End-to-end Res-Unet based reconstruction algorithm for photoacoustic imaging

JINCHAO FENG,1,2 JIANGUANG DENG,1,2 ZHE LI,1,2,3 ZHONGHUA SUN,1,2 HUIJING DOU,1 AND KEBIN JIA1,2,4

1Beijing Key Laboratory of Computational Intelligence and Intelligent System, Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China
2Beijing Laboratory of Advanced Information Networks, Beijing 100124, China
3lizhe1023@bjut.edu.cn
4kebinj@bjut.edu.cn

Abstract: Recently, deep neural networks have attracted great attention in photoacoustic imaging (PAI). In PAI, reconstructing the initial pressure distribution from acquired photoacoustic (PA) signals is a typically inverse problem. In this paper, an end-to-end Unet with residual blocks (Res-Unet) is designed and trained to solve the inverse problem in PAI. The performance of the proposed algorithm is explored and analyzed by comparing a recent model-resolution-based regularization algorithm (MRR) with numerical and physical phantom experiments. The improvement obtained in the reconstructed images was more than 95% in Pearson correlation and 39% in peak signal-to-noise ratio in comparison to the MRR. The Res-Unet also achieved superior performance over the state-of-the-art Unet++ architecture by more than 18% in PSNR in simulation experiments.

1. Introduction

Photoacoustic imaging (PAI), which has the merits of high ultrasonic resolution and strong optical contrast, is a non-invasive and fast developing imaging technology [1]. It is widely used in early-stage cancer diagnosis [2], endogenous chromophores sensing [3], small animal imaging [4], and vessel imaging [5]. In PAI, the tissue is first irradiated by a short-pulsed laser beam to induce the photoacoustic effect [6], then the locally absorbed light is converted into heat, which is further converted to a pressure rise via thermo-elastic expansion. The pressure rise propagates in biological tissue as a photoacoustic wave, which can be detected by ultrasonic transducers, producing an electric signal [7]. To obtain the initial pressure distribution from detected signals, image reconstruction algorithms are necessary.

To date, a wide variety of image reconstruction algorithms have been developed for PAI [8–18]. Conventional algorithms, e.g., back projection (BP) [17,19], are prevalent due to their fast speed. The BP based reconstruction algorithms are popular due to their simplicity and ease of implementation [8,9]. However, the BP algorithm usually requires calculating analytical solution, which only exists for regular geometries [8]. Moreover, reconstructed photoacoustic (PA) images may suffer from severe artifacts, resulting in distorted images. Time reversal (TR) reconstruction algorithms can provide improved quantitative accuracy by modeling ultrasound wave propagation backward in time [10,11]. The TR method requires large amount of data for accurate reconstruction and needs the relatively large computational time for calculation [17,20]. Recently, promising results have been reported using model-based reconstruction algorithms [12–15], even with limited data [16–18]. Model-based reconstruction algorithms seek to minimize the difference between measured signal and predicted signal via the photoacoustic propagation model. Their performances critically depend on a special optimization algorithm. In addition, the regularization parameter may impact the quality of image reconstructed to a large extent [21].
The model-based reconstruction algorithms are categorized into numerical algorithms and require repeated calculation of the forward and adjoint operators during reconstruction. Therefore, they are time-consuming. Moreover, discretized errors are inevitable for solving wave equation governing photoacoustic propagation. Therefore, quantitative accuracy and spatial resolution of reconstructed images are limited. It is therefore substantial for PAI to improve the accuracy and spatial resolution of image reconstruction.

In the last few years, deep learning (DL) especially through convolutional neural networks (CNNs) has been proven to work well in pattern recognition [22], image classification [23–26] and segmentation [27–29]. Some of the recently proposed image reconstruction algorithms in PAI employed the deep learning techniques and have obtained promising results [30–38]. For example, a fully dense Unet was applied to post-process the artifacts in reconstructed images from the TR algorithm [32]. Similarly, a deep learning approach for PAI was recently proposed by Antholzer et al., [33], that uses a filtered back projection (FBP) algorithm as linear preprocessing step followed by a Unet for removing artifacts. Actually, these deep learning-based algorithms are a post-processing approach, and the input of the networks is poor-quality PA images recovered by conventional algorithms. Hence, the performances of these methods are highly dependent on the initial results reconstructed by conventional algorithms.

There are some attempts to directly reconstruct PAI images using a convolutional deep neural network based on Unet-like architectures [39–43]. For example, a double U-net (DU-net) [40] and a novel encoder-decoder based CNN [41] were proposed to directly recover the PA images from raw photoacoustic sensor data. Dominik et al., have used a fully CNN network with a Unet-like architecture to directly reconstruct the initial pressure distribution from the in silico data [42]. The merit of these algorithms is without the need for solving the equations governing PA wave propagation.

The DL-based methods make full advantage of a special deep neural network (DNN) architecture for constructing the mapping from input raw signals to output PA images together with optimization procedures for adjusting the weights of the network during the training phase [30]. DNN is composed of many layers that filter the input raw data to obtain useful information. Moreover, it usually employs multichannel filters, so that there are multiple feature maps at each layer. This increases the expressive power of the network. Also, a large number of stacked layers can enrich the level of features [30]. This suggests an alternative approach for solving PAI problem using DL. However, training a deep learning network is challenging because too many layers will introduce the problem of vanishing (or exploding) gradients [39]. Additionally, degradation of training accuracy appears when the network depth keeps increasing. Furthermore, few studies have thoroughly investigated and analyzed the effects of network architecture for training direct reconstruction algorithms.

In the manuscript, we follow direct reconstruction approach and develop an end-to-end reconstruction algorithm for PAI by modifying the original Unet termed Res-Unet. The Res-Unet incorporates residual blocks into the contracting and expanding paths of the Unet architecture to prevent the deep network from degradation and to improve image quality. Furthermore, a skip connection is added between input sinogram data and output initial pressure images. The end-to-end reconstruction algorithm can reconstruct directly PA images from raw measured data and does not depend on any initial reconstruction such as back projection. The performance of the proposed algorithm is compared with the recently developed model-resolution-based regularization algorithm (MRR) [44], and reconstruction algorithms with different network structures, including Unet [28], inception block [25], dense block [26], recurrent residual block [27], and Unet++ [29].

