2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Learned Parameters and Increment for Iterative Photoacoustic Image Reconstruction via Deep Learning

Zhuoan Li, Hengrong Lan, and Fei Gao*, *Member, IEEE*

Abstract— **Photoacoustic (PA) tomography is a relatively new medical imaging technique that combines traditional ultrasound imaging and optical imaging, which has great application prospects in recent years. To reveal the light absorption coefficient of biological tissues, the images are reconstructed from PA signals by reconstruction algorithms. However, traditional model-based reconstruction method requires a huge number of iterations to obtain relatively good experimental results, which is quite time-consuming. In this paper, we propose to use deep learning method to replace brute parameter adjustment in model-based reconstruction, and speed up the rate of convergence by building convolutional neural networks (CNN). The parameters we defined in our model can be learned automatically. Meanwhile, our method can optimize the increment of gradient in each step of iteration. The numerical experiment validates our method, showing that only three iterations are needed to obtain the satisfactory image quality, which normally requires 10 iterations for tradition method. It demonstrated that efficiency of photoacoustic reconstruction can be greatly improved by our proposed method, compared with traditional model-based methods.**

Index Terms— **Photoacoustic tomography, Deep learning, Convolutional neural networks, Reconstruction.**

I. Introduction

Photoacoustic tomography (PAT) is an emerging biomedical imaging modality, which combines the advantages of optical and ultrasound imaging technologies. It can get biomedical images without ionizing radiation and damaging the tissue. PAT is based on photoacoustic effect, which refers to the generation of ultrasonic waves following light absorption by the sample [1-3]. When illuminating the sample tissue with pulsed laser, transient temperature rise will lead to thermoelastic expansion, which emits ultrasonic waves detected by the ultrasound transducer. Since PAT does not suffer the limitation of light diffusion, it can achieve high spatial resolution at deep penetration in tissues, and demonstrated wide applications these years.

In a photoacoustic computed tomography (PACT) system, ultrasound data collected by sensors will serve as input into image reconstruction algorithm. It will produce maps of the absorbed optical energy density within the sample tissue. High temporal resolution allows PACT to have promising applications in brain imaging, breast cancer diagnosis and whole-body imaging in mice. However, there exists ill-posed PACT reconstruction that artifacts will be shown on the image due to sampling under Nyquist sampling requirement or limited-view conditions. The ill-posed problem cannot be solved traditionally by using the methods like time reversal (TR) and filtered back projection (FBP) [4], while model-based method shows superior performance in ill-posed condition [5-7]. Considering that traditional algorithms need to iterate a huge number of times and adjust the hyper-parameters in order to get better results, the application of model-based method is limited, especially in real-time imaging scenarios.

Recently, deep learning has received a lot of attention from both research institutes and industries. The core of deep learning is feature learning, in which hierarchical networks are used to obtain hierarchical feature information to overcome important challenges that previously required manually designed feature extraction operators to do so. Deep learning has shown advantageous features like time-saving, and it can get excellent image quality in the field of photoacoustic imaging reconstruction [8-11].

In this paper, a simple convolutional neural networks (CNN) is used to avoid the tedious adjustment of the hyper-parameters, and accelerate the rate of convergence in PA image reconstruction. We divide each step of iterative items and optimize the increment of gradient. The numerical experiment shows that our method has a superior performance in a fewer iterative step (3 times) compared with traditional method (10 times), which dramatically reduces the time to get the gradient descent (GD) result for limited-view PA data. Furthermore, the peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) [12] of the image shows that our method has much better performance on PA image reconstruction compared with traditional method.

II. Methodology

A. Forward and inverse model

In PAT system, we develop a model to describe the propagation of the acoustic wave. The photoacoustic pressure follows the equations below.

$$(\partial_u - c^2\Delta)p(x,t) = 0 \tag{1}$$

Besides, we have some initial conditions: $p(x,t=0) = f(x), \partial_t p(x,t=0) = 0$. c represents the speed of sound propagation, Δ stands for the spatial Laplacian and $p(x,t)$ shows the pressure at spatial location x and moment t.

Since the wave propagation equation is such complex, we simply define a linear mapping for the model-based PAT:

$$Ax = y \tag{2}$$

which is forward model of PAT. The corresponding inverse model is the problem we need to solve, so-called image reconstruction. A is a forward operator, x is the image we need

Zhuoan Li, Hengrong Lan and Fei Gao are with the Hybrid Imaging System Laboratory, Shanghai Engineering Research Center of Intelligent Vision and Imaging, Shanghai Engineering Research Center of Energy Efficient and Custom AI IC, School of Information Science and Technology, ShanghaiTech University, Shanghai 201210, China (*corresponding author: gaofei@shanghaitech.edu.cn).

