reconstruction, Photoacoustic imaging, Bandwidth Enhancement, Convolutional Neural Networks.

I. INTRODUCTION

PHOTOACOUSTIC tomography (PAT), also referred to as photoacoustic tomography is an “*imaging from coupled physics*”, has revolutionized biomedical imaging by facilitating high optical contrast along with high spatial resolution [1], [2]. Over the years, PAT has emerged as a non-invasive hybrid imaging modality, effectively unifying the merits of optical excitation and ultrasonic detection facilitating higher imaging depths [3]–[8]. In PAT, the tissue under investigation is irradiated with a nanosecond short laser pulse in near-infrared

(NIR) region. The NIR characteristics like low absorption and scattering coefficient in biological tissues results in deep tissue penetration achieving increased signal to noise ratio (SNR) without damaging the tissue. Light energy incident on the tissue is partially absorbed by it, consequently increasing temperature due to thermo-elastic expansion resulting in emission of acoustic waves. Subsequently, the generated waves propagate across the biological tissue and are detected by wideband ultrasound transducers located at the tissue boundary. Using reconstruction algorithms, information acquired by these transducers gets utilized to compute the initial pressure rise, which is proportional to the absorbed optical (or ultrasonic) energy. Additionally, optical absorption coefficient is highly susceptible to the patho-physiology of the tissue, enabling PAT to distinguish patho-physiological nature of the tissue in high contrast with ease (assuming homogeneous fluence). Hence, photoacoustic (PA) imaging unravels functional and structural information in a scalable manner, which further amplifies its utility for both clinical and pre-clinical applications [6], [9].

The most important part of photoacoustic imaging is to reconstruct accurate photoacoustic (PA) images using the acquired photoacoustic signals. For this, the acoustic inverse problem needs to be solved to estimate the initial pressure at time $t = 0$ using the acoustic measurements acquired through the transducers at time t . Several reconstruction algorithms exist in the literature, these can be broadly classified into (1) *analytical methods*, and (2) *model-based iterative methods* [10]–[13]. Based on spherical Radon transform model, analytical methods like filtered back-projection (BP), delay & sum, Fourier transform based methods as well as model-based time reversal methods although computationally frugal, require massive amounts of data for accurate image reconstruction. Large data requirements commensurate with increased scan time and excessive experimental setup cost, rendering it practically infeasible to meet such huge data requirement. Also, limited aperture of transducers further prohibits the data collection. Hence, analytical methods in these limited-data scenarios result in qualitative imaging [14]–[16]. On the other hand, model-based iterative models perform considerably better than their analytical counterparts in context to limited-data and display robustness to noise. However, these methods are computationally expensive and have high memory constraints. Hence, owing to the setup cost and practical considerations, reconstruction of photoacoustic images is a challenging task given the limited data and memory restrictions for real time

*N. Awasthi and G. Jain contributed equally in this work.

N. Awasthi G. Jain, and P. K. Yalavarthy are with the Department of Computational and Data Sciences, Indian Institute of Science, Bangalore 560012, India (e-mail: navchetana@iisc.ac.in; gauravjain13298@gmail.com; yalavarthy@iisc.ac.in).

G. Jain current affiliation is Department of Computer Science and Engineering, Delhi Technological University, Delhi, India.

S. K. Kalva and M. Pramanik are with the School of Chemical and Biomedical Engineering, Nanyang Technological University, 637459, Singapore (e-mail: sandeepk002@e.ntu.edu.sg; manojit@ntu.edu.sg).

applications. Irrespective of the reconstruction algorithm employed, factors like Signal-to-Noise Ratio (SNR), quality of photoacoustic signals, and limited bandwidth (BW) of the ultrasound transducers affect the performance of these techniques. For any pragmatic experimental setting, transducers are always band-limited and introduce a blurring effect on the reconstructed images.

In order to overcome the inevitable limitation arising due to band-limited ultrasound transducers, an additional step is required to reinstate the quality of reconstructed photoacoustic images. Earlier, a deconvolution-based deblurring algorithm was proposed by Rejesh *et al.* [17]. In this approach, the detected photoacoustic signal is convolved with transducer impulse response to restore the original photoacoustic signal. However, deconvolution-based approaches have been classified as “ill-posed” [18], due to their failure to achieving unique solutions when the data is an approximate and thereby entail regularization. In [19], Cao *et al.* recently proposed an inverse filtering method coupled with multi-sampled signal de-noising technique to recover out-of-band frequency components. This signal recovery not only resulted in increased SNR, but also managed to reveal deeper structural details in reconstructed photoacoustic images.

Over the past few years, deep learning has been extensively employed in the biomedical imaging domain outperforming most of the existing classical approaches. In particular, convolutional neural networks have witnessed proliferation in application to computer-aided detection [20]–[23], segmentation [24]–[27], and image reconstruction [28]–[31]. Antholzer *et al.* [32] proposed a method for improving the quality of the reconstructed photoacoustic images obtained using the back-projection technique via a deep learning based architecture. In [33], [34], the authors introduced a convolutional neural network that was trained to identify the location of the individual point targets for photoacoustic imaging, without requiring traditional geometry based beamforming from the beamformed data for various sound speeds, target locations and absorber sizes. In [35], a deep neural network was proposed for providing high resolution three dimensional photoacoustic images from the limited view data. The model represented an iterative scheme and incorporates gradient information of the data for compensating the limited view artefacts. Davoudi *et al.* [36] proposed a U-Net based deep neural network for getting superior quality photoacoustic images from the substantially under-sampled data or limited view scans.

In this work, a hybrid end-to-end trainable deep neural network model was proposed for super-resolution, denoising as well as bandwidth enhancement of photoacoustic signals for effective reconstruction of quantitatively accurate photoacoustic images. Further, the proposed approach may be applied to ultrasound tomography, tweaking the parameters and fine tuning the model. The super-resolution based techniques using simple convolutional neural networks and U-Net based architectures have been used in Computed Tomography [37], [38] and currently has not been used in photoacoustic tomography. There have been different architectures used for super-resolution such as VDSR [39] (Accurate Image Super-Resolution Using Very Deep Convolutional Networks) which

uses high learning rate and residual learning, RED30 [40] (Image Restoration Using Very Deep Convolutional Encoder-Decoder Networks with Symmetric Skip Connections) which symmetrically link convolutional and de-convolutional layers with skip-layer connections and training loss converges much faster and attains a higher-quality local optimum, and DRCN [41] (Deeply-Recursive Convolutional Network for Image Super-Resolution) which uses recursive-supervision and skip-connection, SR-GAN [42] (Super-resolution Generative Adversarial Networks) which uses a perceptual loss function consisting of an adversarial loss and a content loss to recover photo-realistic textures from heavily down-sampled images on public benchmarks. These architectures have been used for super-resolution of natural images, where as the U-net that was utilized in this work was shown to perform well on medical images [38], and hence these architectures were not utilized. In photoacoustic tomography there have been different deep learning architectures which were used for improving back projection based reconstruction, fusion of different reconstruction techniques [43] but there has not been any work in the sinogram (data) domain to perform the super resolution and bandwidth enhancement using a simple network. The band-width enhancement using neural network was attempted earlier, it has shown limited improvement [44]. This is the first network which proposes super-resolution (increasing the available data), denoising as well as bandwidth enhancement of collected photoacoustic data for improving photoacoustic tomography. The main contributions of the work can be summarized as:

- It is the first kind of single network that performs super-resolution, denoising as well as the bandwidth enhancement of photoacoustic data in the sinogram domain (majority of deep learning networks have been proposed in the image space, i.e. improve the reconstructed image).
- The proposed neural network works on the full sinogram, rather than at pixel-level, thus inherently providing robustness and generalization. The robustness was also shown when the network was trained for numerical phantoms and was shown to perform well even for experimental in-vivo data.
- The architecture proposed with modifications can be used to enhance the experimentally collected raw data (sinogram), which can in turn improve the outcome of an inverse problem, thus inherently removing the bias of the image reconstruction method.
- Introduction of the scaled root mean square loss function for training the network for sinogram data, which contains values with extremely low magnitude, can be extended to other applications that share properties of photoacoustic data. The scaling also helps in preserving the gradient values with small magnitudes and thus the vanishing gradient problem can be solved.
- The method is shown to perform favourably as compared to the other state of the art conventional and deep learning based techniques such as Automated Wavelet Denoising, Super Resolution Convolutional Neural Network, and U-net architectures.

- The photoacoustic data generated through k-WAVE MATLAB Toolbox for training the model was made publicly available to augment research in photoacoustic imaging, encouraging the community to develop methods that can improve the sinogram (rather than doing it indirectly by improving the reconstructed image).

