Check for updates

Dictionary learning sparse-sampling reconstruction method for *in-vivo* 3D photoacoustic computed tomography

FANGYAN LIU,[1,4] XIAOJING GONG,[2,4] LIHONG V. WANG,[3] JINGJING GUAN,[1] LIANG SONG,[2,5] AND JING MENG[1,6]

[1] *Qufu Normal University, School of Information Science and Engineering, 80 Yantai Road North, Rizhao, 276826, China*
[2] *Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Institute of Biomedical and Health Engineering, 1068 Xueyuan Boulevard, Shenzhen, 518055, China*
[3] *California Institute of Technology, Department of Electronic Engineering, Andrew & Peggy Cherng Department of Medical Engineering, Pasadena, CA 91125, USA*
[4] *Equal Contribution*
[5] *liang.song@siat.ac.cn*
[6] *qfmj@163.com*

Abstract: The sparse transforms currently used in the model-based reconstruction method for photoacoustic computed tomography (PACT) are predefined and they typically cannot capture the underlying features of the specific data sets adequately, thus limiting the high-quality recovery of photoacoustic images. In this work, we present an advanced reconstruction model using the K-VSD dictionary learning technique and present the *in vivo* results after adapting the model into the 3D PACT system. The *in vivo* experiments were performed on an IRB approved human hand and two rats. When compared to the traditional sparse transform, experimental results using our proposed method improved accuracy and contrast to noise ration of the reconstructed photoacoustic images, on average, by 3.7 and 1.8 times in the case of 50% sparse-sampling rate, respectively. We also compared the performance of our algorithm against other techniques, and imaging speed was 60% faster than other approaches. Our system would require sparse-transducer array and lower number of data acquisition hardware (DAQs) potentially reducing the cost of the system. Thus, our work provides a new way for reconstructing photoacoustic images, and it would enable the development of new high-speed low-cost 3D PACT for various biomedical applications.

1. Introduction

Photoacoustic computed tomography (PACT) is a major modality for photoacoustic imaging (PAI) [1]. It has the common advantages of optical imaging, such as non-invasive, non-ionizing, high-contrast, and ultrasound imaging, such as high-resolution. In addition, PACT has the multi-scale and multi-resolution imaging abilities for large field-of-vision (FOV) and deep tissues. This imaging modality has been used successfully in many biomedical imaging fields, such as the whole-body imaging of small animals and the early diagnosis of breast cancer [1,2]. More recently, the application of PACT for the detection of vulnerable plaques in the cardiovascular system is also being explored [3,4]. In PACT, the ultrasound signals emitted from the biological tissues are collected using unfocused ultrasound transducers, and then a reconstruction algorithm is employed on these received signals to form the optical-absorption image. The transducer usually contains several hundred densely packed transducer elements to achieve high-quality photoacoustic images. As a result, the cost of fabricating ultrasonic array is very high and increases significantly when high-frequency transducer elements are used [5]. In addition, since the data acquisition card has a limited number of channels, the multiplexing technique is usually used to combine information coming from a

large number of channels available on the transducer. The multiplexing technique leads to long data acquisition time for one B-scan photoacoustic image and multiple laser illuminations may be required. The increased time in forming an image could influence the diagnosing of the disease in clinical applications.

Sparse sampling is an effective way in PACT to decrease the fabrication costs of the ultrasonic array and to reduce the data acquisition time. However, the recovered photoacoustic images with sparse sampling using the traditional reconstruction method, such as the back-projection method (BP), will degrade if reconstruction's technique does not address sparse sampling [6]. Several efforts have been made to improve the image quality of sparse-sampling in PACT [7–9]. The compressed sensing (CS) is a typical model-based technique that incorporates sparsity prior of signals to improve the reconstruction of sparse-sampling. For example, Guo et al. implemented the CS-based photoacoustic imaging of rat brain and subcutaneous blood vessels in vivo [10]; Song et al. proposed an advanced CS reconstruction model with partially known support for acoustic- and optical-resolution PACT [11–13]. The typical predefined sparse representations used in reconstruction model are, Wavelet, Fourier and total variation (TV) transforms. Although these transforms have the advantages of simplicity with known properties, they may not capture all features of the specific image sequence adequately in sparse sampling and could lead to limited artifacts suppression. Recently, our research group developed a principal component analysis (PCA) method for 3D sparse-sampling PACT [14], in which $\sim 30 - 50\%$ fully sampled frames were required to obtain the PCA basis. This method implemented the 3D high-speed imaging of PACT without any iterative process, but the data acquisition time did not reduce significantly.