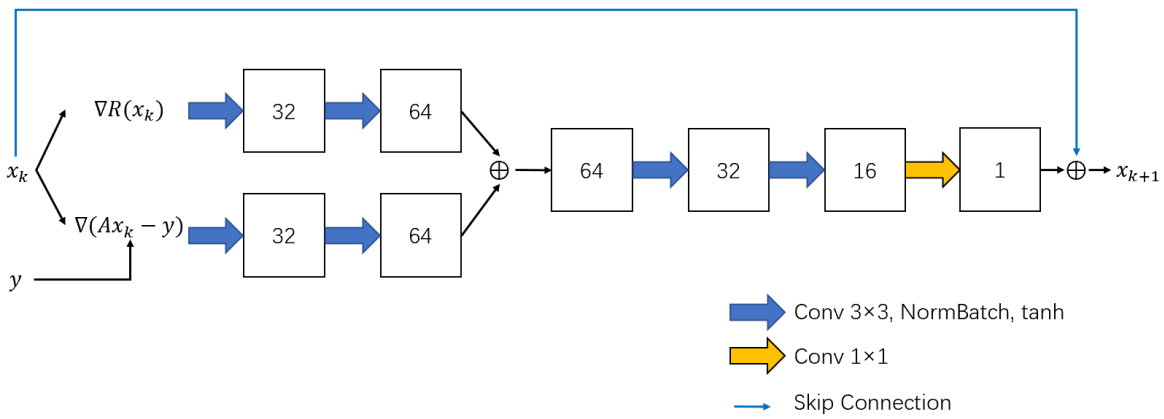

Figure 1. The concrete convolutional neural network structure of our model. The diagram shows one convolutional neural network which representing one iteration of the deep gradient descent and x_k represents the current estimated image. R(x_k) is the regularization item and we use total variation (TV) method to calculate it. ∇ is gradient operator. The blue arrows denote a convolutional layer with 3 × 3 kernels and the resulting channels in each layer are indicated in the squares. The yellow arrow denotes a convolutional layer with 1 × 1 kernels.

to reconstruct, and y represents the measured pressure distribution on time series. Our reconstruction task is mapping the measured pressure data on time series y, which were detected by the acoustic sensors, to a high spatial resolution medical image.

Theoretically, we can simply get the image at the formula level by multiplying y by A^*, where A^* represents the inverse of the forward operator A. Since there exists ill-posed condition, it is really difficult for us to get A^* by direct measurement. In this situation, we try to minimize the least squares error:

$$\min_x \|Ax - y\|_2^2 . \tag{3}$$

to get better reconstruction results.

B. Learnable reconstruction

A common approach to find a solution to Eq. (3) is given by a gradient descent (GD) scheme. It follows the equation shown below

$$x_{k+1} = x_k - \Upsilon \nabla(Ax_k - y), k \geq 0. \tag{4}$$

In order to get the estimate value as close as possible to the ground truth, we can define a learnable coefficient $\Upsilon \in \mathbb{R}$ which is small enough and can be optimized after each iteration. Besides, $F(x_k) \geq F(x_{k+1})$ and the sequence (x_k) will converges to the minimum. To get a better performance in shorter time, we add a regularization item $R(x_k)$ to improve the method. Therefore, the penalty function Eq. (3) can be rewritten as:

$$\min_x \|Ax - y\|_2^2 + \lambda R(x). \tag{5}$$

We calculate regularization item $R(x_k)$ using total variation (TV) method [13]. Similarly, we get best estimate value by gradient descent.