The overview of this paper is arranged as follows. Section 2 briefly introduces the background, and describe the end-to-end Res-Unet architecture used in PAI reconstruction. Furthermore, data generation and evaluation metrics are described at the end of the section. In Section 3, the
Simulation and physical phantom results are shown compared with a model-based algorithm and other DL-based methods. Finally, we discuss some details and conclude this work followed by future work.

2. Methods

2.1. Background

In PAI, the PA wave is excited by a short-pulse laser light. In the case of acoustically homogeneous medium, under the condition of thermal confinement, the equation governing PA wave propagation can be described as [45]:

\[ \frac{1}{v_s^2} \frac{\partial^2 p(r, t)}{\partial t^2} - \nabla^2 p(r, t) = \Gamma \frac{\partial H(r, t)}{\partial t} \] (1)

where \( v_s \) is the sound velocity, \( r \) is the position, \( t \) is the time, \( p \) is the induced pressure, \( \Gamma \) is the Grüneisen coefficient, and \( H \) is the absorbed energy. In general, both \( p \) and \( H \) depend on \( r \) and \( t \).

Equation (1) can be solved with the two initial conditions [46]:

\[ p|_{t=0} = p_0(r), \quad \frac{\partial p}{\partial t}|_{t=0} = 0 \] (2)

where \( p_0(r) \) is the initial pressure distribution given by \( p_0(r) = \Gamma H(r) \).

The aim of image reconstruction in PAI is to recover initial pressure distribution \( p_0 \) from the acquired PA raw signal \( b \) coupled with Eqs. (1) and (2). In this study, we try to reconstruct initial pressure distribution based on a trained end-to-end Unet instead of solving Eqs. (1) and (2). Based on the deep learning framework, the image reconstruction in PAI can be approximated over a supervised training dataset \( \{ b, p_0 \} = \{ b_i, p_{0i} \}_{i=1}^Q \) as:

\[ \arg \min_{\Theta} \sum_{b,p_0} ||N(\Theta, b) - p_0|| \] (3)

where \( N(\cdot) \) is a neural network with the Unet-like architecture, \( \Theta \) is the parameters of the network and \( Q \) is the number of training datasets. Therefore, the reconstruction algorithm based on deep learning in PAI is a data-driven algorithm without modeling photoacoustic propagation. Although here we focus on the Unet-like architecture, it can be replaced by other types of neural network, such as fully convolutional neural networks [47].

2.2. Proposed Res-Unet architecture

By modifying the original Unet architecture [28], we proposed a neural network architecture termed Res-Unet to directly reconstruction PA images from raw measured data without needing any initial reconstruction. Figure 1 shows the detailed architecture of proposed Res-Unet. The Res-Unet incorporates residual blocks [24] into the contracting and symmetric expanding paths of the Unet architecture. The addition of residual blocks mitigates the gradient vanishing problem when the network goes deeper. In addition, this strategy greatly deepens the neural network and further improves its performance [24]. In Res-Unet, comprehensive context information is extracted from input sinogram data and initial pressure distribution is inferred from a symmetric expanding path. In addition, as a refinement of the original U-net, a skip connection is added between input sinogram data and output initial pressure distribution, which means that the network actually learns the difference between input and output.

In the Res-Unet, we define a basic convolution operator (BasicConv) as a composite function of three consecutive operations: a 3×3 convolution (Conv) followed by a batch normalization (BN) [48] and a rectified linear unit (ReLU). Each contracting unit consists of a BasicConv and a residual block. The spatial dimensions of the feature maps are repeatedly reduced via a 2×2
max pooling operator. This strategy enables the Res-Unet to efficiently learn local and global features over different spatial scales. In the expanding path, each expanding unit consists of a 2×2 up-sampling operators followed by a BasicConv and a residual block. Since the size of the input PA raw sensor data is 512×128 and the size of the true image is 128×128, therefore, a layer in each skip connection is introduced to reduce the image dimensions via an anisotropic kernel size of 6 × 3 and striding of 4 × 1.

Figure 2 shows the details of the residual block [24]. The residual block is divided into two branches: trunk branch and mask branch. The trunk branch is composed of three 3×3 convolutions, two BNs, and two ReLus. In the mask branch, a skip connection is added between input and output to perform identity mapping. The sum of trunk branch and mask branch is followed with a BN and a ReLu. The tensor size remains to be constant in the residual block. All residual blocks in the Res-Unet have the same number of convolutional layers to maintain computational efficiency.

2.3. Synthetic data for training and testing
We used a dataset for training and validation, and a second dataset for testing. Datasets are synthetized using an open source software, k-WAVE MATLAB toolbox [49]. For each dataset generated, the experimental settings are shown in Fig. 3(a). The imaging area is discretized into 100×100 pixels with the pixel size of 200 µm. In the following experiments, the medium is assumed to be homogenous with no absorption or dispersion of sound. And the acoustic speed is set as 1500 m/s. The sensor setting is the same as a commercially available PAI system for
small animal imaging (MSOT inVision 128, iThera Medical GmbH, Munich, Germany). 128 ultrasound transducers are placed with equally spaced on a circle with a radius of 40 mm. And the 128 transducers cover an angle of 270° around the imaging object. The center frequency of each transducer is set as 5 MHz with 70% bandwidth. Each transducer records 512 points with 10 MHz sampling frequency. 30 dB white Gaussian noise is added to the synthetized data.

**Fig. 3.** (a) The experimental setup for data acquisition with shaded square region indicating the imaging area; (b) the physical phantom; (c) the reconstructed slice.

