Research Article Vol. 12, No. 12 / 1 Dec 2021 / Biomedical Optics Express 7835

Biomedical Optics EXPRESS

Compressed sensing for photoacoustic computed tomography based on an untrained neural network with shape prior

HENGRONG LAN,[1,2,3] JUZE ZHANG,[1] CHANGCHUN YANG,[1] AND FEI GAO[1,*]

[1] *Hybrid Imaging System Laboratory, Shanghai Engineering Research Center of Intelligent Vision and Imaging, School of Information Science and Technology, ShanghaiTech University, Shanghai 201210, China*
[2] *Chinese Academy of Sciences, Shanghai Institute of Microsystem and Information Technology, Shanghai 200050, China*
[3] *University of Chinese Academy of Sciences, Beijing 100049, China*
[*] *gaofei@shanghaitech.edu.cn*

Abstract: Photoacoustic (PA) computed tomography (PACT) shows great potential in various preclinical and clinical applications. A great number of measurements are the premise that obtains a high-quality image, which implies a low imaging rate or a high system cost. The artifacts or sidelobes could pollute the image if we decrease the number of measured channels or limit the detected view. In this paper, a novel compressed sensing method for PACT using an untrained neural network is proposed, which decreases a half number of the measured channels and recovers enough details. This method uses a neural network to reconstruct without the requirement for any additional learning based on the deep image prior. The model can reconstruct the image only using a few detections with gradient descent. As an unlearned strategy, our method can cooperate with other existing regularization, and further improve the quality. In addition, we introduce a shape prior to easily converge the model to the image. We verify the feasibility of untrained network-based compressed sensing in PA image reconstruction and compare this method with a conventional method using total variation minimization. The experimental results show that our proposed method outperforms 32.72% (SSIM) with the traditional compressed sensing method in the same regularization. It could dramatically reduce the requirement for the number of transducers, by sparsely sampling the raw PA data, and improve the quality of PA image significantly.

© 2021 Optical Society of America under the terms of the OSA Open Access Publishing Agreement

1. Introduction

Photoacoustic imaging (PAI) is a hybrid imaging modality, which originates from the principle of photoacoustic (PA) effect [1–4]. The PA signal is induced by a short-pulsed laser light, which propagates in medium and is detected by the ultrasonic transducers. In recent decades, PAI has enabled many interesting imaging applications including hemoglobin and oxygen saturation detection, small animal imaging, and pre-clinical cancer diagnosis [5–11]. One of the implementations of PAI is photoacoustic computed tomography (PACT), which uses unfocused light to illuminate the tissue, and detects the PA signals by a transducer array.

In PACT, the number of the detector should satisfy the Nyquist sampling theorem. However, increments of the detector will increase the cost of the system. Meanwhile, the transducer could not encircle the field of view (FOV) in some scenes, e.g., the imaging of human carotid. The under-determined setup of the PA image reconstruction is achieved in these cases.

Compressed sensing (CS) has been used to reconstruct the PA image in sparse or limited-view conditions, which could recover the signal/image under the Nyquist sampling rate [12,13]. CS leverages the sparsity of data to reconstruct the PA image based on different optimizations and

#441901 https://doi.org/10.1364/BOE.441901
Journal © 2021 Received 1 Sep 2021; revised 12 Oct 2021; accepted 12 Oct 2021; published 29 Nov 2021

Biomedical Optics EXPRESS

uses different priors to solve this inverse problem [12]. For instance, in [12], the authors firstly compared several sparse representations including Wavelets, Curvelets, Fourier domains, and total variation (TV). Z. Guo et al. adapted CS for PACT reconstruction [13]. In this work, CS method was validated in *in-vivo* experiment with Wavelet basis. A modified Curvelet basis was proposed to reconstruct the sparse data in [14]. Moreover, many applications of CS are presented to achieve one-shot imaging with a single detector [15,16].

Recently, deep learning (DL) is used to reconstruct PA image in sparse view or limited-view conditions [17,18], which learns the features of the object from numerous data. For limited-view issue, Dominik Waibel et al. established a direct scheme to reconstruct the PA image from linear array data [19]. Derek Allman et al. used VGG16 to beamform the raw data to recognize the point sources [20]. For sparse data, Andreas Hauptmann et al. combined model-based reconstruction with DL to reconstruct the sub-sampled PA image [21]. Hengrong Lan et al. proposed Ki-GAN to validate the sparse PA reconstruction [22]. Most post-processing schemes are designed to solve the limited-view or sparse data issues since the degeneration of the quality is significant in these cases. Examples include that Stephan Antholzer et al. used U-Net with residual connection to enhance sparse PA image [23]. Neda Davoudi et al. used U-Net to post-process the sparse PA image using experimental data [24]. Steven Guan et al. proposed a FDU-Net to remove the artifacts of reconstructed image with 10, 15, and 30 detectors [25]. In [26], the authors proposed AS-Net to achieve superior results with sparse data. However, these methods need many paired data to pre-train the model. Namely, DL methods require the training of models with an enormous amount of data. It remains significantly more challenging for PAI since it is hard to acquire a large number of data at its infant stage. Moreover, the trained model has difficulty in generalizing for different data.

Inspired by [27], Deep Image Prior (DIP) are used to resolve inverse problem with an untrained convolutional neural networks (CNN) in medical image [28,29]. Recently, it has been used for CS [30]. In this paper, for the first time, we develop and investigate the potential of such an approach for sparse PACT reconstruction, which can recover a high-quality image with only 50% number of sensors. The additional regularization used in CS can also be used in our method. (TV is demonstrated in this paper.) Furthermore, we introduce a shape prior that penalizes the difference between the output of the model with direct reconstruction. Simulation and experimental data are used to demonstrate this method. The results show that the proposed method outperforms conventional CS with TV prior. This method bridges the gap between two PA reconstruction schemes: DL-based CS reconstruction and model-based priors method. Meanwhile, it could be combined with other conventional CS methods.

We list our contributions as follow:

1. We introduce a CNN model to reconstruct PA image from a few random noise inputs. A CS problem is adapted to an untrained model optimization to approximate the sparse PA signals. This method has the superiority of not requiring a CNN model trained over the dataset and addresses the challenge of deep learning methods for the requirement of training dataset.