The TV regularization term can be designed differently according to the reconstruction goals. Since traditional TV algorithm sometimes fail to preserve the edges details of the image, the TV of parameters in our model was set as

$$TV = \sum_{i,j} \sqrt{\left(A_{i,j} - A_{i-1,j}\right)^2 + \left(A_{i,j} - A_{i,j-1}\right)^2}, \tag{6}$$

where i and j denotes the location indexes of TV. Then the iteration function of GD for the k^{th} step became

$$x_{k+1} = x_k - \Upsilon \nabla(Ax_k - y) - \lambda \nabla TV(x_k), k \geq 0. \tag{7}$$

Since it is difficult to adjust the values of two parameters at the same time, we introduce CNN to our work to make the parameter learnable and let our adjusting process more time-saving. Data training is performed simultaneously using two structurally similar networks to obtain the parameters, Υ and λ separately. In the k-th step, the image is updated from x_k to x_{k+1} using:

$$x_{k+1} = x_k + CNN(\nabla(Ax_k - y), \nabla TV(x_k)), k \geq 0. \tag{8}$$

Our work is first implemented at the simulation level, using MATLAB toolbox k-Wave [14] to set the initial conditions for the model. We generate enough raw training data and train it with the CNN. We simulate in half-ring transducer condition, which contains 64 channels with 19 mm radius to receive the photoacoustic signals. We set the sound speed as 1500 m/s. The whole region is 40 mm × 40 mm with 380 × 380 grids, while our region of interest (ROI) is of 26.95×26.95 mm size. Furthermore, the center frequency of the sensor is 2.5 MHz with 80% fractional bandwidth. Moreover, our reconstruction images have 256 × 256 pixels.

Pytorch [15] is used to train our CNN that makes parameters Υ and λ learnable. For each iteration, MATLAB calculates the gradient of estimate error $Ax_k - y$ as well as regular term $TV(x_k)$, where A is forward operator. Put $\nabla(Ax_k - y)$ and $\nabla TV(x_k)$ into our two CNN that enable to help us get two important learnable parameters Υ and λ, where ∇ denotes gradient descent operator. To constrain iterations, we use MSE loss to train our model:

$$\text{Loss}(x_k) = \sum MSE(x - x_k) \tag{9}$$

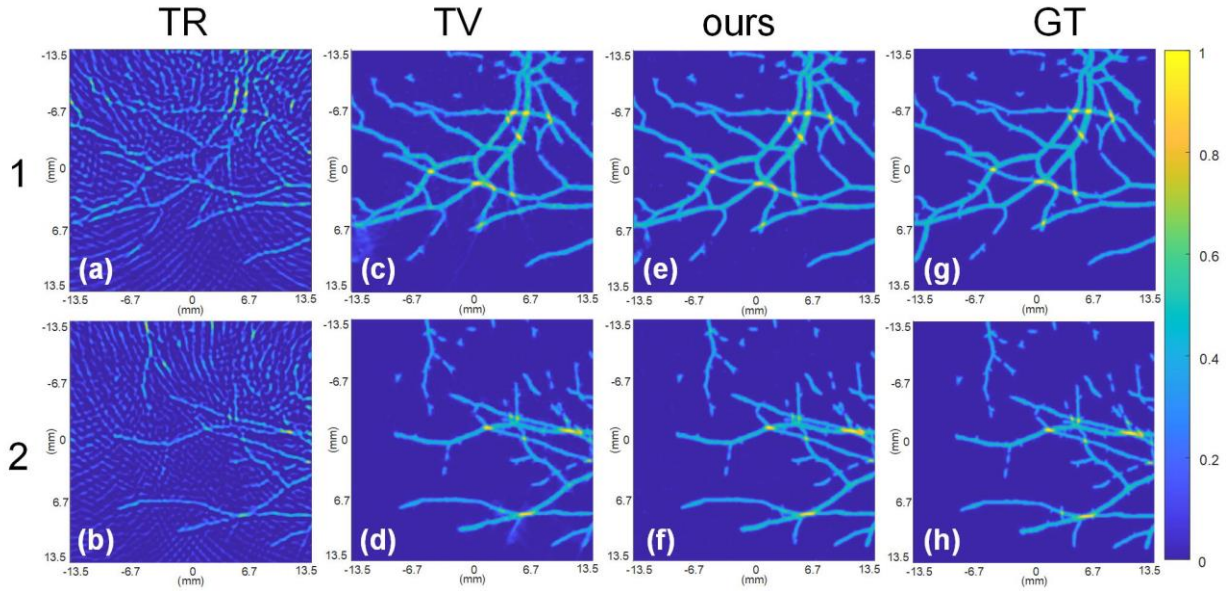

Figure 2. Performance comparison of two examples among different methods. (a) and (b) show the reconstruction results by using time reversal (TR) method, while (c) and (d) traditional total variation (TV) with 10 iterations. Furthermore, (e) and (f) are the images reconstructed using our model. of our model. (g) and (h) are Ground-truth (GT) of each example.