Dictionary learning (DL) is a new approach for adaptive sparse representation of signals and is being widely used in image processing, machine learning and computer vision [15–17]. In medical imaging, several researchers have explored using dictionary learning to reconstruct images. Bai et al. proposed an iterative low-dose CBCT reconstruction method based on the 3-D dictionary to effectively suppress the noise while retaining the structures of the low-dose CBCT image [18]. Wang et al. implemented a real-time dynamic MRI based on a parallel dictionary learning method [19]. Shen et al. developed a dictionary learning-based framework for PET reconstruction [20]. Novel dictionary-learning methods are also being explored in recent years to meet higher processing requirements, and some of them are used in medical imaging fields as well. Karimi et al. proposed a novel two-level structured dictionary for fast processing of 3D medical images and applied it to the 3D computed tomography (CT) restoration [21]. Jiang et al. proposed a super-resolution CT image reconstruction method based on a dictionary jointly trained on low- and high-resolution image patches [22]. Aiming to reconstruct the undersampled dynamic contrast-enhanced (DCE) MRI data, Quantm et al. proposed a data-driven CS reconstruction method based on the convolutional sparse coding algorithm [23]. Ravishankar et al. developed a novel reconstruction framework for MRI based on the sparsifying transform learning [24]. The DL technique is also explored in photoacoustic imaging. Zhou et al. applied DL to the reconstruction of sparse-sampled photoacoustic images, and the effectiveness of DL was verified by the numerical simulations [25]. Ning et al. presented a new DL-based method to remove the reverberant signal in photoacoustic microscopy (PAM) and enabled its depth-resolved imaging of cortical microvasculature in the mouse brain [26].

In this paper, we are proposing a novel DL-based advanced reconstruction algorithm with sparse sampling for in vivo 3D PACT imaging. We adopted a conventional dictionary learning method to show the feasibility of using dictionary learning methods to reconstruct 3D images in PACT. To our knowledge, this is the first time such a DL-based reconstruction method has been proposed for the PACT imaging. In this method, very few frames of the 3D imaging region ($\sim 5\%$) are required to be sampled fully to generate the dictionary, thus, reducing the data acquisition time significantly. The trained DL is then imported into imaging model to reconstruct all other sparse-sampled frames in this 3D region to reconstruct the 3D

volume. Compared to the previous model-based photoacoustic image reconstruction methods, our proposed DL based reconstruction model can recover more accurate and higher quality photoacoustic images, without changing the structure of the experimental device. *In vivo* experiments of a human hand and two rats demonstrated the superiorities of our proposed method.

2. Methods

2.1 Dictionary sparse representation

The sampling image first needs to be divided into many overlapping sub-patches to obtain a dictionary. Assuming the dimension of each patch is $\sqrt{N} * \sqrt{N}$, the sampling data matrix $\mathbf{Y} = [\mathbf{y}_i]_{i=1}^{S}$ is constructed. Where S represents the total number of patches, $\mathbf{y_i}$ is the vector form of the *ith* image patch with size N. Using this data matrix $\mathbf{Y}$, a dictionary $\mathbf{D} = [\mathbf{d}_i]_{i=1}^{K}$ is derived. Due to the overcomplete feature of the dictionary, K is usually much larger than N. Finally, the data $\mathbf{Y}$ is sparsely represented in the dictionary $\mathbf{D}$, and the sparse-coefficient matrix is represented as $\boldsymbol{\alpha} = [\boldsymbol{\alpha}_i]_{i=1}^{S}$. The above process of DL for a given sampling data can be expressed by the following equation [19]:

$$\min_{\mathbf{D},\boldsymbol{\alpha}} \sum_{i=1}^{S} \left\| \mathbf{y}_i - \mathbf{D}\boldsymbol{\alpha}_i \right\|_2^2 \qquad s.t. \ \forall i, \left\| \boldsymbol{\alpha}_i \right\|_0 \leq L_0 \tag{1}$$

where L_0 represents the sparsity constraint, $\boldsymbol{\alpha}_i$ is the sparse representation of $\mathbf{y}_i$ over $\mathbf{D}$. Equation (1) can be rewritten in the following unconstrained form:

$$\min_{\mathbf{D},\boldsymbol{\alpha}} \sum_{i=1}^{S} \left\| \mathbf{y}_i - \mathbf{D}\boldsymbol{\alpha}_i \right\|_2^2 + \mu \sum_{i=1}^{S} \left\| \boldsymbol{\alpha}_i \right\|_0 \tag{2}$$

where μ represents the penalty parameter of the sparsity. There are many advanced DL learning algorithms developed to build DL, including sparsifying transform learning, structured dictionary learning method, and convolutional dictionary learning. In this paper we chose K-SVD method since it is considered to be standard for dictionary learning and is used in a variety of applications. In addition, K-VSD method is simple to use.

2.2 Reconstruction model

Assuming the system matrix of the PACT is $\mathbf{K}$, the measurement data matrix is $\mathbf{Y}$, and the photoacoustic image to be reconstructed is $\mathbf{X}$, then the model-based reconstruction formula with sparsity constraint for PACT can be expressed as:

$$\min_{\mathbf{X}} \mathbf{F} = \left\| \mathbf{KX} - \mathbf{Y} \right\|_2^2 + \lambda \left\| \boldsymbol{\Psi}\mathbf{X} \right\|_1 \tag{3}$$

In Eq. (3), λ represents a regularization parameter and $\boldsymbol{\Psi}$ is a sparse transform. It is often useful to include a total variation (TV) item in the reconstruction model even when other sparse transforms are used in the objective function [12,13,19]. Therefore, the enhanced reconstruction model of Eq. (3) is:

$$\min_{\mathbf{X}} \mathbf{F} = \left\| \mathbf{KX} - \mathbf{Y} \right\|_2^2 + \lambda_1 \left\| \boldsymbol{\Psi}\mathbf{X} \right\|_1 + \lambda_2 \mathbf{TV}(\mathbf{X}) \tag{4}$$

where λ_1 and λ_2 represent the regularization parameters. The TV consists of two parts, row differential transform (represented by $\mathbf{T}_r$) and column differential transform (represented by $\mathbf{T}_c$). So the $\mathbf{TV}(\mathbf{X})$ in Eq. (4) can be expressed as: $\mathbf{TV}(\mathbf{X}) = \left\| \mathbf{T}_r * \mathbf{X} \right\|_1 + \left\| \mathbf{X} * \mathbf{T}_c \right\|_1$. In practical

applications, when a Wavelet transform is adopted as the sparse transform, the reconstruction model of Eq. (4) can be called as Wavelet-TV model.

An advanced reconstruction model of PACT with dictionary sparse representation and TV (DL-TV) was developed to address the sparse-sampling artifacts in the limited-view PACT:

$$\min_{\mathbf{X},\{\boldsymbol{\alpha}_j\}} \mathbf{F} = \frac{1}{2}\|\mathbf{KX} - \mathbf{Y}\|_2^2 + \frac{\lambda_1}{2}\sum_j \|R_j\mathbf{X} - \mathbf{D}\boldsymbol{\alpha}_j\|_2^2 + \mu\sum_j \|\boldsymbol{\alpha}_j\|_0 + \lambda_2 \mathbf{TV}(\mathbf{X}) \qquad (5)$$

