applied sciences

MDPI

Article

Accelerated Correction of Reflection Artifacts by Deep Neural Networks in Photo-Acoustic Tomography

Hongming Shan [1], Ge Wang [1] and Yang Yang [2,*]

[1] Biomedical Imaging Center, Department of Biomedical Engineering, Rensselaer Polytechnic Institute, Troy, NY 12180, USA
[2] Department of Computational Mathematics, Science and Engineering, Michigan State University, East Lansing, MI 48824, USA
* Correspondence: yangy5@msu.edu; Tel.: +1-517-432-0620

Received: 20 May 2019; Accepted: 18 June 2019; Published: 28 June 2019

Abstract: Photo-Acoustic Tomography (PAT) is an emerging non-invasive hybrid modality driven by a constant yearning for superior imaging performance. The image quality, however, hinges on the acoustic reflection, which may compromise the diagnostic performance. To address this challenge, we propose to incorporate a deep neural network into conventional iterative algorithms to accelerate and improve the correction of reflection artifacts. Based on the simulated PAT dataset from computed tomography (CT) scans, this network-accelerated reconstruction approach is shown to outperform two state-of-the-art iterative algorithms in terms of the peak signal-to-noise ratio (PSNR) and the structural similarity (SSIM) in the presence of noise. The proposed network also demonstrates considerably higher computational efficiency than conventional iterative algorithms, which are time-consuming and cumbersome.

Keywords: photo-acoustic tomography; reflection artifacts; deep learning; convolutional neural network; time reversal; Landweber algorithm; U-net

1. Introduction

Photo-Acoustic Tomography (PAT) is an emerging non-invasive modality that has manifested an enormous prospect for some clinical practices [1]. In PAT, the tissue is illuminated with near-infrared light of wavelength 650–900 nm. The absorbed optical energy is transformed into acoustic energy through the photo-acoustic effect, and the generated ultrasound is measured by transducer arrays outside the tissue to retrieve the optical properties of the tissue. The coupling mechanism of optical and ultrasound waves gives multiple advantages over conventional individual imaging modalities. As the acoustic waves experience much less scattering in tissue compared to optical waves, PAT can generate high-resolution images in the presence of strong optical scattering to break the optical diffusion limit [2]. The image can reach submillimeter spatial resolution while preserving intrinsic optical contrast [3].

Typical photo-acoustic signal generation comprises three steps: (1) the tissue absorbs light; (2) the absorbed optical energy causes a temperature rise; (3) thermo-elastic expansion occurs and generates ultrasound. The image formation in PAT is to recover the distribution of the deposited energy, known as the local optical fluence, from the ultrasound signals that are recorded by the sensors deployed around the tissue. As the initial ultrasound pressure is approximately proportional to the optical fluence, it is sufficient to reconstruct the initial pressure from the recorded ultrasound signals.

The quality of PAT images relies on multiple factors, and one of them is the acoustic reflection. The reflection can be caused by either hyperechoic structures or special setting for making a measurement such as in reverberant-field PAT [4]. Conventional PAT reconstruction algorithms, such as the universal back-projection formula [5] and time-reversal-based reconstruction [6–10],

are based on the spherical mean radon transform on canonical geometries and not designed to take the acoustic reflections into account. As a result, the reflected signals are projected along with the real signal back into the image domain, resulting in artifacts that are indistinguishable from the real biological structures. Clinical practitioners who rely on the misleading artifacts to make judgment could come to erroneous conclusions. It is, therefore, essential and of practical significance to design new methods to eliminate the impact of such acoustic reflection in PAT image reconstruction.

Some iterative algorithms were developed to correct the effect of the acoustic reflection. Examples include Fourier analysis [11], averaged time reversal [12], dissipative time reversal [13], and adjoint method [14]. Nonetheless, each of them has its limitation: the Fourier approach applies only to regular domains; averaged/dissipative time reversals, and Landweber iterations can generate high-quality images but time-consuming.

The purpose of this paper is to accelerate and improve the iterative algorithms by incorporating a deep neural network (DNN) structure to reduce the reflection artifacts efficiently and effectively. The entire network contains three parts—feature extraction, artifacts reduction, and reconstruction. A modified version of the U-net with large convolutional filter sizes is used as the backbone of the artifact reduction part to capture the global features by increasing the size of the receptive field. The network is applied to take the place of iterations to accelerate the correction of reflection artifacts. The learning process can simultaneously reduce the image noise as well.

With the rapid increment of computational power, DNNs and deep learning techniques have received considerable attention in recent years for tomographic image reconstruction [15–17]. In particular, they have achieved the state-of-the-art performance in PAT image reconstruction in the scenarios of sparse data [18–20], limited view [21–23], artifacts removal [24–26], as well as other applications [27,28]. It is worth pointing out that the reference [25] considers another type of reflection artifact with different focuses and approaches. The reflection in [25] is caused by point-like targets inside the tissue, while ours by planar reflectors outside the tissue. The network in [25] is mostly based on convolutional neural networks (CNNs) while ours on the U-net structure. The two networks target different problems with associated unique networks and therefore are of independent values. We remark that besides in PAT, DNNs have been successfully applied to other imaging modalities as well, see [29–35] for its applications in CT.

2. Data Generation

This section presents the forward model of PAT and the reflection artifacts induced by the acoustic reflection of signals.

2.1. The Forward Model

We describe the forward model, which is used to generate ultrasound signals for training purpose. Sound hard reflectors, which can bounce the incident ultrasound back with the opposite speed without absorbing any energy, are placed around the tissue to simulate the acoustic reflections. As such reflectors do not dissipate energy, the resultant reflection artifacts are the strongest, hence can be used as a benchmark to test the artifact-correction capability of different algorithms.

Denote by Ω the biological tissue, which is illuminated by rapid pulses of laser, and afterwards, emits the ultrasound. The boundary of the tissue is represented using the conventional notation $\partial\Omega$. The forward ultrasound propagation model, in the presence of sound hard reflectors, reads in 2D as follows:

$$\begin{cases} \left(\partial_t^2 - c^2(r)\Delta\right)p &= 0 \qquad \text{in } (0,T)\times\Omega, \\ p|_{t=0} &= p_0(r), \\ \partial_t p|_{t=0} &= 0, \\ \partial_\nu p|_{[0,T]\times\partial\Omega} &= 0, \end{cases} \tag{1}$$

where $\Delta = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}$ is the Laplace operator, T is the stoppage time, $p_0(r)$ is the initial ultrasound pressure, $p(t,r)$ is the ultrasound pressure of the spatial location r at time t, $c(r)$ is the wave speed, and ∂_ν is the normal boundary derivative. The main difference of this model with the conventional photo-acoustic models (see, for instance [2,36]) is the boundary condition $\partial_\nu p|_{[0,T]\times\partial\Omega} = 0$, which is the standard mathematical model to describe sound hard reflectors. Please note that the Neumann boundary condition corresponds to the sample/water interface, and the Dirichlet boundary condition, which corresponds to the sample/air interface, can be handled similarly. We remark that modeling PAT using the Neumann and Dirichlet boundary conditions was reported in the literature [37,38]. The ultrasound is recorded at the boundary of Ω using 508 sensors deployed evenly around the tissue. This measurement amounts to the temporal boundary value of p, i.e., $p|_{[0,T]\times\partial\Omega}$. The image formation in PAT aims to recover the distribution of the deposited energy that is the local optical fluence. As the initial ultrasound pressure $p_0(r)$ is proportional to the optical fluence, it remains to reconstruct the initial ultrasound pressure $p_0(r)$ from the recorded ultrasound signals.