The rest of the paper is organised as follows: In section II, we briefly introduce existing methods and elucidate the proposed method. In section III, evaluation metrics have been discussed, whereas in section IV, details about the experimental setup has been canvassed. Section V validates the performance of the proposed approach.

II. PHOTOACOUSTIC IMAGE RECONSTRUCTION

In this section the photoacoustic image reconstruction along with few approaches to overcome its limitations is being discussed. Detailed explanation of the proposed deep learning based approach follows this.

A. Forward Problem

In photoacoustic tomography, the forward model provides the computation of data collected by acoustic transducers located at the tissue boundary using Eq. 1, which determines propagation and generation of the acoustic wave inside the tissue [8],

$$\left(\Delta^2 - \frac{1}{c^2} \frac{\partial^2}{\partial t^2}\right) \mathcal{P}(\vec{x}, t) = \frac{-\alpha}{C_p} \frac{\partial \mathcal{H}(\vec{x}, t)}{\partial t} \quad (1)$$

where $\mathcal{P}(\vec{x}, t)$ represents the photoacoustic wave pressure at location $\vec{x}$ at time t , c is the speed of light, C_p denotes the specific heat capacity, α represents the coefficient of thermal expansion and $\mathcal{H}(\vec{x}, t)$ represents the energy deposited per unit volume per unit time. Further, the reconstruction/inverse problem incorporates estimating the initial pressure $\mathcal{P}(\vec{x}, t = 0)$ given the acquired boundary measurements at time t , making it an initial value problem. The following subsection describes the system of equations governing the numerical experimental setup.

B. System Matrix Building

In order to simplify the photoacoustic wave propagation model, it can be represented through a linear system of equations as:

$$\mathbf{A}x = b \quad (2)$$

where $\mathbf{A} \in \mathbb{R}^{m \times n^2}$ is the *system matrix* containing impulse responses, such that each column represents the response of a pixel in the imaging domain for varying time domains. Here m is product of the number of transducers and the number of time samples that each of these transducers acquire and x denotes an image of size $n \times n$ pixels, which is converted to a vector such that $x \in \mathbb{R}^{n^2 \times 1}$. The data collected by each transducer i.e. $b \in \mathbb{R}^{m \times 1}$ is computed using Eq. 2 in the simplified forward model [45], [46]. To obtain the image, Linear Back-projection (LBP) method can be used as: [47]

$$x_{\text{LBP}} = \mathbf{A}^T b \quad (3)$$

where $x_{\text{LBP}} \in \mathbb{R}^{n^2 \times 1}$ is the image vector and $\mathbf{A}^T$ is the transpose of the system matrix. The LBP method is one of the most basic methods for image reconstruction. In this work, the image reconstruction was performed using LBP method due to its low computation time and being analytical in nature. Since the sinogram has been improved using the proposed deep neural network, the LBP will be effective in comparison of reconstructed results as it does not have any parameters (like regularization) that effect the image quality. In the experiments explained in this work, the same setup as comprehensively explained in [45], [46] was employed. The following subsection describes an automated wavelet-based denoising method to improve boundary data in turn improving the reconstructed photoacoustic images.

C. Automated Wavelet Denoising (AWD) Method

The data obtained using the interpolation can be denoised using the maximum overlap DWT (MODWT) [48] method (non-orthogonal transform) using the wavelet denoising [49]. The advantages of MODWT are:

- The sample size is not restricted to powers of 2.
- Zero padding is not required.
- It forms a zero-phase filter and hence lines up with the features of the original signal in contrast to the DWT transform.

The complete discussion can be found in Ref. [48] for interested readers. The application to photoacoustic images has been discussed in Ref. [49]. In this work MODWT has been utilized for denoising the photoacoustic signals obtained after interpolation of the data obtained using 100 detectors. The threshold for denoising is chosen by the universal threshold criteria [50].

D. Super Resolution Convolutional Neural Network (SRCNN)

Image super-resolution [51] has a wide application to translate an image from low resolution to a high resolution image. Previous techniques exploit internal similarities of same image [52]–[54] or learn mapping functions [55]–[57] for translation of low resolution to high resolution image. These techniques are challenged by difficulties in effectively and compactly modeling the data. Sparse coding techniques [58], [59] were used in the past for image based super-resolution. It involves extraction of overlapping patches, encoded by a low resolution dictionary. These coefficients are then used for constructing the dictionary, which helps to build corresponding high resolution image. These techniques either learn and optimize dictionaries [58], [59] or try to model them in different ways [55], [56]. Super Resolution Convolutional Neural Network (SRCNN) [60] was utilized to model one such dictionary, that learns the mapping from low resolution to high resolution image. This architecture was used to first map the noisy, bandlimited patches using 100 detectors to noiseless, full bandwidth patches using 200 detectors. For comparing the performance of SRCNN with the proposed network, the implementation configurations mentioned in the original work [60] were deployed.

Fig. 1: Training configuration for the proposed U-Net based architecture. Data collected through 100 transducers with 70% bandwidth (BW) is added with variable noise resulting in SNR values of data from 10-70 dB is shown in (a). Patches of size 64×64 are cropped from (a) and input into the U-Net based model. Output patch of size 64×64 is then used to compute the scaled-RMSE loss with the corresponding patch from the data collected through 200 transducers and 100% bandwidth (as shown in (b)). This loss is propagated backwards to train the proposed model for bandwidth enhancement, super-resolution, and denoising.

E. Proposed Method

Deep learning has found numerous applications in image processing tasks such as image enhancement, edge detection, and image super-resolution. These have been recently extended to applications in CT imaging for super-resolution of the sinogram data using simple convolutional layers as well as a U-Net based architectures [38]. In this work, a hybrid end-to-end trainable U-Net [61] based model is proposed for super-resolution, denoising as well as bandwidth enhancement of photoacoustic signals. The following sub-subsections discuss the (a) *Architecture*, (b) *Training details*, and (c) *Testing methodology* for the proposed.

a) Architecture: The proposed method trains a U-Net based architecture shown in Fig. 1. The first layer contains 32 convolutional filters of size 3×3 with a stride of 2×2 , which progressively increase in powers of 2 until the bottleneck layer with a maximum of 512 filters. Each layer with corresponding number of 3×3 filters apply two consecutive convolution operations with RELU (Rectified Linear Unit) activations followed by a Max-pooling layer with a window of size 2×2 .

In Ref. [38], pooling layers were utilized for super-resolving the sinogram data of computed tomography (CT) images. The pooling layer that was utilized in this work down-samples the input using max pooling (similar to Ref. [38]). It helps in making shift invariant results by taking only specific values for the input. It also reduces the spatial size of representation and reduce the number of parameters and computation in the network. The shift invariance plays an important role in image segmentation and image classification where the position of the target is arbitrary. After down-sampling the input to reach the bottleneck layer, symmetrically inverse operations were applied using the same filter size and stride to up-sample the bottleneck layer to obtain an output of the same size as the input. Number of filters decrease progressively over the layers in a symmetric fashion. At each layer, zero padding of size '1' was performed to maintain the size before and after convolution operations. Skip connections were added during up-sampling from the corresponding down-sampled layers in order to maintain the overall sinogram structure as shown in the figure above. Note that the original U-Net architecture layers employ softmax activation functions in the

final two layers for segmentation. However, we used ELU (Exponential Linear Unit) (Eq. 5) activation [62]. Although RELUs have been employed extensively for modeling images or improving the reconstruction quality through deep neural networks, it fails to effectively characterize photoacoustic data that contains positive as well as negative values. Hence, in this work ELU activations were employed, which have been shown to outperform RELU activations for specific tasks [62]. The motivation for this stems from the fact that sinogram data may contain negative values, which the RELU activations fail to achieve as evident from Eq. 4.

$$\text{relu}(x) = \begin{cases} x & \text{if } x > 0 \\ 0 & \text{if } x \leq 0 \end{cases} \quad (4)$$

$$\text{elu}(x) = \begin{cases} x & \text{if } x > 0 \\ \alpha(e^x - 1) & \text{if } x \leq 0 \end{cases} \quad (5)$$

where x is the input and α is the regularization parameter for elu . In addition to this, ELUs push the mean unit activations to zero, thereby mimicking batch normalization with an advantage of lower computational complexity. To further evaluate this, we compare three variants of the U-Net architecture, (1) U-Net (Relu): where all activations are ReLUs, (2) U-Net (Elu): where all activations are ELUs, and (3) U-Net (Hybrid): where ELUs have only been used in the final layers. The proposed U-Net (Hybrid) variant was able to provide the generalization by effectively modeling the negative values in sinogram through the final layers, while maintaining the advantages of ReLUs in U-Net architecture.