In order to provide the ResU-net’s generalization ability that can work across various imaging objects, phantoms with distinct characteristics should be used to create training and validation datasets. Here, datasets are created from six types of synthetic phantoms, including discs, breads, spiders, lines, logos, and nature pictures, as shown in Fig. 4. These phantoms are defined as the initial pressure distribution in K-WAVE for generating sinogram data and are taken as the ground truth in training.

**Fig. 4.** Representative images of initial pressure distribution used to generate photosacoustic data.

Since the thick and thin blood vessel or PAT like structures can be represented as a sum of straight and curvy lines [35], four types of phantom images (disc, bread, spider and simple lines) were created for training. The disc dataset was comprised of simple phantoms that contain circular objects with number of between one and ten and varied magnitude (from 0.2 to 0.5 KPa).
These circular objects were randomly placed in different positions inside an area of 20mm x 20mm, and their radii varied from 2 pixels to 12 pixels, yielding a total of 27,000 disc phantom images. 13,000 bread phantom images and 10,800 spider phantom images were downloaded from the “quick draw” dataset [50]. The original image with the size of 640 x 480 pixels was resized into 100 x 100 pixels as training phantom images. Furthermore, 6,000 phantom images with simple lines were created. Phantom images with simple lines were created using the following steps: the first step was creating an image which composed of three multi-order polynomials with lots of random coefficients; the second step is to randomly take four arithmetic operations with random coefficients; the last step is to randomly crop an image with 100 x 100 pixels.

In order to recover imaging objects with amorphous complicated structures, 240 logos phantom images and 15,000 natural phantom images were respectively downloaded from the PASCAL VOC2012 dataset [51] and the internet for training. Training phantom images with size of 100 x 100 pixels were randomly sampled from an original image.

In total 72,658 phantom images were obtained to generate PA signals for training. Each phantom image and its corresponding PA signals were considered as a dataset of input-output pairs. We used 80% datasets as training set, the rest 20% as validation set.

For a second dataset for testing, four numerical phantoms namely (1) Disc, (2) PAT, (3) Blood Vessel, and (4) Derenzo were used, which are never used in training and validation datasets.

2.4. Data preprocessing

Due to limited memory resources, the raw sensor data was simulated in batches. The simulated data for all batches was firstly consolidated, then the order was disrupted. Before training, ground truth pairing with input sinogram data were independently normalized using z-score normalization over the whole dataset according to the following equation [42]:

$$X^* = \frac{X - \mu}{\sigma},$$  \hspace{1cm} (4)

where $X$ is the PA raw signal or ground true image of initial pressure distribution, $X^*$ is the normalized data, $\mu$ denotes the mean of $X$, and $\sigma$ is the standard deviation of $X$.

2.5. Evaluation criterion

To validate the performance of Res-Unet based reconstruction algorithms, Pearson correlation (PC) and peak signal-to-noise ratio (PSNR) are used as the quantitative metrics.

PC is used to measure the correlation between the true and the reconstructed images and is defined as [52]:

$$PC(P_0, P_{recon}) = \frac{\text{cov}(P_0, P_{recon})}{\sigma(P_0)\sigma(P_{recon})},$$  \hspace{1cm} (5)

PSNR is defined as [53]:

$$PSNR(P_0, P_{recon}) = 20 \log \left( \frac{\text{Max}(P_0(i, j))}{\sqrt{\frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} (P_0(i, j) - P_{recon}(i, j))^2}} \right),$$  \hspace{1cm} (6)

where $P_0$ and $P_{recon}$ are true and reconstructed images of initial pressure distribution with size of $N \times N$, respectively; $\sigma$ is the standard deviation, and $\text{cov}$ is the covariance.

2.6. Implementation details

All computations were run on a 64-bit PC, having two RTX 2080 GPUs with 8 GB memory. And the network was implemented in python 3.6 with PyTorch [54]. The Res-Unet was trained using the mean squared error (MSE) as loss function. Furthermore, the Adam optimizer was used for
training the Res-Unet for 50 epochs with the learning rate of $10^{-4}$, the weight decay of $10^{-5}$ and the batch size of 32. Total number of trainable parameters for all layers were 88,061,661. It took approximately 15 hours to train the Res-Unet using the above mentioned PC. After being well trained, it only took less than 1s to reconstruct a PA image.

3. Experiments and results

Performance of the proposed method is compared with both the MRR [44] and variants of the Unet-like architecture, including Sta-Unet [28], Incep-Unet [25], Db-Unet [26], Ds-Unet, R2-Unet [27], and Unet++ [29]. The details of these Unet like architectures are provided in Supplemental Figs. S1-S7.

3.1. Disc phantom

The true Disc phantom used in testing data is shown in Fig. 5(a). It includes 9 circular objects with the same size, having “0” for the background and uniformly-varying magnitudes for objects (from 1 to 8). 40dB Gaussian noise was added to synthetic data. The reconstructed image using the MRR algorithm is shown in Fig. 5(b). Figures 5(c)–5(i) show the reconstructed images by DL-based reconstruction algorithms. And the quantitative comparisons between these algorithms are shown in Table 1.

| Metric | MRR | Sta-Unet | Incep-Unet | Db-Unet | Ds-Unet | R2-Unet | Unet++ | Res-Unet |
|--------|-----|----------|------------|---------|---------|---------|--------|----------|
| PC     | 0.32| 0.79     | 0.96       | 0.96    | 0.96    | 0.97    | 0.93   | 0.98     |
| PSNR   | 15.56| 22.98   | 28.62      | 29.25   | 29.11   | 29.42   | 27.13  | 32.49    |

From Fig. 5, it is obviously observed that significant artifacts exist in reconstructed image for the MRR algorithm, which has the lowest PC (0.32) and PSNR (15.56). The DL-based reconstruction algorithms show significant improvements in reconstructed images with sharp edges and less artifacts, because the DL-based algorithms directly learn the mapping between the input sinogram data and the targeted initial pressure distribution, and more effective priors can be learned through supervised learning. The Sta-Unet yields the worst performance with the lower PC and PSNR values of 0.79 and 22.98, respectively. The proposed Res-Unet achieves the best PC and PSNR values of 0.98 and 32.49, respectively. The introduction of residual blocks not only addresses the degradation problem but also deepens the neural network and further improves accuracy. For the other DL-based algorithms, the values of PC are nearly the same (0.93-0.97), but the PSNR values vary significantly. The higher PSNR of 29.42 is achieved for the R2-Unet, but it is 9.45% lower than that of the Res-Unet. Moreover, some small artifacts can be seen in reconstructed images by the R2-Unet.