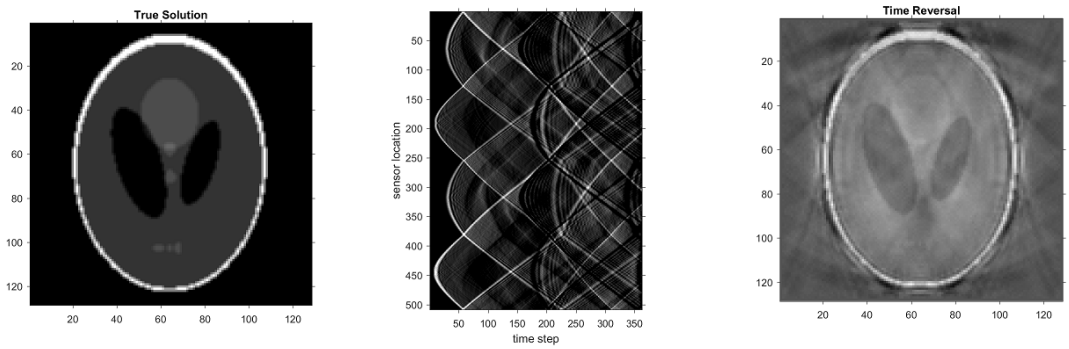
We simulate numerous ultrasound signals to train the neural network. This is accomplished by solving the forward model (1) using the second order finite difference time domain (FDTD) method with the central difference formula. The simulation is implemented using the MATLAB code written by the authors. The computational domain is a 128×128 grid with uniform spacing. The time step is chosen according to the spatial step to fulfill the Courant–Friedrichs–Lewy (CFL) condition. The sound speed $c(r)$ is either a constant or a spatially varying function, with the former modeling the propagation in homogeneous media and the latter in heterogeneous media. The stoppage time is $T = 4$, which is chosen in such a way that the ultrasound generated from any interior source can reach the boundary.

2.2. Reflection Artifacts

Applying the conventional PAT reconstruction directly to the ultrasound signals generated by the forward model (1) could result in artifacts. This is because the conventional reconstruction methods do not take into consideration of the acoustic reflections. We provide a numerical example in this section to illustrate the effect of the acoustic reflections.

We choose the Shepp–Logan phantom (see Figure 1a) as the initial pressure $p_0(r)$ to produce ultrasound signals based on the forward model (1) with $c(r) = 1$. The recorded acoustic signal is shown in Figure 1b. To demonstrate the reflection artifacts by conventional algorithms, we adopt one of the conventional time reversal (TR) methods proposed in [8]. The sound speed in the TR is again $c(r) = 1$. Then the TR method is mathematically equivalent to the 2D universal back-projection formula. The stoppage time is again $T = 4$. The reconstructed image is illustrated in Figure 1c. It is clear that direct application of the TR method leads to numerous artifacts in the reconstructed image, especially near the boundary of the ellipses in the Shepp–Logan phantom. The generation of these artifacts can actually be well understood from the mathematical point of view; see the analysis in [12] for details.

The failure of the conventional TR method indicates that some procedures must be introduced to correct the reflection artifacts in the presence of reflections. There have been some iterative correction algorithms, for instance [10–14,39] as well as methods based on deep learning [25]. In the next section, we propose a novel neural network to remove the reflection artifacts. Its efficacy is compared with two of the popular iterative correction algorithms as well.



(**a**) (**b**) (**c**)

Figure 1. Numerical example to illustrate the effect of the acoustic reflections. (**a**): a 128 × 128 Shepp–Logan phantom as the initial pressure. (**b**): the recorded acoustic signal in the presence of sound hard reflectors. (**c**): reconstruction by conventional TR.

3. Artifacts Correction by Deep Learning

This section presents the proposed neural network for correction of reflection artifacts in PAT imaging. We will make use of the networks to accelerate some conventional iterative algorithms in the following way. Instead of running the iterations, we train the network to learn the map between the reconstructed image by the first iteration and the ground truth. The reconstruction procedure then consists of two steps, given the measured ultrasound signals. The first step is to apply an iterative algorithm to the signals to get the output after the first iteration. The second step is to feed the output into the neural network to obtain the reconstructed image. These two steps preserve the predictability of the model-based iterative algorithm, while at the same time replace the number of iterations by the neural network to achieve improved computational efficiency. The proposed method can be referred to as the deep learning (DL) algorithm or DL reconstruction for short.

3.1. Reflection Artifacts Correction Model

Assume that $\boldsymbol{I}_{\mathrm{IS}}$ is the simulated initial source without any artifacts and $\boldsymbol{I}_{\mathrm{RA}}$ is a PAT image including reflection artifacts, the relationship between them can be expressed as follows:

$$\boldsymbol{I}_{\mathrm{RA}} = \mathcal{F}(\boldsymbol{I}_{\mathrm{IS}}), \tag{2}$$

where function $\mathcal{F}$ denotes the acoustic-reflection-induced process due to the hyperechoic structures or the special setting of making measurement such as in reverberant-field PAT. The reflection artifacts correction model is to seek an approximate inverse, $\mathcal{G} \approx \mathcal{F}^{-1}$, to reduce the reflection artifacts from $\boldsymbol{I}_{\mathrm{RA}}$, i.e.,

$$\boldsymbol{I}_{\mathrm{est}} = \mathcal{G}(\boldsymbol{I}_{\mathrm{RA}}) \approx \boldsymbol{I}_{\mathrm{IS}}. \tag{3}$$

Next, we introduce the proposed network to learn the approximate inverse function $\mathcal{G}$.

3.2. The Proposed Network

The proposed network architecture for reflection artifacts correction is shown in Figure 2, which has following three parts:

1. **Feature extraction** is to extract the feature representation from the input image $\boldsymbol{I}_{\mathrm{RA}}$. This part contains 4 convolutional layers, each of them has 32 convolutional filters of size 3 × 3 with a stride of 1. Zero padding is not used for these four layers.
2. **Artifacts reduction** is to correct the artifacts and remove the noise from the feature-maps obtained above. We use a modified version of U-net coupling with a residual skip connection for this purpose. This part has 6 convolutional layers and 6 deconvolutional layers. Instead of down-sampling or up-sampling operation in original U-net [40], we use the convolution

with a stride of 2 to decrease the size of feature-maps at the 2nd, 4th and 5th convolutional layers, and the deconvolution with a stride of 2 to increase the size of feature-maps at the 1st, 3rd, and 5th deconvolutional layers. The stride of 1 is used at remaining layers. All layers have 32 (de)convolutional filters of size 5×5 to capture the global structures. In addition to this, three conveying links [29,40] copy the former feature-maps and reuse them as the input to a later layer that has the same size of feature-maps via a concatenation operation along channel dimension, which preserve high-resolution features. At last, a residual skip connection [41] enables the above U-net to infer the artifacts and noise from the input feature-maps. The summation between the input feature-maps and the outputs of the U-net are the corrected feature-maps, which is followed by an activation function and then serves as the input to the reconstruction part. Note that zero padding is used throughout this part.

3. **Reconstruction** is to recover the final output from the corrected feature-maps. This part has 4 deconvolutional layers, each of them has 32 filters of size 3×3 with a stride of 1 except for the last layer that has only 1 filter. Zero padding is not used for these four layers.

The rectified linear unit (ReLU) is used throughout this network [42], which is defined as $f(x) = x^+ = \max(0, x)$. The input to the network $\boldsymbol{I}_{\mathrm{RA}}$ is the output of the first iteration of averaged time reversal (ATR), which is denoted as ATR*.

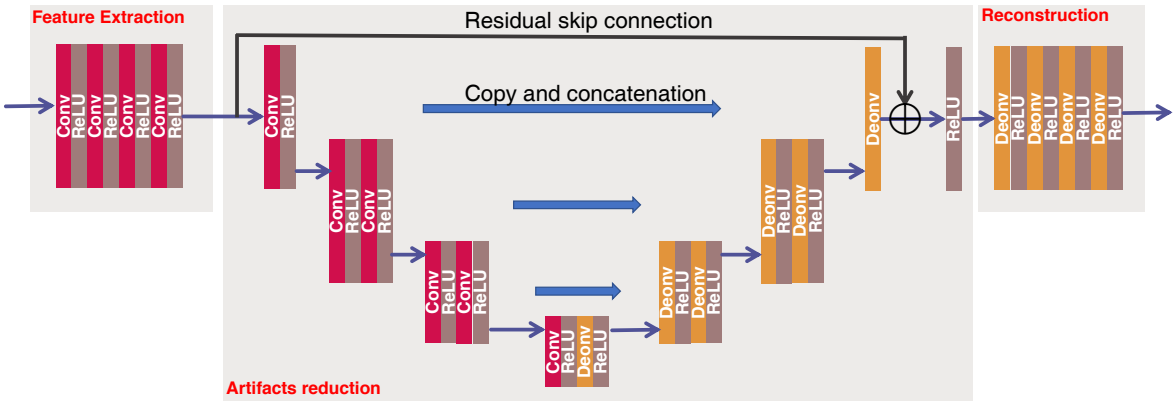


Figure 2. The proposed network structure for reflection artifacts correction. It comprises three parts—feature extraction, artifacts reduction, and reconstruction. In particular, we use a modified version of U-net as the backbone of artifacts reduction part.