b) Training: For training the proposed model for super-resolution and bandwidth enhancement of the sinogram domain data, training data was generated using k-WAVE MATLAB toolbox as mentioned in section IV. Sinogram data obtained corresponding 1,000 Breast images [63] were used to train the model. Sinogram was generated for all these images for limited bandwidth for 100 detectors and full bandwidth for 200 detectors for training the network. The sinogram data was obtained using the limited bandwidth detectors placed at the boundary of the imaging domain. These sinograms obtained were then added with Gaussian noise and sampled noisy sinograms were used as the input for the network. The output consists of fully sampled, full bandwidth and denoised sinograms. Since these output sinograms correspond to at least 200 detectors data and does not contain noise, the reconstruction of photoacoustic image was performed via linear back-projection [47] technique. Note that the aim of this work is to show that the proposed network is capable of enhancing the photoacoustic data (sinogram), so utilization of simple back-projection as the reconstruction method does provide unbiased way of quantifying the improvement of image quality. Reconstructed image quality using filtered back projection or other variants [64] depends on the choice of appropriate filter or parameters in the filter, which may not be straight forward to assess the improvement in the photoacoustic data obtained via the methods discussed here. From each image, 105 patches of size 64×64 were generated by sliding over the 200×512 sinogram data with a stride

of 16×16 . A total of $105 \times 1,000 = 1,05,000$ patches were used for training, where a 20% split for training and validation was done. The training dataset has 84,000 patches while the validation dataset has 21,000 patches. Further, Gaussian noise in the range of 10 dB-70 dB was added in the data while training and the ground truth were the patches generated using the 200 detectors data. Total number of trainable parameters used for training the network were 7,759,787. The network was trained using the proposed scaled root mean square error (scaled-RMSE) loss function, defined as:

$$\mathcal{L} = \frac{1}{N} \sum_{i=1}^N \|\hat{x}_i - \phi(x_i)\|^2 \times \tau \quad (6)$$

where, $\hat{x}_i$ are the predictions of the model and $\phi(x_i)$ are the true signals. Since sinogram data contains values of the order of 0.0001, a loss multiplication factor, $\tau = 10000$ was used to ensure effective back-propagation of the loss to the initial layers. It enables to train a deeper network without worrying about the dying out of loss gradients and hence preventing the problem of vanishing gradients. In Ref. [65] loss scaling was proposed to preserve gradient values with small magnitudes. It was shown that scaling up the gradients will shift them to occupy more of the representable range and hence will result in preserving values that are otherwise close to zero. One of the approaches of scaling is to scale the loss obtained in the forward pass before starting the computation of back-propagation. Once scaling is performed, chain rule affects all the computed values by the same amount and thus no extra operations are required and the vanishing gradient problem can be easily avoided. The scaling factor was determined after extensive experimentation and the optimal value of 10000 was obtained for the proposed method utilizing to the approach mentioned in Ref. [65]. The scaling of the input by the factor has not been tried in the proposed method as the input data needs to be multiplied by the same factor while in case of loss scaling it will be only one multiplication. Further, Adam optimizer was used for training the network with $\gamma = 0.98$ with a step size of 2. The learning rate, number of epochs and the batch size were set to be $1e-3$, 250 and 128 respectively. The model was fine tuned to obtain the best set of hyper-parameters like starting number of filters (32), maximum number of filters (512), batch size, learning rate, γ , step size, epochs, etc. All computations were carried out on a Linux workstation with Intel Xeon Silver 4110 CPU with 2.10 GHz clock speed, having 128 GB RAM and a TITAN RTX GPU with 24 GB memory. It took 65s (approx.) for one epoch, and a total of 4.5 hrs (approx.) to train the model using the above mentioned workstation. PyTorch implementation of the proposed method shall be made available online at <https://sites.google.com/site/sercmig/home/dnnpat>.

c) Testing: For testing, input data obtained using limited bandwidth (70% for numerical data) and low resolution i.e. 100 detector data was first interpolated using nearest neighbour to obtain a sinogram of size 200×512 . This sinogram data was padded using the 'reflect' padding, to obtain a sinogram data of size 512×512 . This was performed to ensure that there is no mismatch between the tensor sizes while concatenating

the skip connections. The padded sinogram data is input into the model to obtain super-resolved sinogram data with full bandwidth. This enhanced data was then reconstructed using back-projection to obtain reconstructed photoacoustic images. These images are then evaluated using figure of merits to analyse the improvement observed with the proposed method. The following section discusses the figures of merit used for evaluation.

III. FIGURES OF MERIT

Quantitative performance of the proposed method has been compared with existing state of the art methods over the reconstructed photoacoustic (PA) images using evaluation metrics namely, Root Mean Square Error (RMSE), Structural Similarity (SSIM) Index, Pearson Correlation (PC) and Signal to Noise Ratio (SNR). For numerical phantoms, RMSE, PC and SSIM were compared, whereas for experimental *in-vivo* phantoms, SNR was employed as the sole criteria of evaluation due to the absence of target initial pressure. The following subsections describe the metrics in detail.

A. Root Mean Square Error (RMSE)

The RMSE evaluates the standard deviation of the residuals as an absolute measure to evaluate the quality of the reconstructed image [66], [67]. Defined as:

$$\text{RMSE}(\mathcal{I}^{\text{target}}, \mathcal{I}^{\text{rec}^n}) = \sqrt{\frac{\sum_{i,j}^{\mathcal{N}} (\mathcal{I}_{i,j}^{\text{target}} - \mathcal{I}_{i,j}^{\text{rec}^n})^2}{\mathcal{N}^2}} \quad (7)$$

where $\mathcal{I}^{\text{target}}$ and $\mathcal{I}^{\text{rec}^n}$ denote the target and reconstructed images, each of pixel size $\mathcal{N} \times \mathcal{N}$. Lower residuals are ideal for a good quality reconstructed images and hence low RMSE values are desirable.

Fig. 2: Illustrative representation of the data collection setup. Computational grid of 501×501 with an array of 200 ultrasound transducers around the data generation grid of 401×401 and a 201×201 pixels reconstruction grid is shown. Transducers are placed at a radial distance of 22.5 mm from the centre of the grid setup.

B. Structural Similarity Index (SSIM)

SSIM amalgamates three comparative measurements namely luminance (l), contrast (c) and structure/ correlation(s) to evaluate the quality of an image with respect to a reference image. In our case, similarity of the reconstructed image (rec^n) is computed with respect to the target image [68], [69]. The three parameters are defined as:

$$l(\text{target}, \text{rec}^n) = \frac{2\mu_{\text{target}}\mu_{\text{rec}^n} + c_1}{\mu_{\text{target}}^2 + \mu_{\text{rec}^n}^2 + c_1} \quad (8)$$

$$c(\text{target}, \text{rec}^n) = \frac{2\sigma_{\text{target}}\sigma_{\text{rec}^n} + c_2}{\sigma_{\text{target}}^2 + \sigma_{\text{rec}^n}^2 + c_2} \quad (9)$$

$$s(\text{target}, \text{rec}^n) = \frac{\sigma_{\text{target}, \text{rec}^n} + c_3}{\sigma_{\text{target}}\sigma_{\text{rec}^n} + c_3} \quad (10)$$

where μ_{target} and μ_{rec^n} represent the mean, σ_{target} and σ_{rec^n} represent the variance of the target and reconstructed images respectively, $\sigma_{\text{target}, \text{rec}^n}$ represents the covariance between the two images, and c_1, c_2 and c_3 are constants such that $c_3 = c_2/2$. SSIM is defined as the weighted combination of the three components, $\text{SSIM} = [l^\alpha \cdot c^\beta \cdot s^\gamma]$. Higher the SSIM, higher the similarity between the reconstructed and target image.

C. Pearson Correlation (PC)

PC is a measure of linear correlation between two entities [70]. Here, correlation between the target image $\mathcal{I}^{\text{target}}$ and reconstructed image $\mathcal{I}^{\text{rec}^n}$ is computed as:

$$\text{PC}(\mathcal{I}^{\text{target}}, \mathcal{I}^{\text{rec}^n}) = \frac{\text{cov}(\mathcal{I}^{\text{target}}, \mathcal{I}^{\text{rec}^n})}{\sigma(\mathcal{I}^{\text{target}}) \cdot \sigma(\mathcal{I}^{\text{rec}^n})} \quad (11)$$

where cov represents the covariance and σ represents the variance. $\text{PC} \in [-1, 1]$, where 1 denotes complete linear correlation, 0 represents no linear correlation and -1 indicates a perfectly complementary relationship.