2. A shape prior is proposed that empirically guides the direction of convergence at the initial iterations. The prior is restricted by the direct reconstructed image. At the first stage, the network initially fits the object; and then, the loss could decrease to fit the artifacts and noise based on empirical DIP. Therefore, the shape prior boosts the model to fit the object before overfitting to artifacts.

3. To implement DIP method in PACT, we decompose the gradient computing process into analytic gradient calculation (the forward and the adjoint operation) and chained gradient calculation (the CNN model). And then, two parts of gradients will be integrated into back-propagation.

Biomedical Optics EXPRESS

4. We demonstrate this method with conventional regularization (TV regularization in this paper). Simulated and experimental results show that our method outperforms conventional unlearned optimization method with the same regularization. Furthermore, our method embodies a robust and shows generalized performance on different data. Other CS methods are also suitable for integrating into our method, not just the TV regularization.

2. Background

2.1. Photoacoustic imaging

In PAI,the initial pressure is excited by a single short laser pulse, which can be expressed as [2]:

$$p_0 = \Gamma_0 \eta_{th} \mu_a F, \tag{1}$$

where Γ_0 is the Gruneisen coefficient, η_{th} is the conversion efficiency from light to heat, μ_a is the optical absorption coefficient, and F is the optical fluence. The pressure propagation in the medium can be described by below equation:

$$(\nabla^2 - \frac{1}{v_s^2}\frac{\partial^2}{\partial t^2})p(r,t) = -\frac{\beta}{C_P}\frac{\partial H(r,t)}{\partial t}, \tag{2}$$

where $p(r,t)$ is the spatiotemporal pressure wave, v_s is the speed of sound, H denotes the heating function, β denotes the thermal coefficient of volume expansion, and C_P denotes the specific heat capacity at constant pressure. To compute PA pressure in any heterogeneous medium, we solve this equation with Green function [2], and derives the forward model:

$$p(\mathbf{r},t) = \frac{1}{4\pi v_s^2}\frac{\partial}{\partial t}\Big[\frac{1}{v_s t}\int d\mathbf{r}' p_0(\mathbf{r}')\delta(t - \frac{|\mathbf{r}-\mathbf{r}'|}{v_s})\Big], \tag{3}$$

where $p_0(\mathbf{r}')$ is the initial pressure at detection position r'.We use a linear operator $\mathcal{A}$ indicates the forward procedure from initial pressure distribution f to the PA signals b:

$$b = \mathcal{A}f + \epsilon, \tag{4}$$

where ϵ is noise.

The light uniformly illuminates the whole target in PACT, which excites the PA signals simultaneously. The transducer array is used to receive the PA data at different positions. The inversion of Eq. (3) can be described from $p(\mathbf{r},t)$ to $p_0(\mathbf{r})$ using universal back-projection (UPB) operation [31]:

$$p_0(\mathbf{r}) = \frac{1}{\Omega_0}\int_{S_0}\Big[2p(\mathbf{r}_0,t) - \frac{2t\partial p(\mathbf{r}_0,t)}{\partial t}\Big]\frac{cos\theta_0}{|\mathbf{r}-\mathbf{r}_0|^2}dS_0, \tag{5}$$

where θ_0 is the angle between the vector pointing to the reconstruction point $\mathbf{r}$ and transducer surface S_0.

2.2. Compressed sensing

In CS, we use Ψ as a proper sparsity transform that results in an overdetermined representation, and the sparsity transform can be represented as:

$$g = \Psi f, \tag{6}$$

where $f \in \mathbb{R}^n$ is original data, and $g \in \mathbb{R}^N$ is coefficient on basis Ψ.

Biomedical Optics EXPRESS

We can project data f into a series of sensing vector b with noise ϵ, and represents compressive measurements as:

$$b = \Phi f + e, \tag{7}$$

we assume it is deterministic noise $\|e\|_2 \leq \epsilon$. We can formulate the original data f obtained by solving the following basis pursuit denoising problem:

$$\min_f \ \|\Psi f\|_1 \ \text{ s.t.} \|\Phi f - b\|_2 \leq \epsilon. \tag{8}$$

Two conditions should be satisfied if we can successfully recover the ground-truth data f_0:

- f_0 is structured: $\|f_0\|_0 \ll N$;
- The two basis sets Ψ and Φ should be incoherent.

In the CS theory, we should find a basis Ψ that sparsely represents f, and minimize the l_1 norm of Ψf promotes sparsity, and the constraint enforces data consistency. In CS-based PACT, a one-step scheme is described to solve the following minimization problem:

$$f^* \longleftarrow \arg\min_f \frac{1}{2}\|\Phi\mathcal{A}f - b\|_2^2 + \lambda \ \|\Psi f\|_1. \tag{9}$$

3. Methods

3.1. Untrained CNN for CS PACT

Given the measured PA signals b and the measurement matrix $\mathcal{M}$ ($\mathcal{M} = \Phi\mathcal{A}$), we has $b = \mathcal{M}f + \epsilon$. We aim to recover f from b:

$$f^* = \arg\min_f \frac{1}{2}\|\mathcal{M}f - b\|_2^2 + \lambda\mathcal{R}(f), \tag{10}$$

in which $\|\mathcal{M}f - b\|_2^2$ is the data consistency term, and $\mathcal{R}(f)$ is the regularization term. In CS, the optimal solution of Eq. (10) depends on $\mathcal{R}(f)$. Namely, we should find a sparse basis as the embedded prior. Some sparse basis has been mentioned before (TV, Wavelets, Curvelets, Fourier), and many papers have studied the use of CNN as NETT regularization for CS-PAT [32,33]. However, a large number of the dataset are required to train the model.

Recently, DIP exposed that a generator model is sufficient to capture a great deal of natural images prior without any learning, which can also be used to recover the compressed signal. In our work, we aim to find a set of weights for the output of CNN to fit the reconstructed image, which is applied to the measurement matrix $\mathcal{M}$ by matching the given sparse measured data b. To implement that, we should design an over-parametrized CNN decoder model D.