3.3. Loss Function

The parameters of the network need to be optimized by minimizing an appropriate loss function. The loss function we choose is the combination of the mean-squared error (MSE) and structural similarity (SSIM) [43].

The MSE between the output of the network, $\boldsymbol{I}_{\mathrm{est}}$, and the initial reference source, $\boldsymbol{I}_{\mathrm{IS}}$, is formally defined as

$$\mathrm{MSE}(\boldsymbol{I}_{\mathrm{est}}, \boldsymbol{I}_{\mathrm{IS}}) = \frac{1}{w \times h} \left\| \boldsymbol{I}_{\mathrm{est}} - \boldsymbol{I}_{\mathrm{IS}} \right\|_F^2, \tag{4}$$

where w and h are the width and height of the image, respectively, and $\| \cdot \|$ refers to Frobenius norm. Although MSE is the most straightforward loss function to optimize the network, the resultant images are usually over-smoothing and lose some details.

In contrast to MSE, the SSIM can measure the similarity between two images in terms of their structures and textures. SSIM index is calculated on various windows of an image. The measure between the window x over $\boldsymbol{I}_{\mathrm{est}}$ and the window y over $\boldsymbol{I}_{\mathrm{IS}}$, based on a common size $k \times k$, is defined as

$$\mathrm{SSIM}(x, y) = \frac{2\mu_x\mu_y + c_1}{\mu_x^2 + \mu_y^2 + c_1} \frac{2\sigma_{xy} + c_2}{\sigma_x^2 + \sigma_y^2 + c_2}, \tag{5}$$

where μ_x and μ_y are the averages of x and y respectively, σ_x^2 and σ_y^2 are the variances of x and y respectively, and σ_{xy} is the covariance of x and y. Also, $c_1 = 0.01^2$ and $c_2 = 0.03^2$ are two constants, which are used to stabilize the division with a weak denominator. The windows size k is 11, as suggested. The SSIM between two images $\boldsymbol{I}_{\text{est}}$ and $\boldsymbol{I}_{\text{IS}}$, $\text{SSIM}(\boldsymbol{I}_{\text{est}}, \boldsymbol{I}_{\text{IS}})$, refers to the average of the SSIM index over all windows.

The loss function is defined as

$$\min_{\boldsymbol{\theta}_{\mathcal{G}}} \mathbb{E}_{(\boldsymbol{I}_{\text{est}}, \boldsymbol{I}_{\text{IS}})} \sqrt{1 + \text{MSE}(\boldsymbol{I}_{\text{est}}, \boldsymbol{I}_{\text{IS}})} \times \left[1 - \text{SSIM}(\boldsymbol{I}_{\text{est}}, \boldsymbol{I}_{\text{IS}})\right], \tag{6}$$

where $\boldsymbol{\theta}_{\mathcal{G}}$ denotes the parameters of the network $\mathcal{G}$. This loss function not only reduces the noise in terms of MSE but also preserve the image structures as measured by SSIM [44]. Various algorithms can solve this minimization problem. In this work, we adopt the Adam algorithm to update the parameters [45]. The gradients of the parameters are computed using a back-propagation algorithm [46].

4. Results

4.1. Experimental Setup

4.1.1. Simulated Dataset

We used 5 cadaver CT scans from Massachusetts General Hospital [47] for simulating PAT dataset. These 5 cadavers were scanned on a GE Discovery 750 HD scanner under 120 kVp x-ray spectra. Noise index (NI) was used by GE to define the image quality, which is approximately equal to the standard deviation of the CT number in the central region of the image of a uniform phantom [48]. In this study, we used a noise index of 10, which represents a normal-dose scanning. Then, the images were reconstructed by the GE commercial iterative reconstruction algorithm, called adaptive statistical iterative reconstruction (ASIR-50%). To reduce the computational cost for simulating the data and training the network, we extracted image patches of size 128×128 from CT scans. More specifically, 64,000 image patches were randomly selected from 3 scans for training purpose. Then, 6400 image patches were randomly selected from 1 scan for validation. Finally, 200 image patches were randomly selected from 1 scan for testing the trained model. Please note that the patients for training, validation, and testing sets were randomly selected from 5 patients CT scans without replacement. Since the Hounsfield unit (HU) used in CT imaging ranges from -1000 to $\sim +2000$, we are interested in the complex tissue structure within $[-160, 240]$ HU window for our PAT simulation. Thus, with this selected HU window, image patches were first normalized into $[0, 1]$ serving as the initial source for PAT imaging, and the output of the first iteration of ATR serves as the input to the network.

4.1.2. Baseline Method

We compare the proposed network-based reconstruction with two state-of-the-art iterative algorithms in correcting reflection artifacts. These iterative algorithms, known as the ATR and adjoint method respectively, are designed to remove the reflection artifacts from the mathematical point of view. The reasons for choosing these two methods are three-fold. Firstly, in contrast to the universal back-projection formula [5] which is mathematically valid only in homogeneous media, these methods can be applied to general heterogeneous media. The universal back-project formula is a special case of the TR method if one sets the sound speed to be constant and applies the Kirchhoff solution formula for the 3D acoustic wave equation. Secondly, the TR and adjoint methods are applicable with arbitrary closed surfaces of sensor arrays. In a typical TR or adjoint process, the measured ultrasound signals are re-transmitted in a temporally reversed order back to the tissue. This can be numerically simulated by solving the wave propagation model backwards to the initial moment while using the measured ultrasound signal as the boundary condition, regardless of the geometry of sensor arrays. Thirdly, TR and adjoint methods are popular in both research and application, for instance [8,9,49,50] for

various TR methods and [51–53] for various adjoint methods. Their implementation has also been included in open source packages such as the k-wave MATLAB package [54]. A sketchy introduction to these iterative algorithms is included in the Appendix A for the convenience of the readers.

4.1.3. Parameter Setting

For our proposed network, the initial learning rate λ was set as 1.0×10^{-3} and was adjusted by every epoch, namely $\lambda_t = \frac{\lambda}{\sqrt{t}}$ at the t-th epoch. For the Adam optimization, the coefficients used for computing running averages of gradients and its square were set as 0.9 and 0.999, respectively. The network was implemented with PyTorch DL library [55] and trained within 60 epochs using four NVIDIA GeForce GTX 1080 Ti GPUs. The batch size was set as 512 during the training.

For these two iterative algorithms, the number of iterations for ATR and Landweber is empirically set to be 10. The regularization parameter in Landweber is empirically chosen as 0.03.

4.1.4. Evaluation Metrics

For the evaluation of image quality, we used the peak signal-to-noise ratio (PSNR) and SSIM in our experiments. The SSIM has been defined in (5) and PSNR is defined as follows:

$$\mathrm{PSNR}(\boldsymbol{I}_{\mathrm{est}}, \boldsymbol{I}_{\mathrm{IS}}) = 10 \log_{10} \Big(\frac{R^2}{\mathrm{MSE}(\boldsymbol{I}_{\mathrm{est}}, \boldsymbol{I}_{\mathrm{IS}})} \Big), \tag{7}$$

where R is the maximum possible pixel value of the images, which is 1 in this study as the images are normalized into $[0, 1]$.

4.2. Results

In this part, we demonstrate the performance of the proposed method in correcting reflection artifacts.

4.2.1. Homogeneous Media

Soft biological tissue is made up mostly of water and therefore can be viewed approximately as a homogeneous medium. The sound speed in water is 1480 m/s, which can be taken as the speed in tissue. The specific value of the constant is irrelevant to the performance of the algorithms, as the value can always be adjusted by choosing different units. What matters is that the speed must be uniform everywhere. Therefore, we simply take the speed as 1 in our numerical test, i.e., we set $c(r) = 1$. Ultrasound signals are generated using the forward model (1). The signals are then exploited in TR, ATR, Landweber iteration, and DL algorithms, respectively, to reconstruct the original image. The input of the neural network is the output of the first iteration of ATR, which is denoted as ATR*.