D. Signal to Noise Ratio (SNR)

SNR is a quantitative measure that assesses the desired signal strength in comparison with the background noise [71]. As mentioned above, the target initial pressure distribution is unavailable for experimental *in-vivo* phantoms, hence SNR is employed to analyze the quantitative improvement. It is defined as:

$$\text{SNR}_r \text{ (dB)} = 20 \times \log_{10} \left(\frac{\mathcal{S}}{\sigma_{\text{noise}}} \right) \quad (12)$$

where $\mathcal{S}$ represents the signal power and σ_{noise} represents the standard deviation in the noise. A higher SNR value (in dB) is desired in the reconstructed photoacoustic image.

IV. NUMERICAL AND EXPERIMENTAL STUDIES

This section describes the numerical and experimental setup for generating data similar to as done in References [43], [46], [72]. A computational grid of 501×501 pixels was employed with 0.1 mm/pixel. For data generation, a high dimensional 401×401 pixels grid was used, while reconstruction was performed on a low dimensional grid of 201×201 pixels to

Fig. 3: Target numerical phantoms used for quantitative evaluation of the proposed method. (a), (b), (c), and (d) represent ground truth images for PAT, modified Derenzo, Blood vessel, and breast phantoms. Corresponding reconstructed images using Linear Backprojection (LBP) with full bandwidth and noise less data are shown in (e), (f), (g), and (h) to serve as reference for best case.

TABLE I: Comparative results obtained for reconstructed PAT phantom images shown in Fig. 4 in terms of figure of merits: Pearson Correlation Coefficient (PC), Structural Similarity (SSIM) Index and Root Mean Square Error (RMSE) for SNR of data being 20, 40 and 60 dB. Bold text indicates the best possible value for the metric used.

SNR of data	20 dB			40 dB			60 dB		
Method/ Metric	PC	SSIM	RMSE	PC	SSIM	RMSE	PC	SSIM	RMSE
Limited BW	0.3077	0.2728	0.1078	0.3293	0.3262	0.1012	0.3293	0.3262	0.1012
Interpolated	0.5317	0.4386	0.0871	0.5738	0.5417	0.0822	0.5746	0.5429	0.0805
Automated Wavelet Denoising	0.5387	0.5376	0.0782	0.6367	0.5761	0.0765	0.6367	0.5761	0.0765
SRCNN	0.6712	0.6474	0.0806	0.7375	0.7096	0.0711	0.7386	0.7100	0.0685
U-Net (Relu)	0.6668	0.6416	0.0756	0.7396	0.7289	0.0637	0.7421	0.7291	0.0642
U-Net (Elu)	0.6676	0.6501	0.1083	0.7388	0.7201	0.0754	0.7419	0.7228	0.0721
U-Net (Hybrid)	0.7306	0.7034	0.0617	0.7721	0.7709	0.0539	0.7727	0.7704	0.0539

avoid the inverse problem. White Gaussian noise was added to the data generated, in order to achieve variable noise levels having SNR values between 20 dB and 60 dB. As depicted in Fig. 2, hundred point-transducers with a BW of 70% having central frequency of 2.25 MHz were placed at the tissue boundary in a circularly equidistant manner at a radius of 22.5 mm. Each transducer sampled 512 observations after every 500 ms i.e. a sampling frequency of 20 MHz. Speed of sound in the medium (tissue under investigation) was assumed to be a uniform with no absorption or dispersion at 1500 m/s. This configuration results in a system matrix $\mathbf{A} \in \mathbb{R}^{51200 \times 40401}$.

For a comprehensive comparison of the proposed method with existing approaches, four distinct numerical phantoms namely (1) PAT, (2) modified Derenzo, (3) Blood Vessel (BV), and (4) Breast were considered. The numerical experimental data was generated using the open source k-WAVE MATLAB toolbox [73]. The PAT phantom constitute the alphabets “PAT” (Fig. 3(a)) and aid in determining the efficacy of recovering sharp objects. The modified Derenzo phantom (Fig. 3(b)) contain circular object groups with differing radii that assist in evaluating the recovery strength of small and large objects. The blood vessel (BV) phantom (Fig. 3(c)), as the name suggests,

resemble the thick and thin blood vessel like structures and help analyse the success in recovery of amorphous complicated structures. Note that a maximum initial pressure distribution of 1 kPa was asserted for the first three phantoms. Finally, the Breast phantom was created using contrast-enhanced imaging data [63], [74]. In this phantom, the initial pressure was varied from 0 to 5 kPa and the expected initial pressure distribution was given in Fig. 3(d). Breast phantoms help in testing the robustness of the proposed approach in cases with complex photoacoustic images bearing variable contrast and noise. The reconstructed images using noiseless data with 200 detectors using these four phantoms was presented in second row of Fig. 3 to serve as reference (to know what is truly achievable).

In addition, performance of the proposed approach is evaluated using two experimental phantoms namely, (1) horse-hair phantom (Fig. 8), and (2) *in-vivo* rat brain phantom (Fig. 9). The trilateral horse-hair phantom data was acquired using Nd:YAG laser while the *in-vivo* rat brain phantom data was acquired using a pulsed laser diode (PLD). Further, we restrict ourselves with any detailed description of the experimental setup and direct the readers to refer to our previous works [75], [76] for elaborate discussions. Note that all experiments conducted on animals were meticulously carried out, adhering

Fig. 4: Reconstructed PAT phantom images with each row representing the SNR of data being 20, 40, and 60 dB and correspondingly given in (a)-(g), (h)-(n) and (o)-(u) respectively. Reconstructed images using data with limited BW (70%) and 100 transducers are shown in (a), (h), and (o). While (b), (i), and (p) show reconstruction results when the data with 100 transducers that are interpolated using nearest-neighbour to obtain a 200 transducer data. (c), (j) and (q) show results obtained using the automated wavelet de-noising approach. The results obtained using the SRCNN based method are shown in (d), (k), and (r) and the results obtained using the U-Net (Relu) are shown in (e), (l), and (s). Finally, (f), (m), and (t) show results obtained using the proposed U-Net (Elu) approach and (g), (n), and (u) show the results obtained using the proposed U-Net (Hybrid) based approach. Table-I, reports the figure of merits of these results in a comparative manner.

TABLE II: Comparative results obtained for reconstructed modified Derenzo phantom images shown in Fig. 5 in terms of figure of merits: Pearson Correlation Coefficient (PC), Structural Similarity (SSIM) Index and Root Mean Square Error (RMSE) for SNR of data being 20, 40 and 60 dB. Bold text indicates the best possible value for the metric used.

SNR	20 dB			40 dB			60 dB		
	PC	SSIM	RMSE	PC	SSIM	RMSE	PC	SSIM	RMSE
Limited BW	0.3525	0.3199	0.1953	0.3616	0.3263	0.1949	0.3616	0.3263	0.1949
Interpolated	0.6544	0.5425	0.1456	0.6661	0.5610	0.1498	0.6662	0.5612	0.1502
Automated Wavelet Denoising	0.6185	0.5110	0.1177	0.7038	0.5851	0.1426	0.7038	0.5851	0.1426
SRCNN	0.7578	0.6477	0.1103	0.7907	0.6690	0.0944	0.7912	0.6697	0.0940
U-Net (Relu)	0.7714	0.6707	0.1112	0.8015	0.6900	0.1063	0.8015	0.6905	0.1062
U-Net (Elu)	0.7645	0.6728	0.0914	0.7842	0.6833	0.0952	0.7714	0.6707	0.1112
U-Net (Hybrid)	0.8020	0.7015	0.1031	0.8305	0.7195	0.1240	0.8305	0.7199	0.1242

to the guidelines and regulations laid out by the Institutional Animal Care & Use Committee of Nanyang Technological University, Singapore (Animal Protocol Number ARF-SBS/NIE-A0263). The following section compares the figures of merits achieved through the proposed method and displays its superiority against state of the art methods along with a visual juxtaposition.

V. RESULTS AND DISCUSSIONS

In this section, detailed discussion about the results obtained in terms of figure of merits is covered. Performance of numerical phantoms has been compared in terms of PC, SSIM and RMSE, while experimental phantoms are compared based on SNR_r (dB) values. This is followed by qualitative analysis

of the images obtained in both the cases. Performance of the proposed approach is compared with both classical and deep learning based state of the art techniques and variants of the U-Net architecture. For a fair evaluation, all compared methods were trained using the same data. Note that the proposed scaled-RMSE loss proved crucial in convergence of all deep learning based approaches discussed in this work.