We also see that it is difficult to accurately recover the target with magnitude of “1” (the white arrows in Fig. 5), even for DL-based reconstruction algorithms. The R2-Unet and Res-Unet might be efficient to recover targets with lower magnitudes. Overall, our results demonstrated that the DL-based algorithms were able to recover PAI images with improved quantitatively accuracy and better image quality, compared with the MRR algorithm.

We further evaluated the performances of the algorithms mentioned above. Different number of circular objects (from one to six) with varied magnitude (from “1” to “5”) and varied radius (from 4 pixels to 10 pixels) were randomly placed inside the imaging area, yielding a cohort of 300 samples. Note that these samples were never used in training or validation datasets. Statistic results for 300 samples are shown in Fig. 6 and summarized in Table 2.

The results demonstrate that the lowest PC value is yielded for the MRR algorithm, and its value is 0.35 ± 0.05. The Sta-Unet has a higher PC of 0.63 ± 0.29, which is 80% higher than the
Fig. 5. Results of the Disc phantom. (a) Ground truth image, (b)-(i) are images reconstructed by algorithms based on the MRR, Sta-Unet, Incep-Unet, Db-Unet, Ds-Unet, R2-Unet, Unet++, and Res-Unet, respectively.

Fig. 6. The statistical results for 300 samples. Boxplot of PC (a) and PSNR (b).
Table 2. Mean ± SD of PC/PSNR for statistical results.

| Method        | MRR   | Sta-Unet | Incep-Unet | Db-Unet | Ds-Unet | R2-Unet | Unet++ | Unet++ |
|---------------|-------|----------|------------|---------|---------|---------|--------|--------|
| PC            | 0.35 ± 0.05 | 0.63 ± 0.29 | 0.95 ± 0.05 | 0.97 ± 0.03 | 0.96 ± 0.04 | 0.96 ± 0.02 | 0.89 ± 0.12 | 0.96 ± 0.04 |
| PSNR          | 24.08 ± 2.82 | 25.20 ± 6.44 | 33.14 ± 4.42 | 35.19 ± 4.33 | 34.08 ± 4.57 | 34.14 ± 2.91 | 30.75 ± 5.57 | 35.56 ± 5.03 |

MRR. Compared to the Sta-Unet, the PC value is significantly improved to 0.89 ± 0.12 for the Unet++, but it is still lower than the other five architectures. As shown in Table 2, it can be seen that the Db-Unet is slightly better than other Unet architectures with higher PC. However, the differences between any two Unet architectures are within ±2%.

Similar improvement occurred in PSNR for the MRR, Sta-Unet and Unet++. However, the PSNR value of Res-Unet (35.56 ± 5.03) are higher than those of other Unet architectures. Overall, the Res-Unet has the best performance in terms of PC and PSNR.

3.2. PAT phantom

The performance of the Res-Unet was further evaluated with a complex PAT phantom, having varied initial pressure distribution from 0 to 4. 40 dB white Gaussian noise was added to synthetic data. The corresponding results are shown in Fig. 7 and Table 3. Similar results as the disc experiment are obtained. Significant artifacts existed in the images reconstructed by the MRR algorithm. The Sat-Unet failed to recover the shape and edges of the PAT image. In this case, the values of PC and PSNR of the Res-Unet are about 2.2% and 6.5% higher than those of the other architectures, as it is evident from the reconstructed images where sharp edges and less artifacts are found when compared with those of other architectures.

Table 3. Quantitative comparisons of different algorithms for the PAT phantom with different levels of noise.

| Noise  | Metric | MRR | Sta-Unet  | Incep-Unet | Db-Unet  | Ds-Unet  | R2-Unet  | Unet++  | Res-Unet  |
|--------|--------|-----|-----------|------------|----------|----------|----------|---------|-----------|
| 60dB   | PC     | 0.31| 0.61      | 0.83       | 0.84     | 0.83     | 0.89     | 0.74    | 0.91      |
|        | PSNR   | 8.07| 12.63     | 14.97      | 15.49    | 15.12    | 17.28    | 14.05   | 17.76     |
| 40dB   | PC     | 0.31| 0.61      | 0.87       | 0.90     | 0.87     | 0.90     | 0.76    | 0.92      |
|        | PSNR   | 8.22| 12.65     | 15.57      | 16.96    | 15.90    | 17.33    | 14.29   | 18.46     |
| 20dB   | PC     | 0.24| 0.54      | 0.68       | 0.71     | 0.68     | 0.79     | 0.60    | 0.77      |
|        | PSNR   | 10.34| 12.11     | 12.95      | 13.56    | 13.14    | 13.78    | 12.53   | 14.46     |

In addition, different levels of noise (20 dB and 60 dB) were added to the synthetic data to further demonstrate their performances. Table 3 lists the PC and PSNR values for different algorithms. From Table 3, higher levels of noise cause significant degradation in reconstructed image quality. For example, the PC and PSNR values of Res-Unet were reduced by 15.4% and 18.6% when the noise level was increased from 60 dB to 20 dB. Although Res-Unet did not yield the highest PC value, the PC value was only 2.5% lower than that of R2-Unet. However, the PSNR value was 4.9% higher than that of R2-Unet.