Based on an independent testing set of 200 CT image patches, we investigated these reconstruction algorithms with 0%, 10%, 20%, 30%, 40% noise added to the ultrasound signals respectively, and reported the box plots of the image quality that was evaluated by PSNR and SSIM in reference to the ground-truth simulated initial source in Figure 3. The experimental results show the DL algorithm has at least two advantages over the others. On the one hand, it is relatively more stable as the noise level increases. Its PSNR outperforms the others at all noise levels, with the only exception of the ideal zero-noise case where ATR provides the best reconstruction. On the other hand, the DL algorithm is time-saving. Although training the neural network takes large amount of time, its output is almost immediate once the training process is accomplished. In contrast, the ATR and Landweber iteration takes several minutes to achieve an image of satisfactory quality.

A few reconstructed images randomly selected from the testing set are displayed in Figure 4. Each column corresponds to the reconstruction using the algorithm labeled at the bottom. The ATR* column consists of images from the first iteration of ATR, which are the inputs to the neural network. The IS column consists of the ground-truth initial source images used to generate the ultrasound signals by the forward model. Among all the algorithms, TR gives the worst result, which can be

expected as it does not resolve reflected artifacts. ATR tends to lose detailed information at a high noise level (see the last row). Landweber performs better in resolving noise as it has regularization effect, which helps remove the high-frequency content of the image. The DL algorithm exhibits superior reconstruction in general.

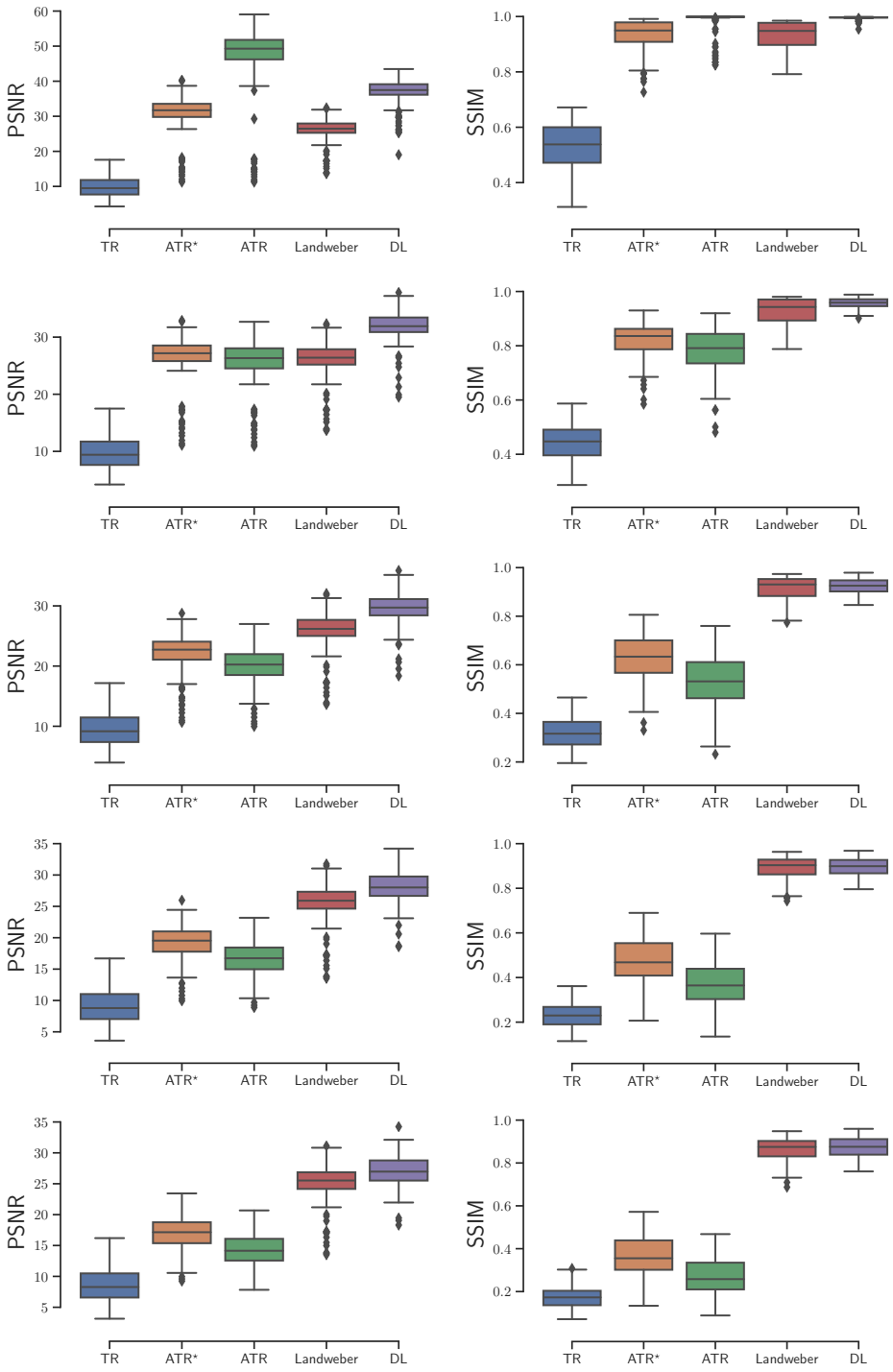


Figure 3. The box plots of the image quality evaluated by PSNR and SSIM on the testing set in a homogeneous medium. From top row to bottom row: noise increases from 0 to 0.4. Diamonds (⋄) indicate outliers.