A. Numerical Phantoms

Target images for the PAT phantom (consisting of 'PAT' letters), Derenzo (consisting of small and large circles), blood vessel (consisting of thin and thick vessel structures), and Breast phantom (consisting of contrast variations) are shown in Fig. 3 (a), (b), (c), and (d) respectively. The image re-

Fig. 5: Reconstructed modified Derenzo phantom images with each row representing the SNR of data being 20, 40, and 60 dB and correspondingly given in (a)-(g), (h)-(n) and (o)-(u) respectively. Reconstructed images using data with limited BW (70%) and 100 transducers are shown in (a), (h), and (o). While (b), (i), and (p) show reconstruction results when the data with 100 transducers that are interpolated using nearest-neighbour to obtain a 200 transducer data. (c), (j) and (q) show results obtained using the automated wavelet de-noising approach. The results obtained using the SRCNN based method are shown in (d), (k), and (r) and the results obtained using the U-Net (Relu) are shown in (e), (l), and (s). Finally, (f), (m), and (t) show results obtained using the proposed U-Net (Elu) approach and (g), (n), and (u) show the results obtained using the proposed U-Net (Hybrid) based approach. Table-II, reports the figure of merits of these results in a comparative manner.

TABLE III: Comparative results obtained for reconstructed Blood vessel phantom images shown in Fig. 6 in terms of figure of merits: Pearson Correlation Coefficient (PC), Structural Similarity (SSIM) Index and Root Mean Square Error (RMSE) for SNR of data being 20, 40 and 60 dB. Bold text indicates the best possible value for the metric used.

SNR	20 dB			40 dB			60 dB		
	PC	SSIM	RMSE	PC	SSIM	RMSE	PC	SSIM	RMSE
Limited BW	0.3124	0.2930	0.1429	0.3326	0.3136	0.1375	0.3326	0.3136	0.1375
Interpolated	0.5879	0.4591	0.1155	0.6224	0.4943	0.1200	0.6233	0.4958	0.1161
Automated Wavelet Denoising	0.4617	0.3721	0.1264	0.6540	0.5208	0.1153	0.6540	0.5208	0.1153
SRCNN	0.7094	0.5769	0.1016	0.7870	0.6607	0.1022	0.7886	0.6624	0.1031
U-Net (Relu)	0.7043	0.5767	0.1016	0.7537	0.6191	0.1084	0.7543	0.6208	0.1072
U-Net (Elu)	0.7154	0.5756	0.0986	0.7621	0.6369	0.0943	0.7628	0.6374	0.0934
U-Net (Hybrid)	0.7328	0.5980	0.1195	0.7911	0.6568	0.1084	0.7922	0.6581	0.1056

construction obtained using 200 detectors and full bandwidth is shown in Fig. 3 (e), (f), (g), and (h) respectively. These images serve as the desired output for our simulations after the reconstruction is performed using the linear back-projection based method.

The reconstruction obtained using limited data for the ‘PAT’ phantom using 100 detectors (to serve as limited data case) is shown in Fig. 4 (a). The reconstruction obtained after interpolation and then denoising are shown in Fig. 4 (b) and Fig. 4 (c) respectively for the ‘PAT’ phantom with SNR of data being 20 dB. The reconstruction obtained for the ‘PAT’ phantom using 200 detectors data using SRCNN is shown in Fig. 4 (d), U-Net (Relu) is shown in Fig. 4 (e), U-Net (Elu) is shown in Fig. 4 (f), and the proposed model U-Net (Hybrid) is shown in

Fig. 4 (g). The improvement obtained in PC, SSIM and RMSE as compared to the interpolated data case are 35.62%, 30.84% and 21.09% respectively in this case of SNR of data being 20 dB. As seen from the Fig. 4 (g), the reconstructed image is relatively artifact free with lesser noise in the background. Similarly the reconstruction obtained for SNR of data being 40 dB for the limited data, interpolated, denoised, SRCNN, U-Net (Relu), U-Net (Elu), and the proposed technique are shown in Figs. 4 (h)-(n) respectively. The improvement obtained in the PC, SSIM and RMSE was 21.26%, 33.81% and 29.54% respectively. As observed in earlier case, the background noise was suppressed to provide improved reconstructed image. Similarly the reconstruction obtained for the case of SNR of data being 60 dB for the discussed methods are shown

Fig. 6: Reconstructed Blood vessel images with each row representing the SNR of data being 20, 40, and 60 dB and correspondingly given in (a)-(g), (h)-(n) and (o)-(u) respectively. Reconstructed images using data with limited BW (70%) and 100 transducers are shown in (a), (h), and (o). While (b), (i), and (p) show reconstruction results when the data with 100 transducers that are interpolated using nearest-neighbour to obtain a 200 transducer data. (c), (j) and (q) show results obtained using the automated wavelet de-noising approach. The results obtained using the SRCNN based method are shown in (d), (k), and (r) and the results obtained using the U-Net (Relu) are shown in (e), (l), and (s). Finally, (f), (m), and (t) show results obtained using the proposed U-Net (Elu) approach and (g), (n), and (u) show the results obtained using the proposed U-Net (Hybrid) based approach. Table-III, reports the figure of merits of these results in a comparative manner.

TABLE IV: Comparative results obtained for reconstructed Breast phantom images shown in Fig. 7 in terms of figure of merits: Pearson Correlation Coefficient (PC), Structural Similarity (SSIM) Index and Root Mean Square Error (RMSE) for SNR of data being 20, 40 and 60 dB. Bold text indicates the best possible value for the metric used.

SNR Method/ Metric	20 dB			40 dB			60 dB		
	PC	SSIM	RMSE	PC	SSIM	RMSE	PC	SSIM	RMSE
Limited BW	0.5388	0.4287	0.1515	0.5804	0.4858	0.1623	0.5804	0.4858	0.1623
Interpolated	0.7720	0.6208	0.1176	0.8156	0.6874	0.1446	0.8158	0.6877	0.1442
Automated Wavelet Denoising	0.8158	0.6877	0.1442	0.8158	0.6877	0.1442	0.8158	0.6877	0.1442
SRCNN	-0.4389	-0.0619	0.2406	-0.4715	-0.0335	0.2436	-0.4714	-0.0326	0.2460
U-Net (Relu)	-0.0836	0.0787	0.2833	-0.0387	0.1574	0.2835	-0.0372	0.1586	0.2870
U-Net (Elu)	0.8114	0.7007	0.1133	0.8893	0.8125	0.0947	0.8949	0.8188	0.0937
U-Net (Hybrid)	0.8219	0.6898	0.1286	0.8947	0.8039	0.0852	0.8976	0.8058	0.0854

in Fig. 4 (o)-(u) respectively. The improvement obtained in the PC, SSIM and RMSE was 21.36%, 33.72% and 29.54% respectively.

The reconstruction obtained using the limited data for the Derenzo phantom using 100 detectors is shown in Fig. 5 (a). The reconstruction obtained after interpolation and then denoising are shown in Figs. 5 (b) and 5 (c) respectively for the Derenzo phantom with SNR of data being 20 dB. The reconstruction obtained for the Derenzo phantom using 200 detectors data using SRCNN is shown in Fig. 5 (d), U-Net (Relu) is shown in Fig. 5 (e), U-Net (Elu) is shown in Fig. 5 (f), and the proposed model U-Net (Hybrid) is shown in Fig. 5 (g) for this case of SNR being 20 dB. The improvement obtained in PC, SSIM and RMSE as compared to the interpolated data case are 22.55%, 29.30% and 12.40%

respectively. The improvement was similar to the case of 'PAT' phantom. Similarly the reconstruction obtained for the case of SNR of data being 40 dB, interpolated, denoised, SRCNN, U-Net (Relu), U-Net (Elu), and the proposed technique are shown in Figs. 5 (h)-(n) respectively. The improvement obtained in the PC, SSIM and RMSE was 18.00%, 22.97% and 13.04% respectively. Similarly the reconstruction obtained for SNR of data being 60 dB for the discussed methods in this work are shown in Fig. 5 (o)-(u) respectively. The improvement obtained in the PC, SSIM and RMSE was 18.00%, 23.03% and 12.90% respectively. The background streak artifact reduction was well achieved using the proposed model compared to other methods.

The reconstruction obtained using the limited data for the blood vessel phantom using 100 detectors is shown in

Fig. 7: Reconstructed Breast phantom with each row representing the SNR of data being 20, 40, and 60 dB and correspondingly given in (a)-(g), (h)-(n) and (o)-(u) respectively. Reconstructed images using data with limited BW (70%) and 100 transducers are shown in (a), (h), and (o). While (b), (i), and (p) show reconstruction results when the data with 100 transducers that are interpolated using nearest-neighbour to obtain a 200 transducer data. (c), (j) and (q) show results obtained using the automated wavelet de-noising approach. The results obtained using the SRCNN based method are shown in (d), (k), and (r) and the results obtained using the U-Net (Relu) are shown in (e), (l), and (s). Finally, (f), (m), and (t) show results obtained using the proposed U-Net (Elu) approach and (g), (n), and (u) show the results obtained using the proposed U-Net (Hybrid) based approach. Table-IV, reports the figure of merits of these results in a comparative manner.