Therefore, we initialize the untrained model $\mathcal{M}D(\Theta, z)$ with a fixed random matrix z, and solve the non-linear least squares solution:

$$\Theta^* = \arg\min_\Theta \frac{1}{2}\|\mathcal{M}D(\Theta, z) - b\|_2^2, \tag{11}$$

Generally, the over-parameterized deep decoder D can fit any image f^*, including unstructured noise. Furthermore, an implicit prior can be expressed if we stop the procedure at the correct stage. Namely, further regularization is unnecessary if the procedure of optimization can be early stopped. Also, we can retain the sparse basis as the regularization term like the model-based optimization:

$$\Theta^* = \arg\min_\Theta \frac{1}{2}\|\mathcal{M}D(\Theta, z) - b\|_2^2 + \lambda\mathcal{R}(D(\Theta, z)). \tag{12}$$

In this work, a convolutional decoder, D, is used as the generator network, and the architecture will be described in the next section. These CNN models can provide a satisfied prior for natural

Biomedical Optics EXPRESS

images in problems such as inpainting and denoising due to their convolutional operations. Therefore, this approach also applies to another differentiable forward operator $\mathcal{A}$, not only PA forward operator.

Note that our method is a learning-free method since it has not the training phase with much data. Meanwhile, this method leverages an untrained generative decoder to optimize the weights Θ of the model. By using different models, our results further support a hypothesis that network structure, not representation learning, is the key component in image reconstruction.

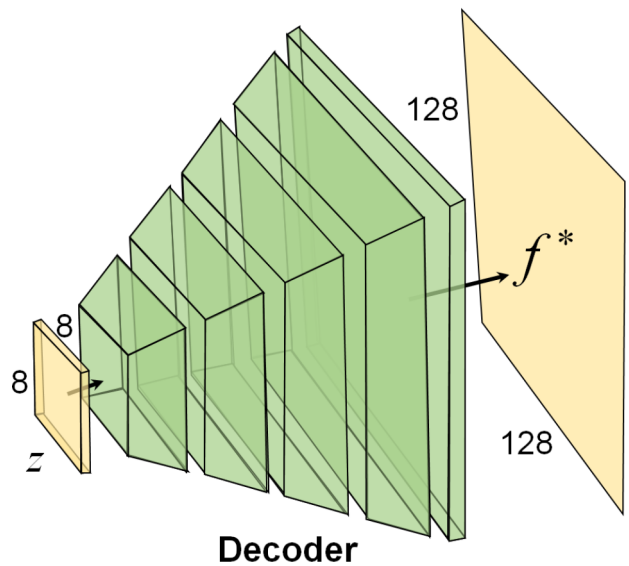

Fig. 1. The architecture of proposed decoder D, and the detail of each layer has been shown in Eq. (12). The input z is a random generated matrix with 8×8 size, and output is the fitted image with 128×128 in our work.

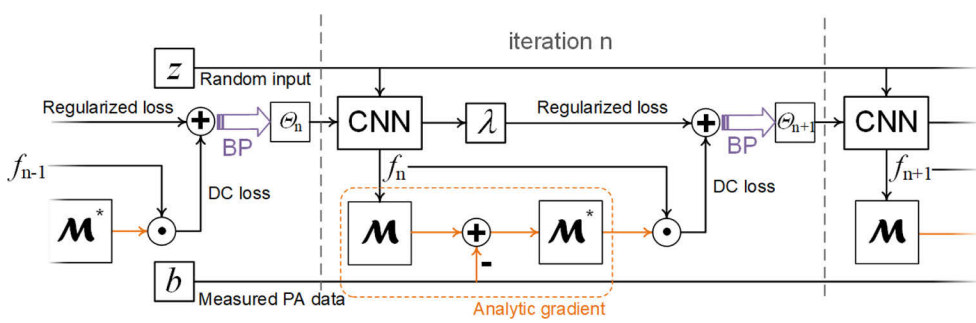

Fig. 2. The procedure of optimization of our method. For each iteration, the data consistency loss should be firstly calculated. Then, the other gradients of loss function are computed to back-propagate the gradient. The CNN indicates decoder model D in this paper, $\mathcal{M}$ contains $\mathcal{A}$ and Φ. BP: back-propagation; DC: data consistency.

3.2. Network architecture

To demonstrate our method, we introduce a CNN that generates an image through convolutions and non-linear operations. Given a random fixed input z, we use a decoder D to generate the final

Biomedical Optics EXPRESS

PA image. For l layers' decoder(in our work, l=5), the model can be defined as:

$$D(z) = \mathrm{ReLU}(\mathrm{BN}(\mathrm{Up}_l(B_l(z)) * \kappa_{3\times3})) * \kappa_{1\times1} \tag{13}$$

where

$$\begin{gathered} B_i = \mathrm{ReLU}(\mathrm{BN}(\mathrm{ReLU}(\mathrm{BN}(B_{i-1} * \kappa_i^1))) * \kappa_i^2) \\ i = 1, \ldots, l, B_0 = z. \end{gathered} \tag{14}$$

Herein, ReLU (Rectified Linear Unit) is an activation function, BN means the batch normalization operation. κ is the convolutional kernel, and κ_i is 3×3 size. Note that B_i contains the coefficient of the convolutional layer.

As mentioned above, a five-layer decoder D is used to fit the initial pressure f^*, and D is implemented here by Eq. (13) as shown in Fig. 1. This architecture is a U-Net [34] without the encoder and skipped connection. In this paper, the input data z is a Gaussian random matrix with 8×8 size, which should be fixed in the procedure of optimization. A decoder model generates an image with 128×128 size through five convolutional layers and de-convolution. For each layer, double combinations of convolution with 3×3 kernel size, BN, and ReLU are used in series and followed by a de-convolution to up-sample the feature map.