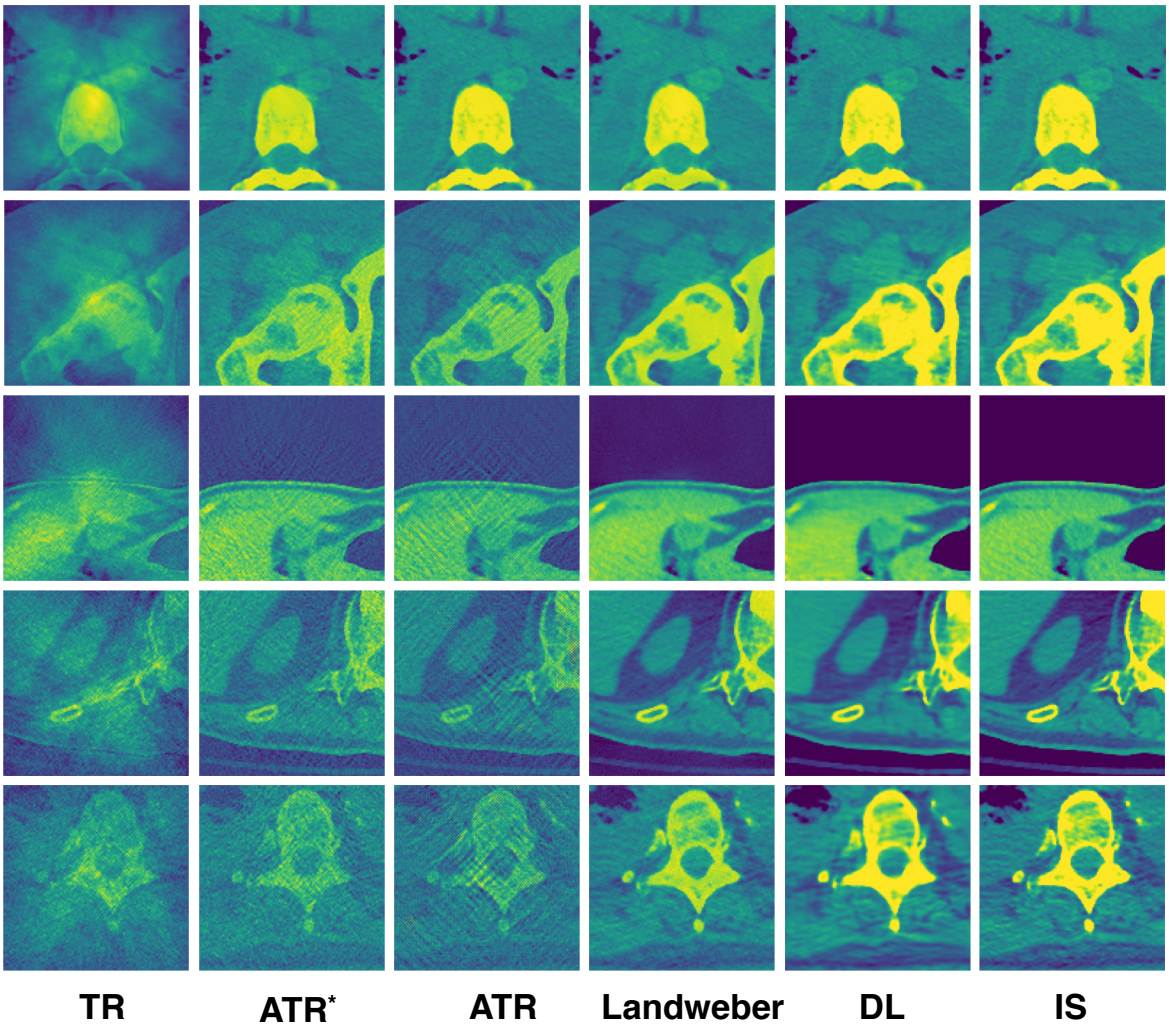


Figure 4. Reconstruction in a homogeneous medium.

4.2.2. Heterogeneous Media

Heterogeneity occurs when the object to be imaged has a complex composition. An example is the transcranial PAT where the sound speed in the skull is 3200 m/s in contrast to 1480 m/s in the water. The speed has jump singularities at the interface of the two constituents. Such singularities can be mingled with those from the initial pressure and appear in the reconstructed image as additional artifacts, casting more challenges for the reconstruction of the initial pressure. We implemented the algorithms in the domain $[-1,1]\times[-1,1]$ and chose the distribution of the sound speed as

$$c = 1 - 0.2\sin(2\pi x) + 0.15\cos(\pi y) + \chi_{(x-0.5)^2+(y-0.5)^2<0.01}, \tag{8}$$

where $\chi_{(x-0.5)^2+(y-0.5)^2<0.01}$ is a function that equals 1 on the disk $\{(x,y) : (x-0.5)^2+(y-0.5)^2 < 0.01\}$ and 0 otherwise. The reason for such choice of c is as follows. The constant 1 models the speed in soft tissue. The smooth term $-0.2\sin(2\pi x) + 0.15\cos(\pi y)$ is added to mimic the slight variation of the sound speed in distinct types of tissue. The non-smooth term $\chi_{(x-0.5)^2+(y-0.5)^2<0.01}$ captures the jumps between different materials such as soft tissue and bones; see Figure 5 for the distribution of the sound speed c.

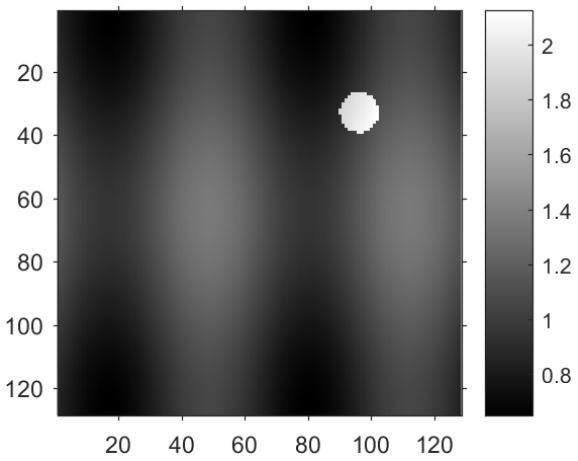


Figure 5. Distribution of the sound speed.

With the variable sound speed, we computed the PSNR and SSIM of the reconstructed images with 0%, 10%, 20%, 30%, 40% noise added to the ultrasound signals respectively, as is shown in Figure 6. The DL algorithm still demonstrates the optimal overall performance. Besides being more stable and time-saving, the DL algorithm also yields the least outliers. Here an outlier is a number in the dataset that is less than $Q_1 - 1.5 \times (Q_3 - Q_1)$ or greater than $Q_3 + 1.5 \times (Q_3 - Q_1)$, where Q_1 is the lower quartile, and Q_3 is the upper quartile.

Some reconstructed images are randomly selected from the testing set again to illustrate the difference of the algorithms; see Figure 7. ATR still suffers from the high noise level, but can resolve the jumps in the speed. Landweber iteration, however, introduces additional artifacts on the top right of the reconstructed image where the sound speed jumps as shown in Figure 5. This is partly due to the limited number of iterations, and it is observed that artifact becomes weaker if the number of iterations is increased. The DL reconstruction resolves simultaneously the high noise and jumps in the speed. It remains visually the closest to the true initial source.

5. Discussions

Our experimental results empirically demonstrated that DL reconstruction in certain situations is superior to the conventional iterative reconstructions, especially when it deals with signals compromised by strong noise. The test on the sound reflectors suggests a great potential of DL-based methods for removing reflection artifacts in the PAT image formation.

However, there are some limitations in this study. First, some of the parameters in these iterative algorithms, such as the number of iterations and the value of the regularization parameter, are not specifically optimized. Varying these parameters may result in somewhat improved performance of the iterative algorithms. However, finding the optimal values of such parameters are highly empirical, and there is no universal approach in general. Second, we empirically used the combination of MSE and SSIM as the loss function to optimize the network, and evaluated the image quality by PSNR and SSIM. The choice of the loss function and image metrics may not be the optimal to capture the visual quality for PAT imaging. Third, we only studied the reflection artifacts under different noise levels and a simple variable sound model, other more complicated conditions such as limited views and more complicated variable sounds can be surely addressed by extending the proposed network.