In Eq. (5), $\mathbf{F}$ is the objective function consisting of four parts——the first part represents the square error between the estimated measurements from the reconstructed signal and the experimentally acquired measurements, the second part represents the square error between the original image and the transformed image by the dictionary, the third part represents the l_0 norm of the sparse signals in the dictionary domain, and the fourth part represents the TV penalty of the signal. R_j is the operator for extracting the *jth* sub-image patch from the original image $\mathbf{X}$. λ_1, λ_2 and μ are the regularization parameters determining the trade-off between the data fidelity and the sparsity. In here, l_0 norm is the number of non-zero elements in the sparse matrix. The optimization of Eq. (5) involves solving for two variables, $\mathbf{X}$ and $\{\alpha_j\}$. Once the dictionary $\mathbf{D}$ is learned from the training data set, we apply a commonly used method —orthogonal matching pursuit (OMP) — to compute the sparse coefficient matrix $\{\alpha_j\}$, in Eq. (5), for all patches of the image $\mathbf{X}$. Once $\mathbf{D}$ and $\{\alpha_j\}$ are determined, $\mathbf{X}$ is updated by conjugate gradient descent (CGD) algorithm. Since $\|\boldsymbol{\alpha}_j\|_0$ is a constant, the gradient computation with respect to $\mathbf{X}$ is not performed on this item [27]. Thus, the gradient computation formula for F with respect to $\mathbf{X}$ in Eq. (5) is:

$$\nabla\mathbf{F}(\mathbf{X}) = \mathbf{K^T}(\mathbf{KX} - \mathbf{Y}) + \lambda_1\sum_j R_j^{\mathrm{T}}(R_j\mathbf{X} - \mathbf{D}\boldsymbol{\alpha_j}) + \lambda_2\nabla(\mathbf{TV}(\mathbf{X})) \qquad (6)$$

Generally, the l_1 norm here is not smooth and not differentiable [28]. To implement the gradient computation on l_1 norm, a method proposed by Lustig is adopted in our paper. In this method, the absolute value of l_1 norm of vector $\mathbf{X}$ is approximated with a smooth function by using the relation $|\mathbf{X}| \approx \sqrt{\mathbf{X}*\mathbf{X} + \mu}$, where μ is a positive smoothing parameter. With this approximation, $\dfrac{d|X|}{dX} \approx \dfrac{X}{\sqrt{X*X + \mu}}$. Let W be a diagonal matrix with the diagonal elements $\mathbf{W}(\mathrm{i,i}) = \sqrt{(\Psi\mathbf{X})_i^*(\Psi\mathbf{X})_i + \mu}$, Ψ is a sparse transform. Then the gradient of l_1 norm of $\mathbf{X}$ can be computed with the formulation: $\nabla|\Psi\mathbf{X}|_1 = \Psi^*W^{-1}\Psi X$. The more detailed description can be found in Ref. 28.

To compute the gradient $\nabla\mathbf{TV}(\mathbf{X})$ in (6), two diagonal matrixes were constructed. They are $\mathbf{W}_r(\mathrm{i,i}) = \sqrt{(\mathbf{T}_r*\mathbf{X})_i((\mathbf{T}_r*\mathbf{X})_i)^*}$ and $\mathbf{W}_c(\mathrm{i,i}) = \sqrt{(\mathbf{X}*\mathbf{T}_c)_i((\mathbf{X}*\mathbf{T}_c)_i)^*}$. The $\nabla\mathbf{TV}(\mathbf{X})$ is then computed by:

$$\nabla(\mathbf{TV}(\mathbf{X})) = (\mathbf{T}_r^{'}*(\mathbf{W}_r)^{-1}*\mathbf{T}_r*\mathbf{X} + \mathbf{X}*(\mathbf{T}_c*(\mathbf{W}_c)^{-1}*\mathbf{T}_c^{'})) \qquad (7)$$