3.3. Blood vessel phantom

Since PAI has been widely used for visualizing blood vessel structures, a blood vessel phantom with varied initial pressure distribution from 0 to 10 was used to demonstrate the performances of different algorithms in reconstructing thin vessels. Note that the phantom was never used in training and validation datasets. The corresponding reconstructed images are shown in Fig. 8, and the quantitative comparisons are shown in Table 4. The results further reveal that the end-to-end DL-based algorithms provide the capability to recover PA images. Again, the blurred image
Fig. 7. Results of the PAT phantom. (a)-(i) are the ground truth image and the images reconstructed by algorithms based on the MRR, Sta-Unet, Incep-Unet, Db-Unet, Ds-Unet, R2-Unet, Unet++, and Res-Unet, respectively.
was obtained by the MRR algorithm, as shown in Fig. 8(b). Similar with above results, obvious artifacts existed in reconstructed images for the Sta-Unet and Unet++, and the Res-Unet yielded the best image quality in terms of PC (0.80) and PSNR (24.07). Compared to the Sta-Unet, the image quality can be improved with Incep-, Db-, Ds- and R2-Unets.

![Fig. 8. Results of the vessel phantom. (a) Ground truth, (b) BP, (c) Sta-Unet, (d) Incep-Unet, (e) Db-Unet, (f) Ds-Unet, (g) R2-Unet, (h) Unet++, and (i) Res-Unet.](image)

| Methods | MRR    | Sta-Unet | Incep-Unet | Db-Unet | Ds-Unet | R2-Unet | Unet++   | Res-Unet |
|---------|--------|----------|------------|---------|---------|---------|----------|----------|
| PC      | 0.40   | 0.63     | 0.75       | 0.73    | 0.72    | 0.74    | 0.70     | **0.80** |
| PSNR    | 15.69  | 21.65    | 22.97      | 22.57   | 22.47   | 22.74   | 22.40    | **24.07**|

### 3.4. Derenzo phantom

In the last simulation experiment, we further evaluated their effectiveness with the Derenzo phantom with varied initial pressure distribution from 0 to 2.5. The corresponding results are shown in Fig. 10, and the quantitative comparisons are summarized in Table 5. The results
confirmed that it is feasible and beneficial for the end-to-end Unet architectures to recover the initial pressure distribution. As shown in Fig. 10(b), the MRR algorithm leads to severally artifacts. Note that the architectures with Sta-, Incep-, Db-, Ds- and Unet++ failed to differentiate small targets, and their reconstructed images were distorted. Compared with the above five Unet architectures, a little image distortion can be seen in the reconstructed images by R2- and Res-Unets. The PC and PSNR values for the R2-Unet are 5% and 5.72% higher than those from the Res-Unet.

| Metric | MRR | Sta-Unet | Incep-Unet | Db-Unet | Ds-Unet | R2-Unet | Unet++ | Res-Unet |
|--------|-----|----------|------------|---------|---------|---------|--------|----------|
| PC     | 0.41| 0.28     | 0.66       | 0.66    | 0.68    | 0.84    | 0.60   | 0.80     |
| PSNR   | 6.57| 9.16     | 10.89      | 11.18   | 11.26   | 14.05   | 10.72  | 13.29    |

Figure 9(b) plots the cross-section profiles along the red dotted line shown in Fig. 10(a). The results demonstrate that the recovered images using the R2-Unet and Res-Unet are closer to their true values compared with other Unet architectures. The results also reveal that the quantitative error increases with the complexity of testing data.

3.5. Physical phantom experiment

To further evaluate the performances of the Res-Unet, a physical phantom experiment was conducted with a MSOT inVision 128 imaging system. Three capillaries containing solution of indocyanine green (ICG) were placed into the imaging chamber of MSOT inVision to acquire PA signals, as shown in Figs. 3(b) and 3(c). The same experimental setting as the simulation experiments was used. In this case, the MRR algorithm failed to reconstruction PA images. Therefore, the result was not shown here. Instead, the reconstructed image from the MSOT system is shown in Fig. 11(a), where aliasing artifacts were produced. However, the Sta-, Db-, Ds- and R2- Unet architectures failed to reconstruct PA images, as shown in Figs. 11(b) and 11(d)–11(f). For the Incep-Unet and Unet++, artifacts were introduced in the reconstructed images, as shown in Figs. 11(c) and 11(g). In contrast, Res-Unet produced PA image with less artifacts (Fig. 11(h)).
Fig. 10. Results of the Derenzo phantom. (a) Ground truth, (b) BP, (c) Sta-Unet, (d) Incep-Unet, (e) Db-Unet, (f) Ds-Unet, (g) R2-Unet, (h) Unet++, and (i) Res-Unet.
Fig. 11. The results of the physical phantom experiment. (a) MSOT, (b) Sta-Unet, (c) Incep-Unet, (d) Db-Unet, (e) Ds-Unet, (f) R2-Unet, (g) Unet++, and (h) Res-Unet.

4. Discussion

Conventional methods for PAI mainly include BP, TR and model-based reconstruction algorithms. PA images reconstructed by the BP algorithms may suffer from severe artifacts, resulting in distorted images. The main drawback of TR reconstruction algorithms is computationally expensive. The performances of model-based reconstruction algorithms strongly depend on a special optimization algorithm. In addition, the model-based reconstruction algorithms require repeated calculation of the forward and adjoint operators during reconstruction. Therefore, they are time-consuming. To overcome these limitations in this paper we develop a new reconstruction approach based on an end-to-end Res-Unet that comes with the following properties: (1) it yields high quality images; (2) it has low numerical complexity. The performance was demonstrated and compared with other DL-based algorithms with numerical and physical phantom experiments. Our results demonstrate that the end-to-end Res-Unet can provide superior performance and has the ability to accurately reconstruct PA images with sharp edges and less artifacts.