Fig. 8: Reconstructed images for *in-vivo* horse hair phantoms for (a) limited BW, 100 transducers, (b) interpolated using nearest neighbour having limited BW but 200 detectors, (c) de-noising using automated wavelet approach, (d) SRCNN, (e) U-Net (Relu), (f) proposed U-Net (Elu), (g) proposed U-Net (Hybrid) approach have been shown. Note that corresponding SNR_r (dB) values have been mentioned below the image.

Fig. 6 (a). The reconstruction obtained after interpolation and then denoising are shown in Figs. 6 (b) and 6 (c) respectively for the blood vessel phantom for the case of SNR of data being 20 dB. The reconstruction obtained for the blood vessel phantom using 200 detectors data using SRCNN is shown in Fig. 6 (d), U-Net (Relu) is shown in Fig. 6 (e), U-Net (Elu) is shown in Fig. 6 (f), and the proposed model U-Net (Hybrid) is shown in Fig. 6 (g) for the case of SNR of data being 20 dB. The improvement obtained in PC, SSIM and RMSE as compared to the interpolated data case are 23.83%, 30.25% and -3.46% respectively. Similarly the reconstructed photoacoustic image obtained with SNR of data being 40 dB, interpolated, denoised, SRCNN, U-Net (Relu), U-Net (Elu), and the proposed technique are shown in Figs. 6 (h)-

(n) respectively. The improvement obtained in the PC, SSIM and RMSE was 20.96%, 26.11% and 5.98% respectively. The reconstruction obtained for SNR of data being 60 dB for the compared techniques are shown in Fig. 6 (o)-(u) respectively. The improvement obtained in the PC, SSIM and RMSE was 21.13%, 26.36% and 8.41% respectively.

Finally, complex photoacoustic data with variable contrast and noise was used to test the ability of the proposed approach to generalise and adapt to the dynamic environment of any experimental setup. The reconstruction obtained using the limited data for the Breast phantom using 100 detectors is shown in Fig. 7 (a). The reconstruction obtained after interpolation and then denoising are shown in Figs. 7 (b) and 7 (c) respectively for the breast phantom corresponding to the

Fig. 9: Reconstructed images for *in-vivo* rat brain data for (a) limited BW, 100 transducers, (b) interpolated using nearest neighbour having limited BW but 200 detectors, (c) de-noising using automated wavelet approach, (d) SRCNN, (e) U-Net (Relu), (f) proposed U-Net (Elu), (g) proposed U-Net (Hybrid) approach have been shown. Note that corresponding SNR_r (dB) values have been mentioned below the image.

case of SNR of data being 20 dB. The reconstruction obtained for this case using 200 detectors data using SRCNN is shown in Fig. 7 (d), U-Net (Relu) is shown in Fig. 7 (e), U-Net (Elu) is shown in Fig. 7 (f), and the proposed model U-Net (Hybrid) is shown in Fig. 7 (g). The improvement obtained in PC, SSIM and RMSE as compared to the interpolated data case are 6.46%, 11.11% and -9.35% respectively. Similarly the reconstruction obtained for SNR of data being 40 dB, interpolated, denoised, SRCNN, U-Net (Relu), U-Net (Elu), and the proposed technique are shown in Figs. 7 (h)-(n) respectively. The improvement obtained in the PC, SSIM and RMSE was 9.69%, 16.94% and 41.07% respectively. The reconstruction obtained for the case of SNR of data being 60 dB for the discussed methods in this work are shown in Fig. 7 (o)-(u) respectively. The improvement obtained in the PC, SSIM and RMSE was 10.02%, 17.17% and 40.77% respectively. Note that U-Net (Relu) (Figs. 7 (e), (l), and (s)) completely fails to model the variable contrast, despite of being trained using the same data as other models. This reinforces the robustness of ELUs in deep neural networks for photoacoustic imaging tasks.

These numerical results indicate that the proposed method was effective in terms of providing good quality photoacoustic data and thus able to provide superior results compared to the existing classical and deep learning based approaches.

B. Experimental Phantoms

The reconstructed photoacoustic images obtained for the horse hair phantom are shown in Fig. 8 (a). The reconstructed results for the interpolated and the denoised data are shown in Fig. 8 (b) and (c) respectively. The reconstructed results obtained using SRCNN is shown in Fig. 8 (d), U-Net (Relu) is shown in Fig. 8 (e), U-Net (Elu) is shown in Fig. 8 (f), and the proposed model U-Net (Hybrid) is shown in Fig. 8 (g). The reconstruction obtained using the proposed technique gives a clear background which is free of artefacts as compared to the other techniques. The SNR_r is improved as compared to the limited bandwidth data reconstruction result by 11.65 dB.

Similar results are obtained for the *in-vivo* rat brain data as shown in the Fig. 9 (a) for hundred detectors data. The images obtained using the interpolated and the denoised method are

shown in Fig. 9 (b) and (c) respectively. The reconstructed results obtained using SRCNN is shown in Fig. 9 (d), U-Net (Relu) is shown in Fig. 9 (e), U-Net (Elu) is shown in Fig. 9 (f), and the proposed model U-Net (Hybrid) is shown in Fig. 9 (g). Again the results obtained using the proposed architecture have shown a artefact free background and hence there is a clear improvement in the image quality. The improvement in SNR_r obtained as compared to the limited data reconstruction is 1.45 dB.

Note that in this work, only linear back projection was utilized for characterising the improvement in data (including bandwidth enhancement, super-resolution and denoising). It might be possible to utilize the model-based schemes, which can provide further improvement in the reconstructed image quality. As the discussion here is limited to improvement of boundary data using deep learning, only analytical reconstruction method (linear back projection) was deployed here to reinforce that improvement in the reconstructed image quality is purely due to improvement in the sinogram data.

VI. CONCLUSION

Deep learning based methods have been explored to improve photoacoustic imaging, where these methods have been deployed in the image space. This work deploys a U-Net based architecture for super-resolution and bandwidth enhancement of sinogram data (boundary photoacoustic data) and in turn it provides improved photoacoustic image quality. The performance of the proposed network with loss function being scaled RMSE, that is capable of providing much needed generalization in the photoacoustic imaging, was systematically compared in both experimental and numerical scenarios with existing classical and deep learning based approaches. The background was free from the streak artefacts and hence superior quality of image reconstruction was obtained. This technique can be applied to other imaging modalities, such as ultrasound tomography, for performing bandwidth enhancement in the data domain. The improvement of image quality due to utilization of the proposed method in the data space in terms of SNR of reconstructed photoacoustic image SNR_r for the *in-vivo* phantoms was as high as 11.65 dB as compared to the reconstruction obtained using data having limited bandwidth. Methods of the type proposed, which are

purely data-driven and having computational complexity being extremely low (for running the trained model), will pave a way for making photoacoustic imaging forerunner in the pre-clinical as well as clinical settings.