2.3 Reconstruction algorithm

The flowchart of the 3D sparse-sampling PACT with DL is shown in Fig. 1. The first stage in the flowchart is data sampling. In this stage, two types of data acquisition strategies were used. One is full sampling, where about 5% of uniformly distributed fully sampled frames are selected from the 3D imaging region, the other is sparse sampling, where sparse sampling is

applied to all other frames in the 3D imaging region. The second stage is the dictionary training. The fully sampled photoacoustic images of the previous stage are reconstructed using BP and combined into a training data. The dictionary $\mathbf{D}$ is computed by running the K-SVD algorithm on the fully-sampled data. The third stage is the PACT reconstruction with DL. In this stage, the 3D PACT imaging is implemented by reconstructing each 2D photoacoustic image. Specifically, to recover ith B-scan with sparse sampling, it is initialized with the results reconstructed by BP. With the system matrix $\mathbf{K}$, raw measurements $\mathbf{Y}$ and the initial image $\mathbf{X}_i$ as inputs, the DL reconstruction process of one B-scan is implemented using an iterative process with CGD optimization method. The major steps of this optimization process are presented in algorithm 1.
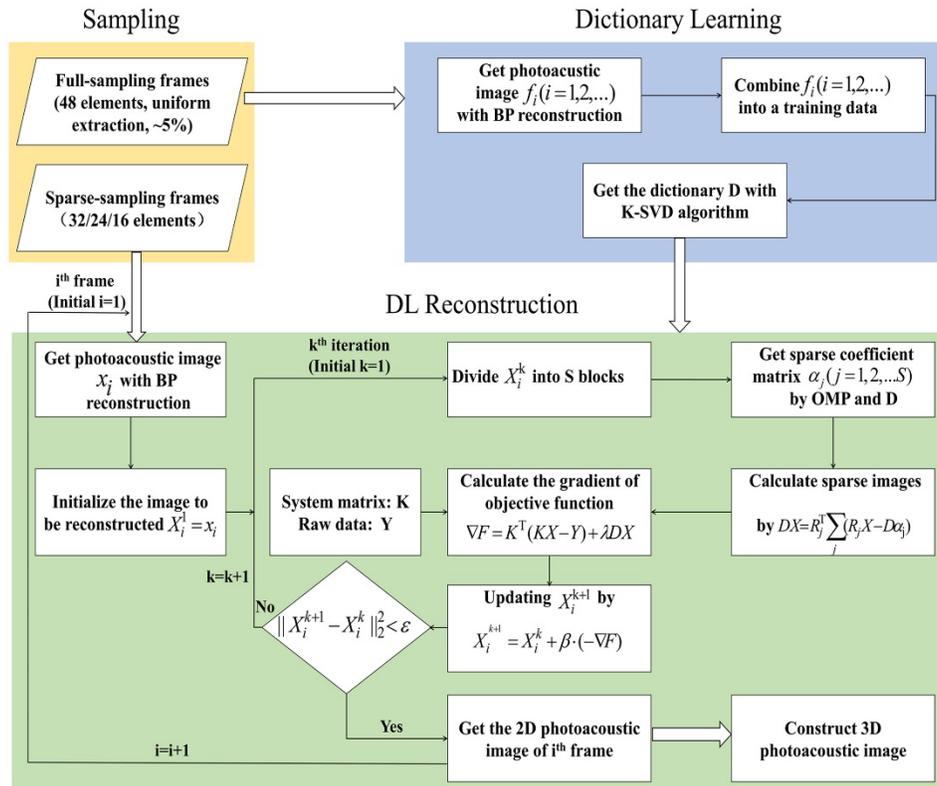

Fig. 1. Flowchart of our proposed algorithm based on learning dictionary.

Algorithm 1: Optimization algorithm of DL-based reconstruction for one B-scan

---

Input: $\mathbf{Y}$ --measurement data, $\mathbf{K}$ --system matrix; λ_1, λ_2 --regularization parameters;

$\quad\quad$ β --iteration step, k --iteration number; ε --stopping threshold; $\mathbf{D}$ --dictionary.