Our results also demonstrate that the Sta-Unet may not be an alternative architecture to reconstruct PA image, even with higher quantitative accuracy than the MRR algorithm. Unet++ provides improved performance, but its ability to recover PA images is limited because the network is difficult to train and is prone to overfitting due to integrating the characteristics of different layers. It is possible to incorporate the inception block, dense block, recurrent residual block or residual block into the Unet architecture to improve its performance. However, the performance of Incep-Unet is poor in terms of PC and PSNR, because it has poor power to extract features. Introducing dense skip can improve the performance in recovering PA images, because each latter row combines the information of the previous rows in the contracting path of Ds-Unet, therefore, more features and potential information are extracted. Compared to the inception block or the dense skip, introduction of dense block can improve the quality of reconstructed images. The benefit is from concatenating feature-maps learned by different layers, which efficiently extract more information. All simulation results show that superior performance with higher image quality is achieved for the Res-Unet. The use of the residual block strongly encourages feature reuse and improves information extraction throughout the
network. Furthermore, regularized effects are imposed by the residual blocks, which reduce the likelihood of overfitting to the training data.

The proposed Res-Unet is trained in supervised learning, which requires ground truth pairing with input sinogram data. Therefore, the performance of the Res-Unet critically depends on the training data. Figure 12 shows the results of the Res-Unet with reduced number of training dataset. From Fig. 12, it is observed that the values of PC and PSNR decrease with the reduction of the number of training data. For the Disc phantom, the PC value was 0.91 when the number of the training data was reduced to 10,000. And for the Derenzo phantom, quantitative performance of the Res-Unet degraded significantly when the number of training data was reduced to 10,000. This is because the limited information is learned from such a small training dataset. Therefore, a large scale and meaningful data set with distinct characteristics is crucial for training. Our results also demonstrate that working with simulated data might not be enough for real measurement data. When the trained networks on simulated data were applied to real measurements, there were obvious degradations or differences in the reconstructed images compared to the real images due to the domain gap [55,56]. Nevertheless, it is difficult to obtain sufficient real data for training. Meanwhile, the ground truth is not available for training real data. One of the possible solutions is to use a pixel-wise technique [57].

![Fig. 12. The reconstruction performances of the Res-Unet with limited training data for the Disc and the Derenzo phantoms. (a) PC; (b) PSNR.](image-url)

When the PA raw sensor data dimension is changed, it cannot be directly inputted to the well-trained Res-Unet. An alternative technique is to resize the raw data dimension into 512×128. We demonstrated its feasibility with the PAT phantom. The corresponding results are shown in Table 6. When the dimension of sensor data is reduced from 512×128 to 512×64 or 512×100, the PC is slightly lower, but within the range of 1%. However, the average PSNR value is reduced 3.36%. When the dimension of sensor data is increased from 512×128 to 512×150, 512×256 or 1024×128, excepted improvement is not obtained in image quality. For high image quality, it would be better to retrain the network.

| Size of sensor data | 512×64 | 512×100 | 512×128 | 512×150 | 512×256 | 1024×128 |
|---------------------|--------|---------|---------|---------|---------|----------|
| PC                  | 0.92   | 0.90    | 0.92    | 0.90    | 0.91    | 0.92     |
| PSNR                | 18.23  | 17.45   | 18.46   | 17.64   | 18.05   | 18.29    |

Table 6. The effects of the dimension of sensor data on image quality using the proposed Res-Unet for PAT phantom.
Another limitation is that the performance of the proposed algorithm is sensitive to sensor placement. Figure 13 and Table 7 show the reconstruction results with deviations of sensor placement. Each variation led to deterioration of reconstruction quality. If larger changes in the sensor placement are expected, it is recommended to either retrain the network or perform an update training.

![Fig. 13. The reconstructed images with deviations of sensor placement for PAT phantom using the Res-Unet.](image)

| Deviation of sensor placement | 0° | −2° | 2° | −4° | 4° | −6° |
|------------------------------|----|-----|----|-----|----|-----|
| PC                           | 0.92 | 0.87 | 0.88 | 0.74 | 0.70 | 0.60 |
| PSNR                         | 18.46 | 16.47 | 16.83 | 13.96 | 13.49 | 12.50 |

Table 7. The results with deviations of sensor placement for PAT phantom using the Res-Unet.

Table 8 shows the number of training parameters and training time for all network architectures. The number of training parameters for the Res-Unet was 88,061,661. It took approximately 15 hours for training from a 512×128-sized dataset. However, for high image quality, more sensors with longer recorded points are desirable. This would need computational resources exceeding our PC (two RTX 2080 GPUs with 8 GB memory).

![Table 8. Number of training parameters and training time for different network architectures.](image)

| Model                  | Sta-Unet | Incep-Unet | Db-Unet | Ds-Unet | R2-Unet | Unet++ | Res-Unet |
|------------------------|-----------|------------|---------|---------|---------|--------|----------|
| # Parameter            | 3507265   | 52795489   | 71310463 | 70467553 | 18427213 | 146202117 | 88061661 |
| GigaFlops              | 5.79      | 9.13       | 22.69    | 11.75   | 25.02   | 19.16   | 17.09    |
| Training time          | 4.25h     | 12h        | 44.33h   | 12h     | 12.5h   | 13.33h  | 15h      |

As mentioned above, there is also an increasing interest to use deep neural network (DNN) as a denoiser to remove artifacts in reconstructed PA images. Actually, it is a postprocessing based reconstruction algorithm. To further demonstrate its performance of the proposed end-to-end reconstruction algorithm, we compared it with a postprocessing method. To perform the postprocessing based reconstruction algorithm, the BP algorithm was firstly used to reconstruct initial pressure distribution, and then a Res-Unet with the architecture shown in Supplemental Fig. S8 was used as a denoiser to remove artifacts. The corresponding results are shown in Fig. 14 and Table 9. As seen from Fig. 14, the performance of our proposed end-to-end reconstruction algorithm is inferior to that of the post-processing reconstruction algorithm.