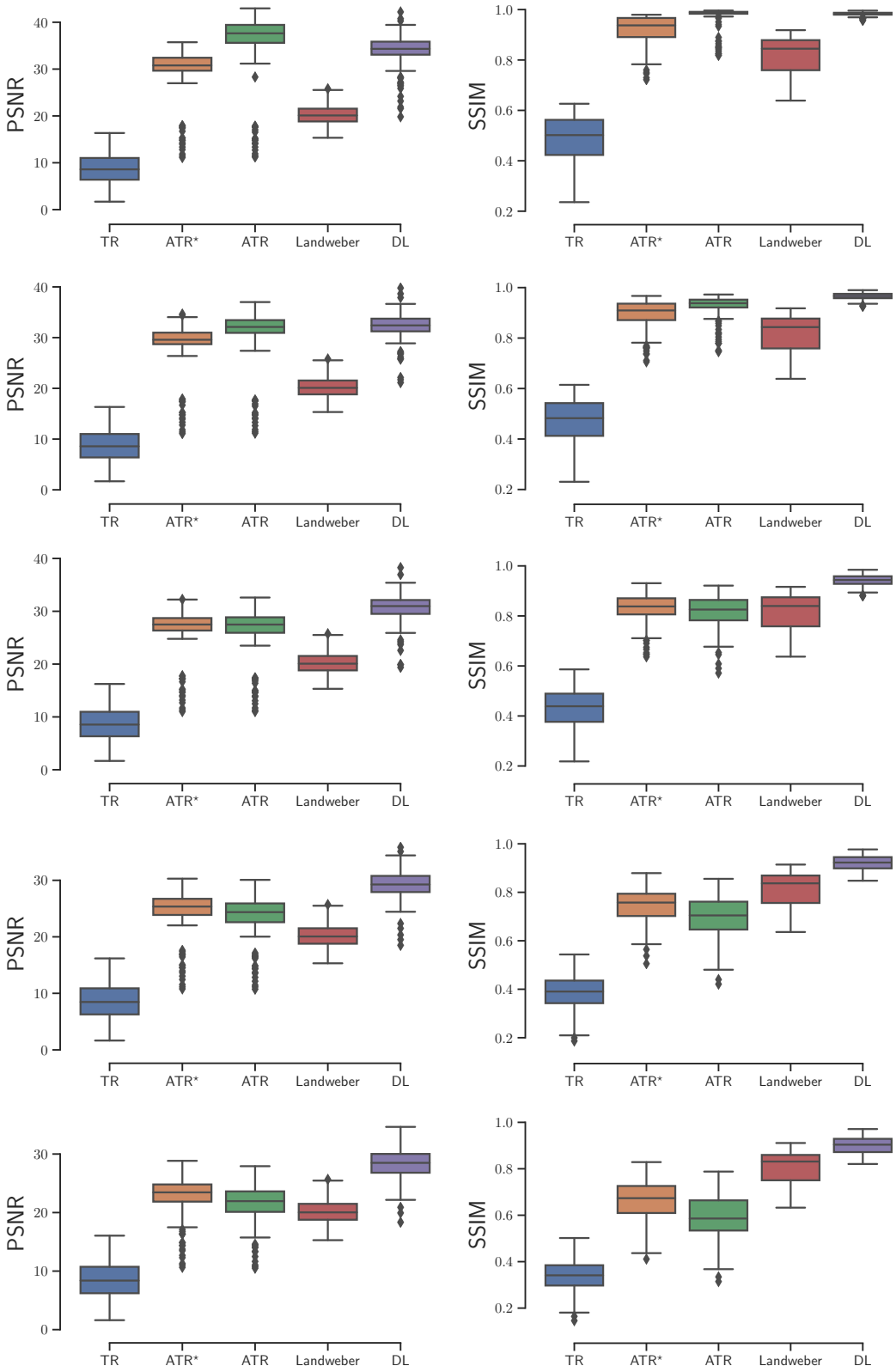


Figure 6. The box plots of the image quality evaluated by PSNR and SSIM on the testing set in a heterogeneous medium. From top row to bottom row: noise increases from 0 to 0.4. Diamonds (⋄) indicate outliers.

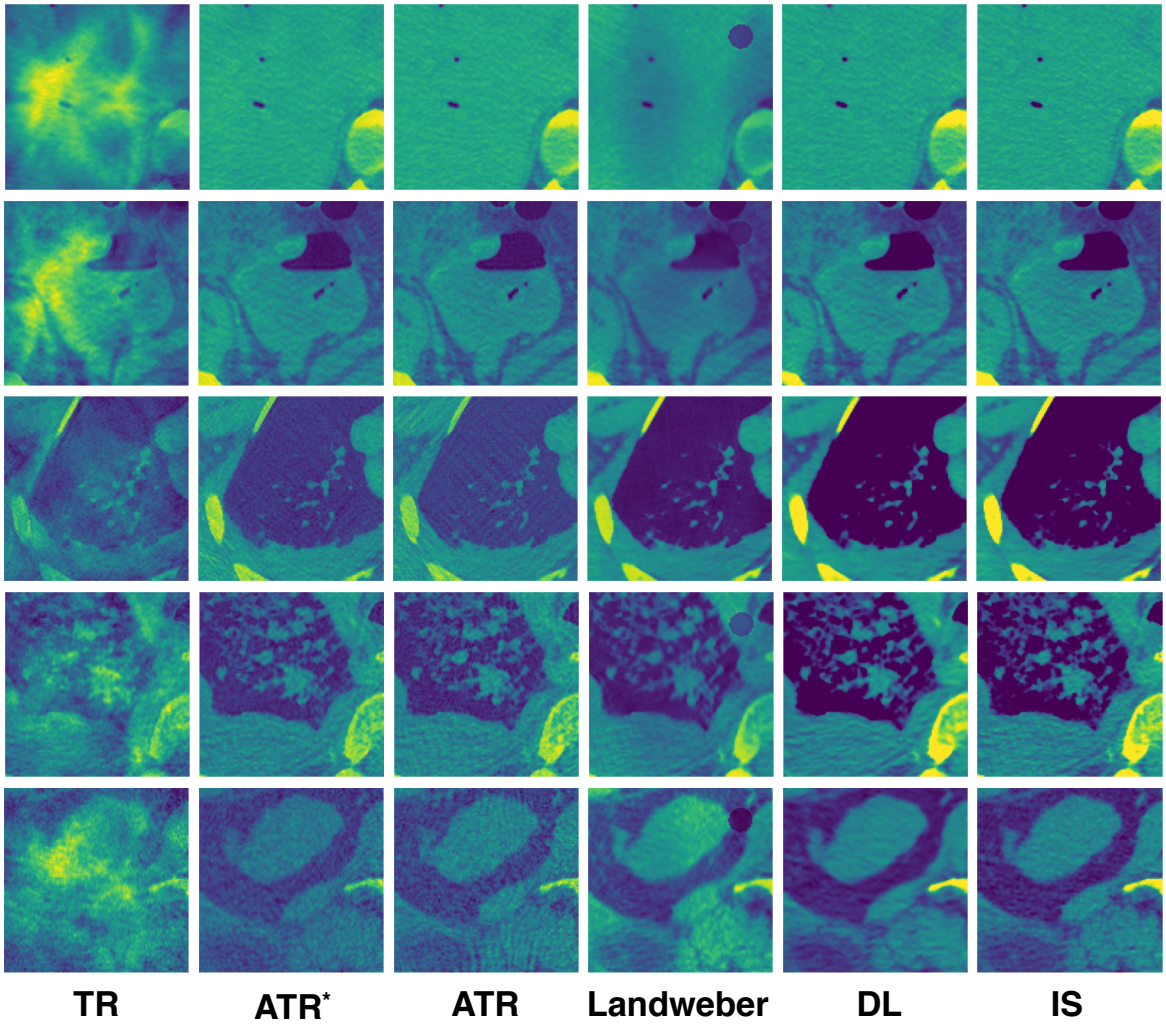


Figure 7. Reconstruction in a heterogeneous Medium.

6. Conclusions

In this article, we have proposed a novel DNN to remove the reflection artifacts in reconstructed PAT images under different noise levels and different media. By directly comparing the proposed network to popular iterative reconstruction algorithms with simulated PAT data from CT scans, the results have showed that the proposed network is able to reconstruct superior images over the conventional iterative reconstructions in typical scenarios in terms of computational efficiency and noise reduction.

The results can be further strengthened in several aspects. One practical and significant question is how to make the network robust to potential malignant attacks. It is well known that DL models are vulnerable to adversarial examples. A more stable training procedure is thus critical to the clinical application of DL methods. Next, the effectiveness and efficiency of the network can be further improved for constrained resources and cloud-end processing. Some other factors, such as limited view, acoustic attenuation, and fluctuation of sound speed, can greatly impact the quality of PAT images. It would be interesting to extend the DL approach to these situations as well.

Author Contributions: H.S. and Y.Y. initiated the project and designed the experiments. H.S. performed machine learning research. Y.Y. performed iterative reconstruction research. H.S. and Y.Y. wrote the paper, and G.W. participated in the discussions and edited the paper.

Funding: Y.Y. was partly supported by the NSF grant DMS-1715178, the Simons travel grant, and the startup fund from the Michigan State University.

Acknowledgments: The authors thank NVIDIA Corporation for the donation of GPUs used for this research. The authors would also like to express their gratitude to the anonymous reviewers for the valuable suggestions and comments which helped considerably improve the exposition of the paper.

Conflicts of Interest: H.S. and G.W. have received unrelated industrial research grants from General Electric and Hologic Inc.

Abbreviations

The following abbreviations are used in this manuscript:

PAT	Photo-Acoustic Tomography
TR	Time Reversal
ATR	Averaged Time Reversal
DL	Deep Learning
CT	Computed Tomography
CNN	Convolutional Neural Network
PSNR	Peak Signal-to-Noise Ratio
SSIM	Structural Similarity
FDTD	Finite Difference Time Domain
CFL	Courant–Friedrichs–Lewy
MSE	Mean-Squared Error
ASIR	Adaptive Statistical Iterative Reconstruction
HU	Hounsfield unit

Appendix A. Artifacts Correction by Iterative Algorithms

In this appendix, we describe the principles of the averaged time reversal and the adjoint method in more details. Interested readers are referred to [8,14] for the accurate and complete exhibition.