We can get better parameters after each iteration, and our model converges only after 3 iterations.

The specific architecture we have used for the CNNs is illustrated in Figure 1. For each iteration, we put $\nabla(Ax_k - y)$ and $\nabla TV(x_k)$ to a similar pipeline which contains 2 layers. Both $\nabla(Ax_k - y)$ and $\nabla TV(x_k)$ are spread to 32 and then 64 channels. We then add two results together and put them in two 3 × 3 convolutional layers to combine 64 channels to 32 channels, then reduce to 16 channels. Finally, it reduced to 1 channel by a 1 × 1 convolutional layer, and the result is added to the current iteration.

III. EXPERIMENTS AND RESULTS

A. Experiments

We perform a few operations to expand the data capacity under condition that the initial acoustic pressure distribution is simulated using the public dataset DRIVE [16]. Our training set and test set consist 4000 images and 400 images, respectively. First, we divided a few complete blood vessel images into four equal parts, then did rotational transform and superpose two parts randomly. Finally, we import these pre-processed data into k-Wave toolbox as our initial pressure distribution.

Adam optimizer is used to train our model. Moreover, the initial learning rate of our model is 0.004, and we adjust it slightly according to the learning situation. The batch size equals to 32 and the number of epochs is 300 in our model.

B. Results

Two traditional algorithms are used to compare with our method: TR and TV with traditional GD. Fig. 2(a) and (b) show the reconstructed image of time reversal (TR) method. Since the measurement surface is often irregular and incomplete with limited ultrasound transducer coverage, we can see there exists a lot of arc-like artifacts across the reconstruction images. Fig. 2(c) and (d) show the result of TV reconstruction method.

Though TV algorithm is able to reconstruct relatively accurate images from sparse data, the reconstructed results shown above are still not as great as our model especially for some capillary parts. The results of TV are obtained after 10 iterations, while our model shows even better reconstructed images after only 3 iterations.

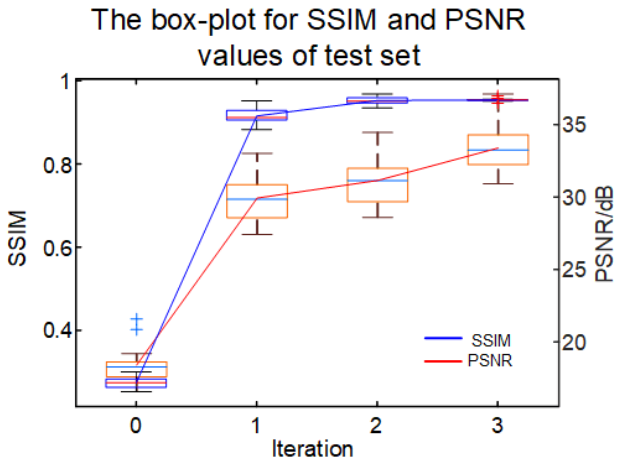

Figure 3. Box-plot after 3 iterations using our model.

Fig. 3 shows the Box-plot of our model, both SSIM and PSNR are greatly improved after the first iteration. Finally, we can find that the mean value of SSIM increases from 0.2735 (1st iteration) to 0.9540 (3rd iteration) and the mean value of PSNR increases from 18.4016 dB to 33.3899 dB after 3 iterations.

Below is a table showing the comparison of two examples' PSNR among TR, TV and our model.

TABLE I. THE PSNR VALUES OF FIGURE 2

PSNR (dB)	TR	TV	Ours
1	17.6558	32.1013	**32.307**
2	19.9657	35.5058	**37.1162**

We also compared SSIM of our model and traditional algorithm in TABLE II.

TABLE II. THE SSIM VALUES OF FIGURE 2

SSIM	TR	TV	Ours
1	0.2487	0.9111	**0.9545**
2	0.2629	0.9523	**0.9636**

The PSNR values in TABLE I show that our method with only 3 iterations can surpass the performance of TV with 10 iterations. The SSIM value in Table II further demonstrate the superiority of our model.

IV. CONCLUSION

In this paper, we introduce deep learning method to accelerate the model-based photoacoustic tomography reconstruction process and simplify the tedious adjustment of parameters. We build a new model that eliminates all hyper-parameters in iterative procedure. It avoids the time-consuming adjustment in model-based methods and speeds up the process of reconstruction. We use k-Wave toolbox of MATLAB to calculate the gradient of two parts and train CNN with Pytorch. The reconstruction results of our model show much better performance compared with the results reconstructed by TV method with 10 iterations. We will further optimize our model by training it with *in vivo* experimental data in the future work.