The output image, multiplied by the matrix $\mathcal{M}$, should be restricted by measured raw PA data b. Namely, we can directly optimize this model by minimizing the data consistency (DC) loss function:

$$\mathcal{L}_{DC}(\Theta) = \frac{1}{2}\|\mathcal{M}D(\Theta, z) - b\|_2^2, \tag{15}$$

3.3. Shape prior

In CS-based PACT, different sparse basis is used. For instance, TV regularization can enforce smoothness as $\mathcal{R}(f)$ in CS theory. Furthermore, the l_1 norm of TV regularization can suppress the small coefficients, whose solution can be sparse. The TV regularization can be described that:

$$\mathrm{TV}(f) = \|\nabla f\|_1, \tag{16}$$

In our method, TV regularization also penalizes the output of the decoder. Therefore, an additional TV term can be contained to restrain the deep generative model, i.e., $\mathrm{TV}(D(\Theta, z))$. Furthermore, this scheme has the advantage that we do not consider the differentiability of the regularization term (l_1 norm) since we optimize the loss function by back-propagation and gradient descent (GD). Now, we rewrite the loss function based on Eq. (12) and Eq. (16):

$$\mathcal{L}(\Theta) = \frac{1}{2}\|\mathcal{M}D(\Theta, z) - b\|_2^2 + \lambda \mathrm{TV}(D(\Theta, z)). \tag{17}$$

We can iterate this procedure and update the weight with GD. The solo data could cause the stochastic direction of the gradient. We further propose a shape prior to improve the performance and create a robust, efficient objective function. Considering that a direct texture of the target could provide a guided optimization at the beginning phase, we calculate the error between the rough image and output of D as shape prior (SP).

In our work, shape prior is proposed to penalize the output of the model, and the rough texture is created by sparse conventional reconstruction. Therefore, we estimate the prior with the decoder model by minimizing the least squares loss $\|D(\Theta, z) - f_d\|_2^2$, and f_d comes from the conventional reconstruction. We combine this prior with the loss function in Eq. (17). Finally,

Biomedical Optics EXPRESS

we optimize the weights of the Decoder model by minimizing the final loss function as follow :

$$\begin{aligned}\mathcal{L}_{final}(\Theta) &= \frac{1}{2}\|\mathcal{M}D(\Theta, z) - b\|_2^2 \\ &+ \lambda_1 \mathrm{TV}(D(\Theta, z)) + \lambda_2 \frac{1}{2}\|D(\Theta, z) - f_d\|_2^2,\end{aligned} \tag{18}$$

λ_1 and λ_2 are hyperparameters, which are chosen for different data empirically. f_d is a direct reconstruction from UBP [31] or time-reversal [35]. Note that, for PACT, this function cannot optimize the model directly since the gradient of the data consistency term cannot be tracked completely with the chain rule. We introduce a decomposed gradient descent to resolve this problem, which will be described in the next section.

Moreover, our main result shows that the estimate Θ, obtained by running gradient descent on the loss until convergence, yields an output $D(\Theta, z)$ which is close to f^*, i.e., $D(\Theta, z) \approx f^*$.

3.4. Implementation

Generally, the proximal gradient method is used to solve TV minimization in Eq. (10):

$$f^{n+1} = \mathrm{prox}_{\mathcal{R},\alpha}\Big(f^n - \alpha\mathcal{A}^*\Phi^T(\Phi\mathcal{A}f^n - b)\Big), \tag{19}$$

which is a classical and widely-used solution for TV minimization [36,37]. The forward and adjoint operator $\mathcal{A}$ and $\mathcal{A}^*$ are used to compute the gradient of the data consistency, which has been implemented in the MATLAB toolbox k-Wave [38]. For DL model, we do not consider the analytic gradient of loss function since GD and back-propagation are used to update the weight at each iteration. However, in our work, the data consistency contains $\mathcal{A}$ and D. Namely, $\mathcal{A}$ and $\mathcal{A}^*$ cannot be back-propagated, and the gradient of D cannot be calculated directly.

On the other hand, the forward operator can be discretized and written as a matrix, which is limited by computing resources. The matrix-related gradient cannot be tracked since the size of the matrix is large. Therefore, no matter whether we use the function or matrix, we cannot directly update the weights by back-propagation. The key solution is to decompose the gradient calculation of forward operation and DL model.

To decouple the data consistency term, we compute the gradient of Eq. (15) for D:

$$\frac{\partial \mathcal{L}_{DC}(\Theta)}{\partial D} = \mathcal{M}^*\big(\mathcal{M}D(\Theta, z) - b\big), \tag{20}$$

which can be calculated by k-Wave based on the output of D. We rewrite the derivative of Eq. (15) for Θ based on chain rule:

$$\frac{\partial \mathcal{L}_{DC}(\Theta)}{\partial \Theta} = \frac{\partial \mathcal{L}(\Theta)}{\partial D}\frac{\partial D}{\partial \Theta}. \tag{22}$$

For $\partial D/\partial\Theta$, the weights automatically optimize based on the chain of the gradient. Therefore, the gradient is decomposed into two terms, the first term can be computed by Eq. (20), and the second term can be updated by back-propagation. To transfer the gradient of the first term, we multiply these two terms and update the weight of DL model by back-propagation. Thus, the gradient of the data consistency term can be transferred to Θ. The procedure of this optimization can be described in Algorithm 1. We can calculate this loss and optimize the weights by back-propagation. We decompose the procedure as back-propagation and analytic gradient descent. For each iteration, the data consistency loss (lines 3 in Algorithm 1) is calculated based on the output of D since $\partial D/\partial\Theta$ can be regarded as a constant for Θ. Next, this loss needs to be combined with other regularizations to form the final loss. Finally, the back-propagation is used to update the weights. In Fig. 2, we further illustrate the pipeline of this optimization, and CNN is our decoder model D in this paper.

Biomedical Optics EXPRESS

Algorithm 1 The untrained CNN reconstructs the CS-PACT image

Input: The measurement matrix, $\mathcal{M}(\mathcal{M} = \mathcal{A}\Phi)$, the direct reconstructed image, f_d, the measured PA data, b, the hyperparameters, λ_1 and λ_2.

Output: The final reconstructed image, f^*.

Initialize a random input of model z, and the weights of model Θ_0;
for $i = 1$ to n **do**
  $\mathcal{L}_{DC}(\Theta_{i-1}) = \left[\mathcal{M}^*\left(\mathcal{M}D(\Theta_{i-1}, z) - b\right)\right]D(\Theta_{i-1}, z)$;
  $\mathcal{L}_{TV}(\Theta_{i-1}) = \mathrm{TV}(D(\Theta_{i-1}, z))$;
  $\mathcal{L}_{SP}(\Theta_{i-1}) = 0.5 \times \|D(\Theta_{i-1}, z) - f_d\|_2^2$;
  $\mathcal{L}_{final}(\Theta_{i-1}) = \mathcal{L}_{DC}(\Theta_{i-1}) + \lambda_1\mathcal{L}_{TV}(\Theta_{i-1}) + \lambda_2\mathcal{L}_{SP}(\Theta_{i-1})$;
  $\Theta_i \leftarrow \mathrm{BP}\left(\mathcal{L}_{final}(\Theta_{i-1})\right)$; // Update Θ using back-propagation;
end for
$f^* = D(\Theta_n, z)$.