REFERENCES

- [1] S. Arridge and O. Scherzer, "Imaging from coupled physics," *Inverse Problems*, vol. 28, no. 8, p. 080201, 2012.
- [2] L. V. Wang and S. Hu, "Photoacoustic tomography: in vivo imaging from organelles to organs," *Science*, vol. 335, no. 6075, pp. 1458–1462, 2012.
- [3] J. Xia, C. Huang, K. Maslov, M. A. Anastasio, and L. V. Wang, "Enhancement of photoacoustic tomography by ultrasonic computed tomography based on optical excitation of elements of a full-ring transducer array," *Optics letters*, vol. 38, no. 16, pp. 3140–3143, 2013.
- [4] M. Pramanik, G. Ku, C. Li, and L. V. Wang, "Design and evaluation of a novel breast cancer detection system combining both thermoacoustic (ta) and photoacoustic (pa) tomography," *Medical Physics*, vol. 35, no. 6, pp. 2218–2223, 2008.
- [5] L. V. Wang, "Tutorial on photoacoustic microscopy and computed tomography," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 14, no. 1, pp. 171–179, 2008.
- [6] P. K. Upputuri and M. Pramanik, "Recent advances toward preclinical and clinical translation of photoacoustic tomography: a review," *Journal of Biomedical Optics*, vol. 22, no. 4, p. 041006, 2017.
- [7] L. Li, L. Zhu, C. Ma, L. Lin, J. Yao, L. Wang, K. Maslov, R. Zhang, W. Chen, J. Shi, and L. V. Wang, "Single-impulse panoramic photoacoustic computed tomography of small-animal whole-body dynamics at high spatiotemporal resolution," *Nature Biomedical Engineering*, vol. 1, no. 5, p. 0071, 2017.
- [8] Y. Zhou, J. Yao, and L. V. Wang, "Tutorial on photoacoustic tomography," *Journal of Biomedical Optics*, vol. 21, no. 6, p. 061007, 2016.
- [9] J. L. Su, B. Wang, K. E. Wilson, C. L. Bayer, Y.-S. Chen, S. Kim, K. A. Homan, and S. Y. Emelianov, "Advances in clinical and biomedical applications of photoacoustic imaging," *Expert Opinion on Medical Diagnostics*, vol. 4, no. 6, pp. 497–510, 2010.
- [10] K. Wang, R. Su, A. A. Oraevsky, and M. A. Anastasio, "Investigation of iterative image reconstruction in three-dimensional photoacoustic tomography," *Physics in Medicine and Biology*, vol. 57, no. 17, pp. 5399–5423, 2012.
- [11] A. Rosenthal, V. Ntziachristos, and D. Razansky, "Acoustic inversion in photoacoustic tomography: A review," *Current Medical Imaging Reviews*, vol. 9, no. 4, pp. 318–336, 2013.
- [12] C. Huang, K. Wang, L. Nie, L. V. Wang, and M. A. Anastasio, "Full-wave iterative image reconstruction in photoacoustic tomography with acoustically inhomogeneous media," *IEEE Transactions on Medical Imaging*, vol. 32, no. 6, pp. 1097–1110, 2013.
- [13] K. Wang, S. A. Ermilov, R. Su, H.-P. Brecht, A. A. Oraevsky, and M. A. Anastasio, "An imaging model incorporating ultrasonic transducer properties for three-dimensional photoacoustic tomography," *IEEE Transactions on Medical Imaging*, vol. 30, no. 2, pp. 203–214, 2011.
- [14] Y. Xu, L. V. Wang, G. Ambartsoumian, and P. Kuchment, "Reconstructions in limited-view thermoacoustic tomography," *Medical Physics*, vol. 31, no. 4, pp. 724–733, 2004.
- [15] S. R. Arridge, M. M. Betcke, B. T. Cox, F. Lucka, and B. E. Treeby, "On the adjoint operator in photoacoustic tomography," *Inverse Problems*, vol. 32, no. 11, p. 115012, 2016.
- [16] S. Arridge, P. Beard, M. Betcke, B. Cox, N. Huynh, F. Lucka, O. Ogunlade, and E. Zhang, "Accelerated high-resolution photoacoustic tomography via compressed sensing," *Physics in Medicine and Biology*, vol. 61, no. 24, pp. 8908–8940, 2016.
- [17] N. A. Rejesh, H. Pullagurra, and M. Pramanik, "Deconvolution-based deblurring of reconstructed images in photoacoustic/thermoacoustic tomography," *Journal of the Optical Society of America A*, vol. 30, no. 10, pp. 1994–2001, 2013.
- [18] S. M. Riad, "The deconvolution problem: An overview," *Proceedings of the IEEE*, vol. 74, no. 1, pp. 82–85, 1986.
- [19] M. Cao, T. Feng, J. Yuan, G. Xu, X. Wang, and P. L. Carson, "Spread spectrum photoacoustic tomography with image optimization," *IEEE transactions on biomedical circuits and systems*, vol. 11, no. 2, pp. 411–419, 2016.
- [20] B. Bier, M. Unberath, J.-N. Zaech, J. Fotouhi, M. Armand, G. Osgood, N. Navab, and A. Maier, "X-ray-transform invariant anatomical landmark detection for pelvic trauma surgery," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2018, pp. 55–63.
- [21] M. Aubreville, M. Stoeve, N. Oetter, M. Goncalves, C. Knipfer, H. Neumann, C. Bohr, F. Stelzle, and A. Maier, "Deep learning-based detection of motion artifacts in probe-based confocal laser endomicroscopy images," *International journal of computer assisted radiology and surgery*, vol. 14, no. 1, pp. 31–42, 2019.
- [22] A. Akselrod-Ballin, L. Karlinsky, S. Alpert, S. Hasoul, R. Ben-Ari, and E. Barkan, "A region based convolutional network for tumor detection and classification in breast mammography," in *Deep Learning and Data Labeling for Medical Applications*. Springer, 2016, pp. 197–205.
- [23] F.-C. Ghesu, B. Georgescu, Y. Zheng, S. Grbic, A. Maier, J. Hornegger, and D. Comaniciu, "Multi-scale deep reinforcement learning for real-time 3d-landmark detection in ct scans," *IEEE transactions on pattern analysis and machine intelligence*, vol. 41, no. 1, pp. 176–189, 2017.
- [24] H. R. Roth, L. Lu, A. Farag, H.-C. Shin, J. Liu, E. B. Turkbey, and R. M. Summers, "Deeporgan: Multi-level deep convolutional networks for automated pancreas segmentation," in *International conference on medical image computing and computer-assisted intervention*. Springer, 2015, pp. 556–564.
- [25] P. Moeskops, M. A. Viergever, A. M. Mendrik, L. S. de Vries, M. J. Benders, and I. Išgum, "Automatic segmentation of mr brain images with a convolutional neural network," *IEEE transactions on medical imaging*, vol. 35, no. 5, pp. 1252–1261, 2016.
- [26] S. Chen, H. Roth, S. Dorn, M. May, A. Cavallaro, M. M. Leil, M. Kachelrieß, H. Oda, K. Mori, and A. Maier, "Towards automatic abdominal multi-organ segmentation in dual energy ct using cascaded 3d fully convolutional network," *arXiv preprint arXiv:1710.05379*, 2017.
- [27] J. J. Nirschl, A. Janowczyk, E. G. Peyster, R. Frank, K. B. Margulies, M. D. Feldman, and A. Madabhushi, "Deep learning tissue segmentation in cardiac histopathology images," in *Deep Learning for Medical Image Analysis*. Elsevier, 2017, pp. 179–195.
- [28] G. Wang, J. C. Ye, K. Mueller, and J. A. Fessler, "Image reconstruction is a new frontier of machine learning," *IEEE transactions on medical imaging*, vol. 37, no. 6, pp. 1289–1296, 2018.
- [29] M. T. McCann, K. H. Jin, and M. Unser, "Convolutional neural networks for inverse problems in imaging: A review," *IEEE Signal Processing Magazine*, vol. 34, no. 6, pp. 85–95, 2017.
- [30] Z. Zhang, X. Liang, X. Dong, Y. Xie, and G. Cao, "A sparse-view ct reconstruction method based on combination of densenet and deconvolution," *IEEE transactions on medical imaging*, vol. 37, no. 6, pp. 1407–1417, 2018.
- [31] N. Awasthi, R. Pardasani, S. K. Kalva, M. Pramanik, and P. K. Yalavarthy, "Sinogram super-resolution and denoising convolutional neural network (srcn) for limited data photoacoustic tomography," *arXiv preprint arXiv:2001.06434*, 2020.
- [32] S. Antholzer, M. Haltmeier, and J. Schwab, "Deep learning for photoacoustic tomography from sparse data," *Inverse problems in science and engineering*, vol. 27, no. 7, pp. 987–1005, 2019.
- [33] A. Reiter and M. A. L. Bell, "A machine learning approach to identifying point source locations in photoacoustic data," in *Photons Plus Ultrasound: Imaging and Sensing 2017*, vol. 10064. International Society for Optics and Photonics, 2017, p. 100643J.
- [34] D. Allman, A. Reiter, and M. A. L. Bell, "Photoacoustic source detection and reflection artifact removal enabled by deep learning," *IEEE Transactions on Medical Imaging*, vol. 37, no. 6, pp. 1464–1477, 2018.
- [35] A. Hauptmann, F. Lucka, M. Betcke, N. Huynh, J. Adler, B. Cox, P. Beard, S. Ourselin, and S. Arridge, "Model-based learning for accelerated, limited-view 3-d photoacoustic tomography," *IEEE Transactions on Medical Imaging*, vol. 37, no. 6, pp. 1382–1393, 2018.
- [36] N. Davoudi, X. L. Deán-Ben, and D. Razansky, "Deep learning photoacoustic tomography with sparse data," *Nature Machine Intelligence*, vol. 1, no. 10, pp. 453–460, 2019.