However, the end-to-end Res-Unet avoids the computation of photoacoustic propagation, and the implementation of PAI reconstruction requires less than one second with the well-trained network. Therefore, the end-to-end Res-Unet can be used to reconstruct PA images in real time.
Fig. 14. The reconstruction results for different phantoms using the post-processing Res-Unet.

Table 9. Quantitative comparisons between the algorithms with a post-processing Res-Unet and end-to-end Res-Unet.

| Metric | Method      | Disc | PAT | Vessel | Derenzo |
|--------|-------------|------|-----|--------|---------|
|       | End-to-end  | 0.98 | 0.92| 0.80   | 0.80    |
| PC     | Post-processing | 0.99 | 0.96| 0.95   | 0.93    |
| PSNR   | End-to-end  | 32.49| 18.46| 24.07  | 13.29   |
|        | Post-processing | 36.69| 21.94| 29.43  | 17.57   |

Meanwhile, the training time could be greatly reduced using Apex, a Pytorch extension with NVIDIA-maintained utilities to streamline mixed precision and distributed training.

5. Conclusion

In the manuscript, we designed and experimentally demonstrated an end-to-end Res-Unet for directly reconstructing PA images from raw sensor data by utilizing the original U-net as the deep neural network architecture, which is composed of a contracting path and a symmetric expanding path. The contracting path is used to extract comprehensive context from sinogram data, and the symmetric expanding path is used to infer initial pressure distribution. In addition, we add a skip connection between input and output, which means that the network actually learns the difference between input and output. Furthermore, in order to prevent the deep network from degradation and improve image quality, the residual learning mechanism is adopted. The proposed algorithm takes the entire raw signals as inputs, so the reconstruction makes the best of all the input PA signals via learning weights/filters to mitigate ill-posedness of image reconstruction [30]. The Res-Unet successfully produced high-quality images with sharp edges and less artifacts and outperformed the MRR algorithm and state-of-art Unet and Unet++. These features make the end-to-end Res-Unet promising alternative to conventional PAI reconstruction methods. Despite the potential of the end-to-end Res-Unet in recovering PAI images, further work should focus on improving reconstructed image from experimental data by deploying the trained network from simulation data. Recently, Navchetan Awasthi et al., have demonstrated the feasibility for the denoising performed in the data-domain followed by an image reconstruction algorithm [35]. From the point of view, a network working as denoiser may be used as the starting block in the Res-Unet. And the feasibility and effectiveness will be addressed in future research.

Funding

National Natural Science Foundation of China (81871394); Beijing Municipal Commission of Education (KM201810005030).
Disclosures

The authors have no relevant financial interests in the manuscript.

See Supplement 1 for supporting content.

References

1. L. V. Wang and J. Yao, “A practical guide to photoacoustic tomography in the life sciences,” Nat. Methods 13(8), 627–638 (2016).
2. G. Fei and Y. Zheng, “A correlated microwave-acoustic imaging method for early-stage cancer detection,” in International Engineering in Medicine and Biology Conference (EMBC 2012), pp. 480–483.
3. F. Guo, Q. Peng, X. Feng, B. Gao, and Y. Zhang, “Single-wavelength blood oxygen saturation sensing with combined optical absorption and scattering,” IEEE Sensors J. 16(7), 1943–1948 (2016).
4. G. Ku, X. Wang, X. Xie, G. Stoica, and L. V. Wang, “Imaging of tumor angiogenesis in rat brains in vivo by photoacoustic tomography,” Appl. Opt. 44(5), 770–775 (2005).
5. K. Chulhong, C. Favazza, and L. V. Wang, “In vivo photoacoustic tomography of chemicals: high-resolution functional and molecular optical imaging at new depths,” Chem. Rev. 110(5), 2756–2782 (2010).
6. G. Ku, X. Wang, X. Xie, G. Stoica, and L. V. Wang, “Imaging of tumor angiogenesis in rat brains in vivo by photoacoustic tomography,” Appl. Opt. 44(5), 770–775 (2005).
7. K. Wang and M. A. Anastasio, Photoacoustic and Thermoacoustic Tomography: Image Formation Principles (Springer, 2015).
8. M. Xu and L. V. Wang, “Universal back-projection algorithm for photoacoustic computed tomography,” J. Acoust. Soc. Am. 112(4), 1536–1544 (2002).
9. M. Bhatt, A. Acharya, and P. K. Yalavarthy, “Computationally efficient error estimate for evaluation of regularization in photoacoustic tomography,” J. Biomed. Opt. 21(10), 106002 (2016).
10. Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” Nature 521(7553), 436–444 (2015).
11. A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, “Dermatologist-level classification of skin cancer with deep neural networks,” Nature 542(7639), 115–118 (2017).
12. K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in IEEE Conference on Computer Vision and Pattern Recognition (IEEE, 2016), pp. 770–778.
13. C. Szegedy, S. Ioffe, and V. Vanhoucke, “Inception-v4, Inception-Resnet and the impact of residual connections on learning,” arXiv:1602.07261 (2016).
14. G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, “Densely connected convolutional networks,” in IEEE Conference on Computer Vision and Pattern Recognition (IEEE, 2017), pp. 4700–4708.
15. M. Z. Alom, M. Hasan, C. Yakopcic, T. M. Taha, and V. K. Asari, “Recurrent residual convolutional neural network based on U-net (R2U-net) for medical image segmentation,” arXiv:1802.06955 (2018).
28. 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 (MICCAI 2015)*, pp. 234–241.

29. Z. Zhou, M. M. R. Siddiquee, N. Tajbahksh, and J. Liang, “UNet++: a nested U-Net architecture for medical image segmentation,” arXiv:1807.10165 (2018).

30. K. H. Jin, M. T. McCann, E. Froustey, and M. Unser, “Deep convolutional neural network for inverse problems in imaging,” *IEEE Trans. Image Process.* 26(9), 4509–4522 (2017).

31. V. S. Kadimsetty, S. Gupta, S. Ganapathy, and P. K. Yalavarthy, “Convolutional neural network-based robust denoising of low-dose computed tomography perfusion maps,” *IEEE Trans. Radiat. Plasma Med. Sci.* 3(2), 137–152 (2019).