In this paper, the optimizer of all experiments is RMSRrop, and the size of output image is 128×128 . We implement this algorithm including $\mathcal{M}$, $\mathcal{M}^*$, and the framework $D(\Theta, z)$ by MATLAB. The initial step rate is 0.001. All methods are implemented on a 64-bit operating system with an Intel Core i7-6700 CPU and an NVIDIA GTX 1080 Ti GPU.

4. Experiments

To experimentally validate our approach, simulation data and *in-vivo* data are used. Furthermore, we compare our method with other methods. To validate CS-based PACT, the data is 128 channels, and a 50% random sub-sampling matrix is used to sub-sample the channel number.

Conventional method with TV norm is compared with our method, which leverages Eq. (19) to solve the objective function. Furthermore, we also compare our method with conventional Tikhonov regularization. Since our method is unlearned, we only compare it to other unlearned methods. Meanwhile, some comparison experiments are demonstrated, e.g., different regularization. Moreover, through experiments, we illustrate the effects of the priors and determine the appropriate number of iterations.

4.1. Synthetic setup

For the synthetic data, we use k-Wave to generate the data. 128 elements circular ultrasound (US) array receives the PA signals with 14.5mm radius. The pixel number of the initial pressure map is 380×380, and the total grid size is 30mm $\times$ 30mm. The sampling rate of PA signal is 40 MSa/s, and noise is added to signal with 40 dB SNR. The center frequency of the US transducer is set as 2.5 MHz with 80% bandwidth, and the speed of sound is 1500 m/s. The reconstructed region is 30 mm $\times$ 30 mm with 128×128 pixels.

4.2. In-vivo data

Moreover, we also compare our method with the conventional method on the *in-vivo* data, which contains the brain of mice and the cross-section of the human finger. All data is acquired from a self-built PACT system in Fig. 3. The transducer array is a customized 128-elements full-view circular with 30mm radius (2.5 MHz, Doppler Inc.), which is placed in a 3-D printed water tank. The laser source is a pulsed laser (720 nm wavelength, 10 Hz repetition rate), which illuminates the object by a fiber optic bundle as Fig. 3 shows. The data sampling rate of the data acquisition system is 40 MSa/s. The region of image reconstruction is 30 mm $\times$ 30 mm with 128×128 pixels.

Biomedical Optics EXPRESS

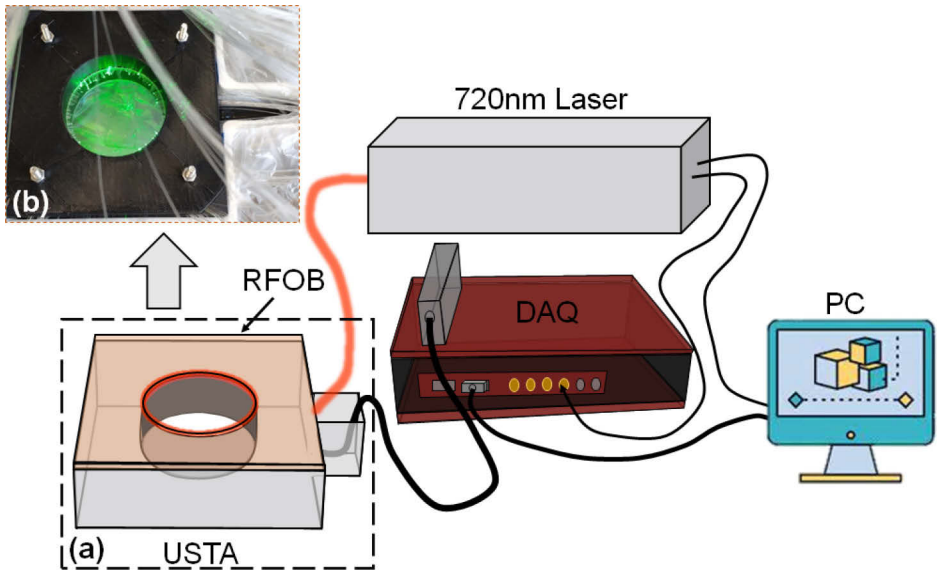

Fig. 3. Schematics of the PACT system. The black box region (a) has a detailed photograph in (b). RFOB: ring-shaped fiber optics bundle; USTA: ultrasonic transducer array; DAQ: data acquisition system; PC: personal computer.

5. Results

5.1. Synthetic results

We firstly validate our method on the synthetic data, λ_1 and λ_2 are 0.006 and 0.05 respectively. In Fig. 4, we show the results of TV method and our method. Specially, our method is minimized by Eq. (18), which refers to this function by default in this paper. This holds by simply running TV method until convergence (300 iterations). All methods are implemented on a system with Intel i7-7600 processor with 3.40 GHz, 32 GB memory, and a GTX 1080Ti graphics processing unit. For each iteration, the conventional TV costs 0.23s, our proposed method costs 0.29s and the Tikhonov method costs 0.07s. Note that, for all experiments, the number of iterations is 700, which is terminated by pre-running different iterations. Moreover, the procedure of optimization can automatically stop when the value of metrics starts to decline if we use an additional metrics of quality. Due to the reduction in the number of channels, the background of the conventional result is polluted, which causes a poor contrast compared with Fig. 4. For our approach, most structures of the object are reconstructed well with few artifact. It shows that the decoder D fits the object at the initial phase, and the artifacts are reconstructed after continuously optimizing. For Tikhonov method, it is less sensitive to edges compared with TV, which causes blurry edges for artifacts. We can compute the Structural Similarity (SSIM) of these results to quantitatively compare the performance. The SSIM values of Fig. 4 are 0.6312, 0.8377 and 0.3618 respectively, which also indicates our method outperforms the conventional TV method over 32%.

Furthermore, we should validate the effects of different priors and the appropriate iteration times. A series of comparison experiments are set up.