REFERENCES

[1] L. V. Wang, "Tutorial on Photoacoustic Microscopy and Computed Tomography," *IEEE Journal of Selected Topics in Quantum Electronics,* vol. 14, no. 1, pp. 171-179, 2008, doi: 10.1109/jstqe.2007.913398.

[2] L. V. Wang and S. Hu, "Photoacoustic tomography: in vivo imaging from organelles to organs," *Science,* vol. 335, no. 6075, pp. 1458-62, Mar 23 2012, doi: 10.1126/science.1216210.

[3] H. Zhong, T. Duan, H. Lan, M. Zhou, and F. Gao, "Review of Low-Cost Photoacoustic Sensing and Imaging Based on Laser Diode and Light-Emitting Diode," *Sensors (Basel),* vol. 18, no. 7, Jul 13 2018, doi: 10.3390/s18072264.

[4] M. Xu and L. V. Wang, "Universal back-projection algorithm for photoacoustic computed tomography," *Phys Rev E Stat Nonlin Soft Matter Phys,* vol. 71, no. 1 Pt 2, p. 016706, Jan 2005, doi: 10.1103/PhysRevE.71.016706.

[5] Y. Zhang, Y. Wang, and C. Zhang, "Total variation based gradient descent algorithm for sparse-view photoacoustic image reconstruction," *Ultrasonics,* vol. 52, no. 8, pp. 1046-55, Dec 2012, doi: 10.1016/j.ultras.2012.08.012.

[6] C. Huang, K. Wang, L. Nie, L. V. Wang, and M. A. Anastasio, "Full-wave iterative image reconstruction in photoacoustic tomography with acoustically inhomogeneous media," *IEEE Trans Med Imaging,* vol. 32, no. 6, pp. 1097-110, Jun 2013, doi: 10.1109/TMI.2013.2254496.

[7] J. Prakash, D. Sanny, S. K. Kalva, M. Pramanik, and P. K. Yalavarthy, "Fractional Regularization to Improve Photoacoustic Tomographic Image Reconstruction," *IEEE Trans Med Imaging,* Dec 24 2018, doi: 10.1109/TMI.2018.2889314.

[8] C. Yang, H. Lan, F. Gao, and F. Gao, "Review of deep learning for photoacoustic imaging," *Photoacoustics,* vol. 21, p. 100215, Mar 2021, doi: 10.1016/j.pacs.2020.100215.

[9] A. Hauptmann and B. Cox, "Deep Learning in Photoacoustic Tomography: Current approaches and future directions," *arXiv preprint arXiv:2009.07608,* 2020.

[10] H. Lan *et al.*, "Ki-GAN: Knowledge Infusion Generative Adversarial Network for Photoacoustic Image Reconstruction In Vivo," in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019*, (Lecture Notes in Computer Science, 2019, ch. Chapter 31, pp. 273-281.

[11] A. Hauptmann *et al.*, "Model-Based Learning for Accelerated, Limited-View 3-D Photoacoustic Tomography," *IEEE Trans Med Imaging,* vol. 37, no. 6, pp. 1382-1393, Jun 2018, doi: 10.1109/TMI.2018.2820382.

[12] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE transactions on image processing,* vol. 13, no. 4, pp. 600-612, 2004.

[13] Y. Dong, T. Görner, and S. Kunis, "An algorithm for total variation regularized photoacoustic imaging," *Advances in Computational Mathematics,* vol. 41, no. 2, pp. 423-438, 2014, doi: 10.1007/s10444-014-9364-1.

[14] B. E. Treeby and B. T. Cox, "k-Wave: MATLAB toolbox for the simulation and reconstruction of photoacoustic wave fields," *Journal of biomedical optics,* vol. 15, no. 2, pp. 021314-021314-12, 2010.

[15] A. Paszke *et al.*, "Pytorch: An imperative style, high-performance deep learning library," in *Advances in neural information processing systems*, 2019, pp. 8026-8037.

[16] J. Staal, M. D. Abramoff, M. Niemeijer, M. A. Viergever, and B. van Ginneken, "Ridge-based vessel segmentation in color images of the retina," *IEEE Trans Med Imaging,* vol. 23, no. 4, pp. 501-9, Apr 2004, doi: 10.1109/TMI.2004.825627.