- [37] H. Lee, J. Lee, and S. Cho, "View-interpolation of sparsely sampled sinogram using convolutional neural network," in *Medical Imaging 2017: Image Processing*, vol. 10133. International Society for Optics and Photonics, 2017, p. 1013328.
- [38] H. Lee, J. Lee, H. Kim, B. Cho, and S. Cho, "Deep-neural-network-based sinogram synthesis for sparse-view ct image reconstruction," *IEEE Transactions on Radiation and Plasma Medical Sciences*, vol. 3, no. 2, pp. 109–119, 2018.
- [39] J. Kim, J. Kwon Lee, and K. Mu 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.
- [40] X. Mao, C. Shen, and Y.-B. Yang, “Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections,” in *Advances in Neural Information Processing Systems*, 2016, pp. 2802–2810.
 - [41] J. Kim, J. Kwon Lee, and K. Mu Lee, “Deeply-recursive convolutional network for image super-resolution,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 1637–1645.
 - [42] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang *et al.*, “Photo-realistic single image super-resolution using a generative adversarial network,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 4681–4690.
 - [43] N. Awasthi, K. R. Prabhakar, S. K. Kalva, M. Pramanik, R. V. Babu, and P. K. Yalavarthy, “Pa-fuse: deep supervised approach for the fusion of photoacoustic images with distinct reconstruction characteristics,” *Biomedical Optics Express*, vol. 10, no. 5, pp. 2227–2243, 2019.
 - [44] S. Gutta, V. S. Kadimesetty, S. K. Kalva, M. Pramanik, S. Ganapathy, and P. K. Yalavarthy, “Deep neural network-based bandwidth enhancement of photoacoustic data,” *Journal of biomedical optics*, vol. 22, no. 11, p. 116001, 2017.
 - [45] C. B. Shaw, J. Prakash, M. Pramanik, and P. K. Yalavarthy, “Least squares qr-based decomposition provides an efficient way of computing optimal regularization parameter in photoacoustic tomography,” *Journal of Biomedical Optics*, vol. 18, no. 8, p. 080501, 2013.
 - [46] N. Awasthi, S. K. Kalva, M. Pramanik, and P. K. Yalavarthy, “Vector extrapolation methods for accelerating iterative reconstruction methods in limited-data photoacoustic tomography,” *Journal of Biomedical Optics*, vol. 23, no. 7, p. 071204, 2018.
 - [47] C. G. Hoelen and F. F. de Mul, “Image reconstruction for photoacoustic scanning of tissue structures,” *Applied Optics*, vol. 39, no. 31, pp. 5872–5883, 2000.
 - [48] D. B. Percival and A. T. Walden, *Wavelet methods for time series analysis*. Cambridge university press, 2006, vol. 4.
 - [49] S. H. Holan and J. A. Viator, “Automated wavelet denoising of photoacoustic signals for circulating melanoma cell detection and burn image reconstruction,” *Physics in Medicine & Biology*, vol. 53, no. 12, p. N227, 2008.
 - [50] D. L. Donoho and J. M. Johnstone, “Ideal spatial adaptation by wavelet shrinkage,” *Biometrika*, vol. 81, no. 3, pp. 425–455, 1994.
 - [51] M. Irani and S. Peleg, “Improving resolution by image registration,” *CVGIP: Graphical models and image processing*, vol. 53, no. 3, pp. 231–239, 1991.
 - [52] G. Freedman and R. Fattal, “Image and video upscaling from local self-examples,” *ACM Transactions on Graphics (TOG)*, vol. 30, no. 2, p. 12, 2011.
 - [53] D. Glasner, S. Bagon, and M. Irani, “Super-resolution from a single image,” in *2009 IEEE 12th International Conference on Computer Vision*. IEEE, pp. 349–356.
 - [54] J. Yang, Z. Lin, and S. Cohen, “Fast image super-resolution based on in-place example regression,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2013, pp. 1059–1066.
 - [55] M. Bevilacqua, A. Roumy, C. Guillemot, and M. L. Alberi-Morel, “Low-complexity single-image super-resolution based on nonnegative neighbor embedding,” 2012.
 - [56] H. Chang, D.-Y. Yeung, and Y. Xiong, “Super-resolution through neighbor embedding,” in *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004.*, vol. 1. IEEE, 2004, pp. I–I.
 - [57] W. T. Freeman, E. C. Pasztor, and O. T. Carmichael, “Learning low-level vision,” *International Journal of Computer Vision*, vol. 40, no. 1, pp. 25–47, 2000.
 - [58] J. Yang, J. Wright, T. Huang, and Y. Ma, “Image super-resolution as sparse representation of raw image patches,” in *2008 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2008, pp. 1–8.
 - [59] J. Yang, J. Wright, T. S. Huang, and Y. Ma, “Image super-resolution via sparse representation,” *IEEE Transactions on Image Processing*, vol. 19, no. 11, pp. 2861–2873, 2010.
 - [60] C. Dong, C. C. Loy, K. He, and X. Tang, “Learning a deep convolutional network for image super-resolution,” in *European Conference on Computer Vision*. Springer, 2014, pp. 184–199.
 - [61] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.
 - [62] D.-A. Clevert, T. Unterthiner, and S. Hochreiter, “Fast and accurate deep network learning by exponential linear units (elus),” *arXiv preprint arXiv:1511.07289*, 2015.
 - [63] Y. Lou, W. Zhou, T. P. Matthews, C. M. Appleton, and M. A. Anastasio, “Generation of anatomically realistic numerical phantoms for photoacoustic and ultrasonic breast imaging,” *Journal of biomedical optics*, vol. 22, no. 4, p. 041015, 2017.
 - [64] P. Burgholzer, J. Bauer-Marschallinger, H. Grün, M. Haltmeier, and G. Paltauf, “Temporal back-projection algorithms for photoacoustic tomography with integrating line detectors,” *Inverse Problems*, vol. 23, no. 6, p. S65, 2007.
 - [65] P. Mickevičius, S. Narang, J. Alben, G. Diamos, E. Elsen, D. Garcia, B. Ginsburg, M. Houston, O. Kuchaiev, G. Venkatesh *et al.*, “Mixed precision training,” *arXiv preprint arXiv:1710.03740*, 2017.
 - [66] N. Gandhi, M. Allard, S. Kim, P. Kazanzides, and M. A. L. Bell, “Photoacoustic-based approach to surgical guidance performed with and without a da vinci robot,” *Journal of Biomedical Optics*, vol. 22, no. 12, p. 121606, 2017.
 - [67] P. P. Pai, A. De, and S. Banerjee, “Accuracy enhancement for noninvasive glucose estimation using dual-wavelength photoacoustic measurements and kernel-based calibration,” *IEEE Transactions on Instrumentation and Measurement*, vol. 67, no. 1, pp. 126–136, 2017.
 - [68] Z. Wang, A. C. Bovik, H. R. Sheikh, E. P. Simoncelli *et al.*, “Image quality assessment: from error visibility to structural similarity,” *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.
 - [69] S. Guan, A. Khan, S. Sikdar, and P. Chitnis, “Fully dense unet for 2d sparse photoacoustic tomography artifact removal,” *IEEE Journal of Biomedical and Health Informatics*, 2019.
 - [70] J. Prakash, A. S. Raju, C. B. Shaw, M. Pramanik, and P. K. Yalavarthy, “Basis pursuit deconvolution for improving model-based reconstructed images in photoacoustic tomography,” *Biomedical Optics Express*, vol. 5, no. 5, pp. 1363–1377, 2014.
 - [71] L. Li, L. Zhu, Y. Shen, and L. V. Wang, “Multiview hilbert transformation in full-ring transducer array-based photoacoustic computed tomography,” *Journal of Biomedical Optics*, vol. 22, no. 7, p. 076017, 2017.
 - [72] N. Awasthi, S. K. Kalva, M. Pramanik, and P. K. Yalavarthy, “Image-guided filtering for improving photoacoustic tomographic image reconstruction,” *Journal of Biomedical Optics*, vol. 23, no. 9, p. 091413, 2018.
 - [73] B. E. Treeby and B. T. Cox, “k-wave: Matlab toolbox for the simulation and reconstruction of photoacoustic wave fields,” *Journal of Biomedical Optics*, vol. 15, no. 2, p. 021314, 2010.
 - [74] Y. Lou, K. Mitsuhashi, C. M. Appleton, A. Oraevsky, and M. A. Anastasio, “Generation of anatomically realistic numerical phantoms for photoacoustic breast imaging,” in *Photons Plus Ultrasound: Imaging and Sensing 2016*, vol. 9708. International Society for Optics and Photonics, 2016, p. 97084O.
 - [75] S. K. Kalva and M. Pramanik, “Experimental validation of tangential resolution improvement in photoacoustic tomography using modified delay-and-sum reconstruction algorithm,” *Journal of Biomedical Optics*, vol. 21, no. 8, p. 086011, 2016.
 - [76] S. K. Kalva, P. K. Upputuri, and M. Pramanik, “High-speed, low-cost, pulsed-laser-diode-based second-generation desktop photoacoustic tomography system,” *Optics Letters*, vol. 44, no. 1, pp. 81–84, 2019.