5.1.1. Iteration times

We use an untrained model D to compare the performance of different numbers of iterations without any regularization. Different iteration times have been validated from 100 to 8000 as Fig. 5 shows. As the number of iterations increases, the main object is reconstructed firstly

Biomedical Optics EXPRESS

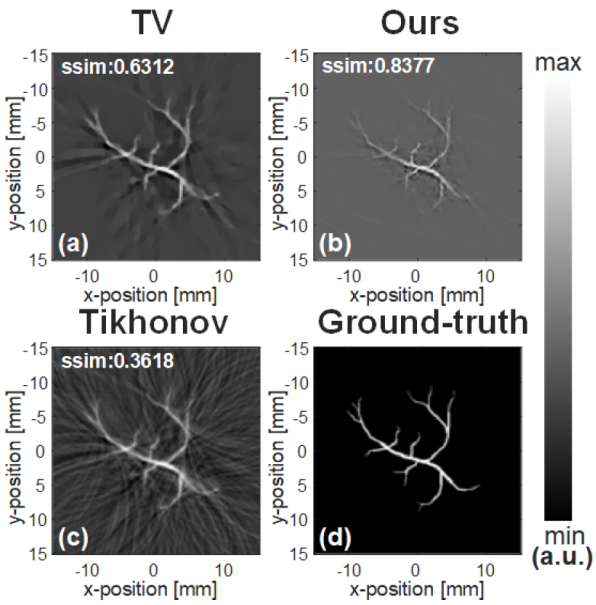

Fig. 4. The synthetic vessel result (50% sub-sampling rate), (a) the conventional TV method; (b) our approach with TV prior; (c) the conventional Tikhonov method; (d) the ground-truth of generated image

(from 1 to 500), and the best reconstruction is achieved between 500 and 1000. And then, the reconstruction result starts to blur (after 1000) since the artifacts near the object are appearing. It also indicates our model will first fit the major structure and then fit the noise and the artifacts. We should appropriately stop the iteration to obtain a better result [27,30]. We should further quantitatively evaluate these results.

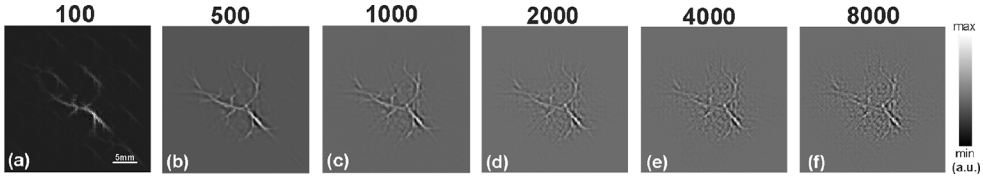

Fig. 5. The generated PA images at different iterations, without any regularization.

Three metrics are used to quantify the performance of different results, which are SSIM, Peak Signal to Noise Ratio (PSNR), and Signal to Noise Ratio (SNR). Table 1 demonstrates the quantitative results of these different iterations. The quantitative results show that the reconstruction quality first increases and then decreases with the number of iterations increasing. Namely, the model fits the correct object at the beginning, and the best quality is 500 iterations in Table 1. Therefore, for PACT, the best iterative times could be selected from 100 to 1000. After comparing different iterations, we determined to use 700 iterations for all experiments without unnecessary artifacts.

5.1.2. Comparison experiments

For DIP, the DL model can express the implicit prior generally, thus the additional prior term. In this section, we evaluate the effects of two priors (TV and SP). The two parameters are same with before. The synthetic vessel results have been shown in Fig. 6. The two results show similar

Biomedical Optics EXPRESS

Table 1. The quantitative results of different iteration

Iterations	100	500	1000	2000	4000	8000
SSIM	0.7418	0.8236	0.7958	0.7591	0.7130	0.6405
PSNR	21.2158	21.3040	21.1562	20.8575	20.6561	20.5045
SNR	1.7469	1.8350	1.6873	1.3886	1.1872	1.0356

reconstructions from Fig. 6. The background of D's result has some noises as the yellow arrows indicated in Fig. 6(a). By contrast, Fig. 4 has a purer background and higher contrast compared with Fig. 6. To further evaluate the contrast of these results, we can compute the contrast-to-noise rate (CNR) for these results.

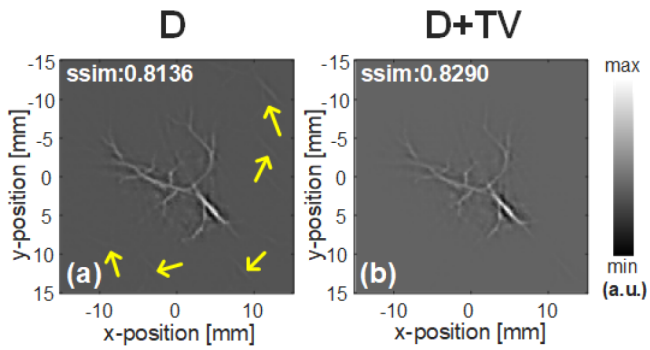

Fig. 6. Comparison experimental results, (a) the vessel result only using network $D(\Theta, z)$; (b) the vessel result using network $D(\Theta, z)$ and TV prior (without SP). The yellow arrows indicate some noises in background.

We list the quantitative results of Fig. 4 and Fig. 6 in Table 2. The first three columns of the table indicate that each of the two priors items has improved the reconstructed quality. Although the conventional sparse basis can be surpassed only using a deep model, different regularization can further boost the robustness and efficiency of this method. Compared with the decoder D without regularization terms, the decoder D with regularization performs better in terms of noise suppression, i.e., higher SNR (3.1046 and 1.6595). Similarly, the results of the quantitative comparisons also reflect the same performance from Table 2, and our method has higher contrast compared with others. These results further show that effective priors can improve the performance of untrained CNN.