Appendix A.1. Averaged Time Reversal

The averaged time reversal method is proposed in [12] as a remedy to the conventional TR to remove the reflection artifacts in the latter. The essential improvement comes from the introduction of an averaging process to the measured data before reversing the time. The rational, roughly speaking, is that artifacts generated by acoustic reflections have either positive amplitude or negative amplitude, depending on the number of times they touch the boundary. A suitable averaging process along the time direction is thus able to annihilate reflected artifacts with opposite signs. The process can be viewed as a pre-conditioning to the TR. To be a bit more precise, let $\mathbf{\Lambda}$ be the linear operator

$$\mathbf{\Lambda}: \qquad p_0 \mapsto u|_{[0,T]\times\partial\Omega} \tag{A1}$$

where u is the solution of the forward model (1). In other words, $\mathbf{\Lambda}$ sends the initial pressure to the measured data. This is a linear operator that is completed determined by the form of the partial differential equation in (1). If the forward model is discretized, $\mathbf{\Lambda}$ would just be a matrix. Reconstructing the initial pressure from the ultrasound signals amounts to inverting the matrix $\mathbf{\Lambda}$. Direct inversion is normally impossible due to the large dimensionality of this matrix. Instead, it is shown in [12] that one can introduce an averaged time reversal process $\boldsymbol{A}$ such that the initial pressure p_0 can be reconstructed iteratively from the measured data $u|_{[0,T]\times\partial\Omega}$ by the relation

$$p_0 = \sum_{k=0}^{\infty} (\mathbf{A}\mathbf{\Lambda} - \boldsymbol{Id})^k \mathbf{A}(u|_{[0,T]\times\partial\Omega}) \tag{A2}$$

where $\boldsymbol{Id}$ is the identity operator (or identity matrix if the forward model is discretized). This algorithm is known as the averaged time reversal (ATR). It is mathematically proved in [12] that ATR can correct the reflection artifacts caused by conventional time reversal methods. It is one of the baseline methods we used in the paper to compare with the DL reconstruction.

Appendix A.2. Landweber Iteration

The Landweber iteration, also known as the Landweber algorithm, is an algorithm proposed in the 1950s by Landweber [56] to solve ill-posed linear systems of the form $\mathbf{\Lambda}\boldsymbol{x} = \boldsymbol{b}$ where $\mathbf{\Lambda}$ is a (not necessarily square) matrix. It is a regularization method that can be viewed as iteratively solving the unconstrained optimization problem

$$\min_{\boldsymbol{x}} \frac{1}{2}\|\mathbf{\Lambda}\boldsymbol{x} - \boldsymbol{b}\|_2^2, \tag{A3}$$

which leads to the update scheme

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - \gamma\mathbf{\Lambda}^*(\mathbf{\Lambda}\boldsymbol{x}_k - \boldsymbol{b}), \qquad k = 0, 1, 2, \ldots \tag{A4}$$

where $\mathbf{\Lambda}^*$ is the adjoint operator of $\mathbf{\Lambda}$ and γ is a regularization parameter. It is well known the Landweber iteration is convergent if $0 < \gamma < \frac{2}{\sigma_1^2}$ where σ_1 is the largest singular value of $\mathbf{\Lambda}$, and it converges to the projection of the true solution on the orthogonal complement of the null space of $\mathbf{\Lambda}$, see the analysis in [14] for instance. In our case, the vector $\boldsymbol{x}$ is the discretized version of p_0, and the vector $\boldsymbol{b}$ is the discretized version of the measured data $u|_{[0,T]\times\partial\Omega}$. The iteration exhibits the effect of semi-convergence: it converges before reaching a certain number of iterations but then diverges once beyond. The optimal value of γ and the stopping rule are largely empirical. Some guiding principles exist, but we do not intend to discuss them in this article.

References

1. Xia, J.; Yao, J.; Wang, L.V. Photoacoustic tomography: Principles and advances. *Electromagn. Waves* **2014**, *147*, 1. [CrossRef]
2. Wang, L.V.; Beare, G.K. Breaking the Optical Diffusion Limit: Photoacoustic Tomography. In *Frontiers in Optics*; Optical Society of America: Rochester, NY, USA, 2010; p. FWY2.
3. Yao, J.; Wang, L.V. Photoacoustic tomography: Fundamentals, advances and prospects. *Contrast Media Mol. Imag.* **2011**, *6*, 332–345. [CrossRef] [PubMed]
4. Cox, B.; Beard, P. Photoacoustic tomography with a single detector in a reverberant cavity. *J. Acoust. Soc. Am.* **2009**, *125*, 1426–1436. [CrossRef] [PubMed]
5. Xu, M.; Wang, L.V. Universal back-projection algorithm for photoacoustic computed tomography. *Phys. Rev. E* **2005**, *71*, 016706. [CrossRef] [PubMed]
6. Hristova, Y.; Kuchment, P.; Nguyen, L. Reconstruction and time reversal in thermoacoustic tomography in acoustically homogeneous and inhomogeneous media. *Inverse Probl.* **2008**, *24*, 055006. [CrossRef]
7. Hristova, Y. Time reversal in thermoacoustic tomography—An error estimate. *Inverse Probl.* **2009**, *25*, 055008. [CrossRef]
8. Stefanov, P.; Uhlmann, G. Thermoacoustic tomography with variable sound speed. *Inverse Probl.* **2009**, *25*, 075011. [CrossRef]
9. Stefanov, P.; Uhlmann, G. Thermoacoustic tomography arising in brain imaging. *Inverse Probl.* **2011**, *27*, 045004. [CrossRef]
10. Stefanov, P.; Yang, Y. Thermo and Photoacoustic Tomography with variable speed and planar detectors. *SIAM J. Math. Anal.* **2017**, *49*, 297–310. [CrossRef]