32. S. Guan, A. A. Khan, S. Sikdar, and P. V. Chitnis, “Fully dense Unet for 2-D sparse photoacoustic tomography artifact removal,” *IEEE J. Biomed. Health Inform.* 24(2), 568–576 (2020).

33. S. Anholzer, M. Hallmeier, and J. Schwab, “Deep learning for photoacoustic tomography from sparse data,” *Inverse Probl. Sci. Eng.* 27(7), 1005–1019 (2019).

34. M. W. Kim, G. Jeng, I. Pelivanov, and M. O’Donnell, “Deep-learning image reconstruction for real-time photoacoustic system,” *IEEE Trans. Med. Imaging* 1 (2020), doi: 10.1109/TMI.2020.2938385.

35. N. Awasthi, G. Jain, S. K. Kalva, M. Pramanik, and P. K. Yalavarthy, “Deep neural network based sinogram super-resolution and bandwidth enhancement for limited-data photoacoustic tomography,” *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* 1 (2020), doi: 10.1109/TUFFC.2020.2977210.

36. D. Allman, A. Reiter, and M. A. Lediju Bell, “Photoacoustic source detection and reflection artifact removal enabled by deep learning,” *IEEE Trans. Med. Imaging* 37(6), 1464–1477 (2018).

37. T. Kirchner, J. Gröhl, and L. Maier-Heina, “Context encoding enables machine learning-based quantitative photoacoustics,” *J. Biomed. Opt.* 23(05). 1 (2018).

38. N. Davoudi, X. Luís Deán-Ben, and D. Razansky, “Deep learning photoacoustic tomography with sparse data,” *Nat. Mach. Intell.* 1(10), 453–460 (2019).

39. C. Cai, K. Deng, C. Ma, and J. Luo, “End-to-end deep neural network for optical inversion in quantitative photoacoustic imaging,” *Opt. Lett.* 43(12), 2752–2755 (2018).

40. H. Lan, C. Yang, D. Jiang, and F. Gao, “Reconstruct the photoacoustic image based on deep learning with multi-frequency ring-shape transducer array,” in *International Engineering in Medicine and Biology Conference (EMBC 2019)*, pp. 7115–7118.

41. K. Johnstonbaugh, S. Agrawal, D. A. Durairaj, M. Homewood, S. P. K. Karri, S. Kothapalli, A. A. Oraevsky, and L. V. Wang, “Novel deep learning architecture for optical fluence dependent photoacoustic target localization,” *Proc. SPIE* 10878, 108781L (2019).

42. D. Waibel, J. Gröhl, F. Jensee, T. Kirchner, K. Maier-Hein, and L. Maier-Hein, “Reconstruction of initial pressure from limited view photoacoustic images using deep learning,” *Proc. SPIE* 10494, 104942S (2018).

43. H. Lan, D. Jiang, C. Yang, and F. Gao, “Y-Net: a hybrid deep learning reconstruction framework for photoacoustic imaging in vivo,” arXiv:1908.00795 (2019).

44. D. R. Sammy, J. Prakash, S. K. Kalva, M. Pramanik, and P. K. Yalavarthy, “Spatially variant regularization based on model resolution and fidelity embedding characteristics improves photoacoustic tomography,” *J. Biomed. Opt.* 23(10), 1 (2018).

45. A. Rosenthal, V. Nitziachristos, and D. Razansky, “Acoustic inversion in photoacoustic tomography: a review,” *Curr. Med. Imaging Rev.* 9(4), 318–336 (2014).

46. S. Gutta, V. S. Kadimsetty, S. K. Kalva, M. Pramanik, S. Ganapathy, and P. K. Yalavarthy, “Deep neural network-based bandwidth enhancement of photoacoustic data,” *J. Biomed. Opt.* 22(11), 1 (2017).

47. M. A. Dedmari, S. Conjeti, S. Estrada, P. Ehses, T. Stöcker, and M. Reuter, “Complex fully convolutional neural networks for MR image reconstruction,” *International Workshop on Machine Learning for Medical Image Reconstruction, MLMR 2018*, pp. 30–38, 2018.

48. S. Ioffe and C. Szegedy, “Batch normalization: accelerating deep network training by reducing internal covariate shift,” in *International Conference on Machine Learning (ICML 2015)*, pp. 448–456.

49. B. E. Treeby and B. T. Cox, “k-Wave: MATLAB toolbox for the simulation and reconstruction of photoacoustic wave fields,” *J. Biomed. Opt.* 15(2), 021314 (2010).

50. https://console.cloud.google.com/storage/browser/quickdraw_dataset.

51. http://home.bharathh.info/pubs/codes/SBD/download.html.

52. A. P. Cuadros and G. R. Arce, “Coded aperture optimization in compressive X-ray tomography: a gradient descent approach,” *Opt. Express* 25(20), 23833–23849 (2017).

53. A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, and A. Lerer, “Automatic differentiation in pytorch,” in *Conference on Neural Information Processing Systems (NIPS 2017)*, pp. 1–4.

54. K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” arXiv:1409.1556 (2015).

55. J. Tremblay, A. Prakash, D. Acuna, M. Brophy, V. Jampani, C. Anil, T. To, E. Cameracci, S. Boochoon, and S. Birchfield, “Training deep networks with synthetic data: bridging the reality gap by domain randomization,” arXiv:1804.06516 (2018).
56. J. Gröhl, T. Kirchner, T. Adler, and L. Maier-Hein, “Confidence estimation for machine learning-based quantitative photoacoustics,” J. Imaging 4(12), 147 (2018).

57. J. Gröhl, T. Kirchner, T. Adler, and L. Maier-Hein, “Estimation of blood oxygenation with learned spectral decoloring for quantitative photoacoustic imaging (LSD-qPAI),” arXiv:1902.05839 (2019).