Table 2. The quantitative results of comparison experiments

	D	D+TV	D+TV+SP	TV	Tikhonov
SSIM	0.8136	0.8290	**0.8377**	0.6312	0.3618
PSNR (dB)	21.1285	21.7119	**22.5736**	18.1459	15.1819
SNR (dB)	1.6595	1.4430	**3.1046**	-1.323	-4.6120
CNR (dB)	-5.9295	-4.8350	**-2.2060**	-9.6199	-63.5153

5.2. *In-vivo results*

In addition, we demonstrate our method on *in-vivo* data, λ_1 and λ_2 are 0.005 and 0.1, respectively. Firstly, a real mice brain data is validated. We also compare TV method with our method. Figure 7(a)-(d) shows the brain imaging results, where the TV obtains the result with 300

Biomedical Optics EXPRESS

iterations. To further evaluate the results, we also demonstrate the full-view results for all *in-vivo* results, which use TV method to reconstruct the image with 128 channels full-view PA data. Obviously, the conventional method cannot suppress the artifacts only using 64 channels data from Fig. 7(a). The untrained CNN model method shows a superior result, purer background creates a higher contrast in Fig. 7(b). The yellow arrows in Fig. 7(a), (b) show that the vessel in the sulcus has a clear shape compared with TV result, which contains a few artifacts in Fig. 7(a). For Fig. 7, the Tikhonov regularization is not sensitive for artifacts in this sparse condition compared with TV. Note that the required number of detectors N is related to the size of region of interest and the acoustic wavelength ($N\lambda/2 = \pi D$) to satisfy Nyquist criteria [8]. For this setup (D=30mm, λ=400 μm), the number of detectors should be greater than 470. Although full-view result is formed from 128 channel data, the number of detectors still cannot satisfy the Nyquist criteria in this work. Therefore, the full-view result still has artifacts as Fig. 7(d) and 7(h) shown. For the background, our method even outperforms the full-view result (Fig. 7(d)) in the way of artifacts. Since the artifacts still exist in the full-view results, we do not further calculate the quantitative metrics for these full-view results.

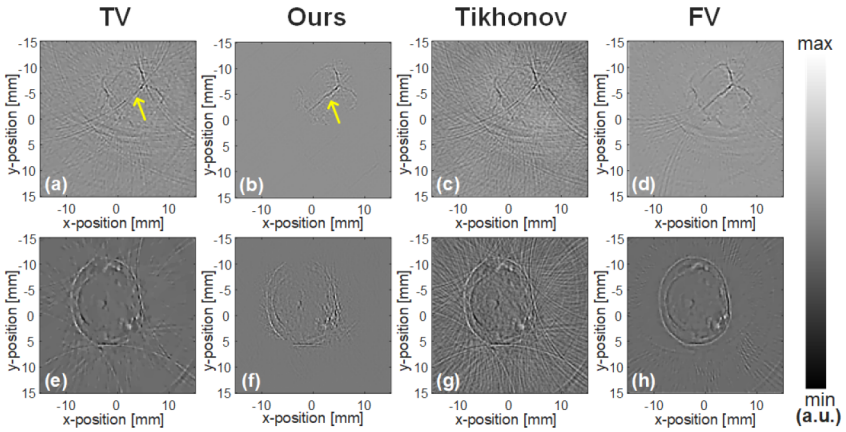

Fig. 7. The *in-vivo* results with 50% sub-sampling rate, (a)-(d) the results of mice brain, (e)-(h) the results of cross section of finger. (a), (e) the iterative TV method; (b), (f) our approach with TV prior; (c), (g) the iterative Tikhonov method; (d), (h) The full-view results, FV: full-view

We further use a cross-section of the human finger to demonstrate different methods. We also compare these two different methods in Fig. 7(e)-(h). Similarly, some artifacts are retained in the result of conventional method as shown in Fig. 7(e),(g). The 50% sub-sampling rate, i.e., 64 channels, causes a blurry result showing that the object is disturbed by the artifacts. For our result, Fig. 7(f) eliminates most of the obvious artifacts compared with Fig. 7(e) and (g). Note that the artifacts near the objects are also beginning to be reconstructed from Fig. 7(f). However, for the SNR and contrast, our method still outperforms the conventional method. In addition, these results further verify the merits of this method, which will first fit the target and then fit the noise and the artifacts. It implies this method can fit any signal with appropriately stopping.

6. Conclusion

In this paper, we introduce an untrained CNN model to reconstruct a sparse PACT image, which outperforms unlearned methods without plenty of data. In addition, a direct reconstructed image is used to penalize the output of DL model. This prior improves the reconstruction error and efficiency. We further demonstrate how to implement this method on PACT, which further decomposes the analytic gradient and chained gradient in the data consistency term. The

Biomedical Optics EXPRESS

experimental results show our approach outperforms the conventional CS method with the same sparse basis. Note that DIP method can fit any signal given an over-parameterized model in empirical. This method provides insight for CS-based PACT, and explores more solid works combined with other conventional CS methods. Meanwhile, we will compare another sparse basis in future works.

Funding. National Natural Science Foundation of China (61805139); United Imaging Intelligence (2019X0203-501-02).

Disclosures. The authors declare no conflicts of interests.

Data availability. Data underlying the results presented in this paper are not publicly available at this time but may be obtained from the authors upon reasonable request.