11. Holman, B.; Kunyansky, L. Gradual time reversal in thermo-and photo-acoustic tomography within a resonant cavity. *Inverse Probl.* **2015**, *31*, 035008. [CrossRef]
12. Stefanov, P.; Yang, Y. Multiwave tomography in a closed domain: Averaged sharp time reversal. *Inverse Probl.* **2015**, *31*, 065007. [CrossRef]
13. Nguyen, L.V.; Kunyansky, L.A. A dissipative time reversal technique for photoacoustic tomography in a cavity. *SIAM J. Imag. Sci.* **2016**, *9*, 748–769. [CrossRef]
14. Stefanov, P.; Yang, Y. Multiwave tomography with reflectors: Landweber's iteration. *Inverse Probl. Imag.* **2017**, *11*, 373–401. [CrossRef]
15. Wang, G. A perspective on deep imaging. *IEEE Access* **2016**, *4*, 8914–8924. [CrossRef]
16. Wang, G.; Ye, J.C.; Mueller, K.; Fessler, J.A. Image reconstruction is a new frontier of machine learning. *IEEE Trans. Med. Imag.* **2018**, *37*, 1289–1296. [CrossRef] [PubMed]
17. Zhu, B.; Liu, J.Z.; Cauley, S.F.; Rosen, B.R.; Rosen, M.S. Image reconstruction by domain-transform manifold learning. *Nature* **2018**, *555*, 487. [CrossRef] [PubMed]
18. Antholzer, S.; Haltmeier, M.; Schwab, J. Deep Learning for Photoacoustic Tomography from Sparse Data. *Inverse Probl. Sci. Eng.* **2019**, *27*, 987–1005. [CrossRef] [PubMed]
19. Antholzer, S.; Haltmeier, M.; Nuster, R.; Schwab, J. Photoacoustic image reconstruction via deep learning. In *Photons Plus Ultrasound: Imaging and Sensing 2018*; International Society for Optics and Photonics: Bellingham, WA, USA, 2018; Volume 10494.
20. Antholzer, S.; Schwab, J.; Haltmeier, M. Deep Learning Versus ℓ^1-Minimization for Compressed Sensing Photoacoustic Tomography. In Proceedings of the IEEE International Ultrasonics Symposium (IUS), Kobe, Japan, 22–25 October 2018; pp. 206–212.
21. Hauptmann, A.; Lucka, F.; Betcke, M.M.; Huynh, N.; Adler, J.; Cox, B.T.; Beard, P.C.; Ourselin, S.; Arridge, S.R. Model-Based Learning for Accelerated, Limited-View 3-D Photoacoustic Tomography. *IEEE Trans. Med. Imag.* **2018**, *37*, 1382–1393. [CrossRef] [PubMed]
22. Waibel, D.; Gröhl, J.; Isensee, F.; Kirchner, T.; Maier-Hein, K.; Maier-Hein, L. Reconstruction of initial pressure from limited view photoacoustic images using deep learning. In *Photons Plus Ultrasound: Imaging and Sensing 2018*; International Society for Optics and Photonics: Bellingham, WA, USA, 2018; Volume 10494.
23. Schwab, J.; Antholzer, S.; Nuster, R.; Paltauf, G.; Haltmeier, M. Deep Learning of truncated singular values for limited view photoacoustic tomography. In *Photons Plus Ultrasound: Imaging and Sensing 2019*; International Society for Optics and Photonics: San Francisco, CA, USA, 2019; Volume 10878, p. 1087836.
24. Guan, S.; Khan, A.; Sikdar, S.; Chitnis, P. Fully Dense UNet for 2D sparse photoacoustic tomography artifact removal. *IEEE J. Biomed. Health Inform.* **2019**.[CrossRef]
25. Allman, D.; Reiter, A.; Bell, M.A.L. Photoacoustic source detection and reflection artifact removal enabled by deep learning. *IEEE Trans. Med. Imag.* **2018**, *37*, 1464–1477. [CrossRef]
26. Allman, D.; Reiter, A.; Bell, M. Exploring the effects of transducer models when training convolutional neural networks to eliminate reflection artifacts in experimental photoacoustic images. In *Photons Plus Ultrasound: Imaging and Sensing 2018*; International Society for Optics and Photonics: Bellingham, WA, USA, 2018; Volume 10494, p. 104945H.
27. Kelly, B.; Matthews, T.P.; Anastasio, M.A. Deep learning-guided image reconstruction from incomplete data. *arXiv* **2017**, arXiv:1709.00584.
28. Schwab, J.; Antholzer, S.; Nuster, R.; Haltmeier, M. Real-time photoacoustic projection imaging using deep learning. *arXiv* **2018**, arXiv:1801.06693.
29. Shan, H.; Zhang, Y.; Yang, Q.; Kruger, U.; Kalra, M.K.; Sun, L.; Cong, W.; Wang, G. 3-D convolutional encoder-decoder network for low-dose CT via transfer learning from a 2-D trained network. *IEEE Trans. Med. Imag.* **2018**, *37*, 1522–1534. [CrossRef] [PubMed]
30. You, C.; Yang, Q.; Shan, H.; Gjesteby, L.; Li, G.; Ju, S.; Zhang, Z.; Zhao, Z.; Zhang, Y.; Cong, W.; et al. Structurally-Sensitive Multi-Scale Deep Neural Network for Low-Dose CT Denoising. *IEEE Access* **2018**, *6*, 41839–41855. [CrossRef] [PubMed]
31. Gjesteby, L.; Yang, Q.; Xi, Y.; Shan, H.; Claus, B.; Jin, Y.; De Man, B.; Wang, G. Deep learning methods for CT image-domain metal artifact reduction. In *Developments in X-ray Tomography XI*; International Society for Optics and Photonics: San Diego, CA, USA, 2017; Volume 10391, p. 103910W.

32. Gjesteby, L.; Shan, H.; Yang, Q.; Xi, Y.; Claus, B.; Jin, Y.; De Man, B.; Wang, G. Deep Neural Network for CT Metal Artifact Reduction with a Perceptual Loss Function. In Proceedings of the 5th International Conference on Image Formation in X-Ray Computed Tomography, Salt Lake City, UT, USA, 20–23 May 2018.
33. You, C.; Li, G.; Zhang, Y.; Zhang, X.; Shan, H.; Ju, S.; Zhao, Z.; Zhang, Z.; Cong, W.; Vannier, M.W.; et al. CT Super-resolution GAN Constrained by the Identical, Residual, and Cycle Learning Ensemble (GAN-CIRCLE). *IEEE Trans. Med. Imag.* **2019**. [CrossRef]
34. Lyu, Q.; You, C.; Shan, H.; Zhang, Y.; Wang, G. Super-resolution MRI and CT through GAN-circle. In *Developments in X-Ray Tomography XI*; International Society for Optics and Photonics: San Diego, CA, USA, 2019.
35. Shan, H.; Padole, A.; Homayounieh, F.; Kruger, U.; Khera, R.D.; Nitiwarangkul, C.; Kalra, M.K.; Wang, G. Competitive performance of a modularized deep neural network compared to commercial algorithms for low-dose CT image reconstruction. *Nat. Mach. Intell.* **2019**, *1*, 269–276. [CrossRef]
36. Kuchment, P.; Kunyansky, L. Mathematics of thermoacoustic tomography. *Eur. J. Appl. Math.* **2008**, *19*, 191–224. [CrossRef]
37. Ammari, H.; Bossy, E.; Jugnon, V.; Kang, H. Mathematical modeling in photoacoustic imaging of small absorbers. *SIAM Rev.* **2010**, *52*, 677–695. [CrossRef]
38. Ammari, H.; Asch, M.; Bustos, L.G.; Jugnon, V.; Kang, H. Transient wave imaging with limited-view data. *SIAM J. Imag. Sci.* **2011**, *4*, 1097–1121. [CrossRef]
39. Acosta, S.; Montalto, C. Multiwave imaging in an enclosure with variable wave speed. *Inverse Probl.* **2015**, *31*, 065009. [CrossRef]
40. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*; Springer: Cham, Munich, Germany, 2015; pp. 234–241.
41. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 770–778.
42. Glorot, X.; Bordes, A.; Bengio, Y. Deep sparse rectifier neural networks. In Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics, Fort Lauderdale, FL, USA, 11–13 April 2011; pp. 315–323.
43. Wang, Z.; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612. [CrossRef] [PubMed]
44. Zhao, H.; Gallo, O.; Frosio, I.; Kautz, J. Loss functions for image restoration with neural networks. *IEEE Trans. Comput. Imag.* **2016**, *3*, 47–57. [CrossRef]
45. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.
46. Rumelhart, D.E.; Hinton, G.E.; Williams, R.J. Learning representations by back-propagating errors. *Cognit. Model.* **1988**, *5*, 1. [CrossRef]
47. Yang, Q.; Kalra, M.K.; Padole, A.; Li, J.; Hilliard, E.; Lai, R.; Wang, G. Big data from CT scanning. *JSM Biomed. Imag.* **2015**, *2*, 1003.
48. McCollough, C.H.; Bruesewitz, M.R.; Kofler Jr, J.M. CT dose reduction and dose management tools: Overview of available options. *Radiographics* **2006**, *26*, 503–512. [CrossRef]
49. Qian, J.; Stefanov, P.; Uhlmann, G.; Zhao, H. An efficient Neumann series-based algorithm for thermoacoustic and photoacoustic tomography with variable sound speed. *SIAM J. Imag. Sci.* **2011**, *4*, 850–883. [CrossRef]
50. Treeby, B.E.; Zhang, E.Z.; Cox, B.T. Photoacoustic tomography in absorbing acoustic media using time reversal. *Inverse Probl.* **2010**, *26*, 115003. [CrossRef]
51. Belhachmi, Z.; Glatz, T.; Scherzer, O. A direct method for photoacoustic tomography with inhomogeneous sound speed. *Inverse Probl.* **2016**, *32*, 045005. [CrossRef]
52. Javaherian, A.; Holman, S. Direct quantitative photoacoustic tomography for realistic acoustic media. *Inverse Probl.* **2019**. [CrossRef]
53. Javaherian, A.; Holman, S. A continuous adjoint for photo-acoustic tomography of the brain. *Inverse Probl.* **2018**, *34*, 085003. [CrossRef]
54. Treeby, B.E.; Cox, B.T. k-Wave: MATLAB toolbox for the simulation and reconstruction of photoacoustic wave-fields. *J. Biomed. Opt.* **2010**, *15*, 021314. [CrossRef] [PubMed]

55. Paszke, A.; Gross, S.; Chintala, S.; Chanan, G.; Yang, E.; DeVito, Z.; Lin, Z.; Desmaison, A.; Antiga, L.; Lerer, A. Automatic differentiation in PyTorch. In Proceedings of the NIPS 2017 Autodiff Workshop, Long Beach, CA, USA, 9 December 2017.
56. Landweber, L. An iteration formula for Fredholm integral equations of the first kind. *Am. J. Math.* **1951**, *73*, 615–624. [CrossRef]

© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).