References

1. Y. Zhou, J. Yao, and L. V. Wang, "Tutorial on photoacoustic tomography," J. Biomed. Opt. **21**(6), 061007 (2016).
2. L. V. Wang, "Tutorial on photoacoustic microscopy and computed tomography," IEEE J. Sel. Top. Quantum Electron. **14**(1), 171–179 (2008).
3. L. V. Wang and S. Hu, "Photoacoustic tomography: in vivo imaging from organelles to organs," Science **335**(6075), 1458–1462 (2012).
4. L. V. Wang and J. Yao, "A practical guide to photoacoustic tomography in the life sciences," Nat. Methods **13**(8), 627–638 (2016).
5. J. Shah, S. Park, S. R. Aglyamov, T. Larson, L. Ma, K. V. Sokolov, K. P. Johnston, T. E. Milner, and S. Y. Emelianov, "Photoacoustic imaging and temperature measurement for photothermal cancer therapy," J. Biomed. Opt. **13**(3), 034024 (2008).
6. H. Lan, T. Duan, D. Jiang, H. Zhong, M. Zhou, and F. Gao, "Dual-contrast nonlinear photoacoustic sensing and imaging based on single high-repetition-rate pulsed laser," IEEE Sens. J. **19**(14), 5559–5565 (2019).
7. F. Gao, Q. Peng, X. Feng, B. Gao, and Y. Zheng, "Single-wavelength blood oxygen saturation sensing with combined optical absorption and scattering," IEEE Sens. J. **16**(7), 1943–1948 (2016).
8. L. Li, L. Zhu, C. Ma, L. Lin, J. Yao, L. Wang, K. Maslov, R. Zhang, W. Chen, J. Shi, and L. V. Wang, "Single-impulse panoramic photoacoustic computed tomography of small-animal whole-body dynamics at high spatiotemporal resolution," Nat. Biomed. Eng. **1**(5), 0071 (2017).
9. L. Lin, P. Hu, J. Shi, C. M. Appleton, K. Maslov, L. Li, R. Zhang, and L. V. Wang, "Single-breath-hold photoacoustic computed tomography of the breast," Nat. Commun. **9**(1), 2352–2359 (2018).
10. Z. Wu, F. Duan, J. Zhang, S. Li, H. Ma, and L. Nie, "In vivo dual-scale photoacoustic surveillance and assessment of burn healing," Biomed. Opt. Express **10**(7), 3425–3433 (2019).
11. J. Lv, Y. Xu, L. Xu, and L. Nie, "Quantitative functional evaluation of liver fibrosis in mice with dynamic contrast-enhanced photoacoustic imaging," Radiology **300**(1), 89–97 (2021).
12. J. Provost and F. Lesage, "The application of compressed sensing for photo-acoustic tomography," IEEE Trans. Med. Imaging **28**(4), 585–594 (2009).
13. Z. Guo, C. Li, L. Song, and L. V. Wang, "Compressed sensing in photoacoustic tomography in vivo," J. Biomed. Opt. **15**(2), 021311 (2010).
14. B. Pan, S. R. Arridge, F. Lucka, B. T. Cox, N. Huynh, P. C. Beard, E. Z. Zhang, and M. M. Betcke, "Photoacoustic reconstruction using sparsity in curvelet frame," arXiv preprint arXiv:2011.13080 (2020).
15. N. Huynh, F. Lucka, E. Z. Zhang, M. M. Betcke, S. R. Arridge, P. C. Beard, and B. T. Cox, "Single-pixel camera photoacoustic tomography," J. Biomed. Opt. **24**(12), 121907 (2019).
16. Y. Guo, B. Li, and X. Yin, "Single-shot compressed photoacoustic tomographic imaging with a single detector in a scattering medium," Phys. Rev. Appl. **13**(4), 044009 (2020).
17. C. Yang, H. Lan, F. Gao, and F. Gao, "Review of deep learning for photoacoustic imaging," Photoacoustics **21**, 100215 (2021).
18. A. Hauptmann and B. T. Cox, "Deep learning in photoacoustic tomography: Current approaches and future directions," J. Biomed. Opt. **25**(11), 112903 (2020).
19. D. Waibel, J. Gröhl, F. Isensee, 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).
20. D. Allman, A. Reiter, and M. A. L. Bell, "Photoacoustic source detection and reflection artifact removal enabled by deep learning," IEEE Trans. Med. Imaging **37**(6), 1464–1477 (2018).
21. A. Hauptmann, F. Lucka, M. Betcke, N. Huynh, J. Adler, B. Cox, P. Beard, S. Ourselin, and S. Arridge, "Model-based learning for accelerated, limited-view 3-D photoacoustic tomography," IEEE Trans. Med. Imaging **37**(6), 1382–1393 (2018).
22. H. Lan, K. Zhou, C. Yang, J. Cheng, J. Liu, S. Gao, and F. Gao, "Ki-gan: knowledge infusion generative adversarial network for photoacoustic image reconstruction in vivo," in *International conference on medical image computing and computer-assisted intervention*, (Springer, 2019), pp. 273–281.
23. S. Antholzer, M. Haltmeier, and J. Schwab, "Deep learning for photoacoustic tomography from sparse data," Inverse Probl. Sci. Eng. **27**(7), 987–1005 (2019).

Biomedical Optics EXPRESS

24. N. Davoudi, X. L. Deán-Ben, and D. Razansky, "Deep learning optoacoustic tomography with sparse data," Nature Machine Intelligence **1**(10), 453–460 (2019).
25. 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).
26. M. Guo, H. Lan, C. Yang, and F. Gao, "As-net: fast photoacoustic reconstruction with multi-feature fusion from sparse data," arXiv preprint arXiv:2101.08934 (2021).
27. D. Ulyanov, A. Vedaldi, and V. Lempitsky, "Deep image prior," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, (2018), pp. 9446–9454.
28. S. Dittmer, T. Kluth, P. Maass, and D. O. Baguer, "Regularization by architecture: a deep prior approach for inverse problems," J. Math Imaging Vis. **62**(3), 456–470 (2020).
29. D. O. Baguer, J. Leuschner, and M. Schmidt, "Computed tomography reconstruction using deep image prior and learned reconstruction methods," Inverse Problems **36**(9), 094004 (2020).
30. R. Heckel and M. Soltanolkotabi, "Compressive sensing with un-trained neural networks: gradient descent finds a smooth approximation," in *International Conference on Machine Learning*, (PMLR, 2020), pp. 4149–4158.
31. M. Xu and L. V. Wang, "Universal back-projection algorithm for photoacoustic computed tomography," Phys. Rev. E **71**(1), 016706 (2005).
32. H. Li, J. Schwab, S. Antholzer, and M. Haltmeier, "Nett: solving inverse problems with deep neural networks," Inverse Problems **36**(6), 065005 (2020).
33. S. Antholzer, J. Schwab, J. Bauer-Marschallinger, P. Burgholzer, and M. Haltmeier, "Nett regularization for compressed sensing photoacoustic tomography," Proc. SPIE **10878**, 108783B (2019).
34. O. Ronneberger, P. Fischer, and T. Brox, "U-net: convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*, (Springer, 2015), pp. 234–241.
35. B. E. Treeby, E. Z. Zhang, and B. T. Cox, "Photoacoustic tomography in absorbing acoustic media using time reversal," Inverse Problems **26**(11), 115003 (2010).
36. Y. Dong, T. Görner, and S. Kunis, "An algorithm for total variation regularized photoacoustic imaging," Adv. Comput. Math. **41**(2), 423–438 (2015).
37. 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 **32**(6), 1097–1110 (2013).
38. 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).