Output: $\mathbf{X}$ -- reconstructed photoacoustic images

Initialization: $k = 1$; $\mathbf{X}_i^k = x_i$, x_i is the initial 2D image of the ith frame reconstructed by BP;

Repeat:

$\quad$ Step 1: Divide $\mathbf{X}_i^k$ into S sub-patches;

$\quad$ Step 2: Compute sparse-coefficient matrix $[\boldsymbol{\alpha}_j]_{j=1}^{S}$ using OMP and $\mathbf{D}$;

$\quad$ Step 3: Compute the gradient $\nabla\mathbf{F}$ of the objective function in Eq. (5) using Eqs. (6) and (7);

$\quad$ Step 4: Update $\mathbf{X}_i^k$ by $\mathbf{X}_i^{k+1} = \mathbf{X}_i^k + \beta * (-\nabla\mathbf{F})$;

$\quad$ Step 5: if $\left\|\mathbf{X}_i^{k+1} - \mathbf{X}_i^k\right\|_2^2 < \varepsilon$

$\quad\quad$ then $k = k + 1$ and goto step 1;

$\quad\quad$ else EXIT

---

2. 4 Imaging system and in vivo experiments

All in vivo experiments in our work were performed using a linear ultrasonic-array based PACT system. The major components of this system include: (1) a Q-switched Nd: YLF laser-pumped tunable dye laser with a laser repetition rate of 10 Hz and a pulse width of 6ns; (2) a 48-element linear ultrasound array (30MHz center frequency, 70% bandwidth) with each element of size 800μm × 200μm; (3) a high-frequency 8-channel PCI data acquisition card. The system is equipped with a container filled with deionized water, and a low-density polyethylene film-sealed window underneath the container for laser irradiation and signal acquisition. The film underneath and deionizing water provides good coupling for receiving ultrasound signals. In this system, the data acquisition for each two-dimensional images requires 6 laser pulses, and the three-dimensional data is obtained by mechanical scanning of the two-dimensional ultrasonic probe.

For in-vivo imaging experiments, the dye laser output was tuned to 584 nm—an isosbestic point at which the oxy- and deoxy-hemoglobin absorb the light equally. For small animal imaging, the Sprague Dawley rat (Harlan Laboratories, Inc., USA) was used. For anesthesia, intra-dermal injection of ketamine (85 mg/kg) and xylazine (15 mg/kg) mixture was used. For human hand imaging, the optical fluence on the skin surface was set to ~0.5 mJ/cm2 per pulse, well below the ANSI recommended maximum permissible exposure (MPE) of 20 mJ/cm2 for a single pulse in the visible spectral range [29]. Since the ANSI safety limit for this pulse width region is based dominantly on the thermal mechanism, our compliance with the ANSI standards guarantees no thermal damage to the tissue. All animal and human experiments described here were carried out in compliance protocols approved by Washington University, St. Louis, USA.

3. Results

3.1 In vivo imaging

To verify the performance of the proposed sparse-sampling PACT with DL, in vivo imaging of vascular bed in the back of two rats (Rat-A and Rat-B) and a human hand were performed. For a reconstructed 3D volume, 166 B-scan frames were acquired, with each reconstructed B-scan image in a 3D volume consisting of 128 × 128 pixels (corresponding to a cross-section of ~6.4 mm × 1.6 mm). For constructing a global dictionary, 8 homogeneously distributed frames from the entire 3D imaging region were extracted and fully sampled as discussed before (here, full sample means all transducers in the ultrasonic array are used to acquire data). The reconstructed photoacoustic images by BP were used as the training data to derive a dictionary. The sample images and the corresponding dictionary of the human hand and two rats are shown in Fig. 2. Figures 2(A) - (C) are the maximum-amplitude-projection (MAP, note: the projection is performed along the depth direction) photoacoustic images of the human hand and two rats with full sampling data, respectively, and they are used as the reference images for all the following sections in this paper. Figures 2(1) – 2(8) are homogeneously distributed 8 sampling frames on the entire 3D imaging region.
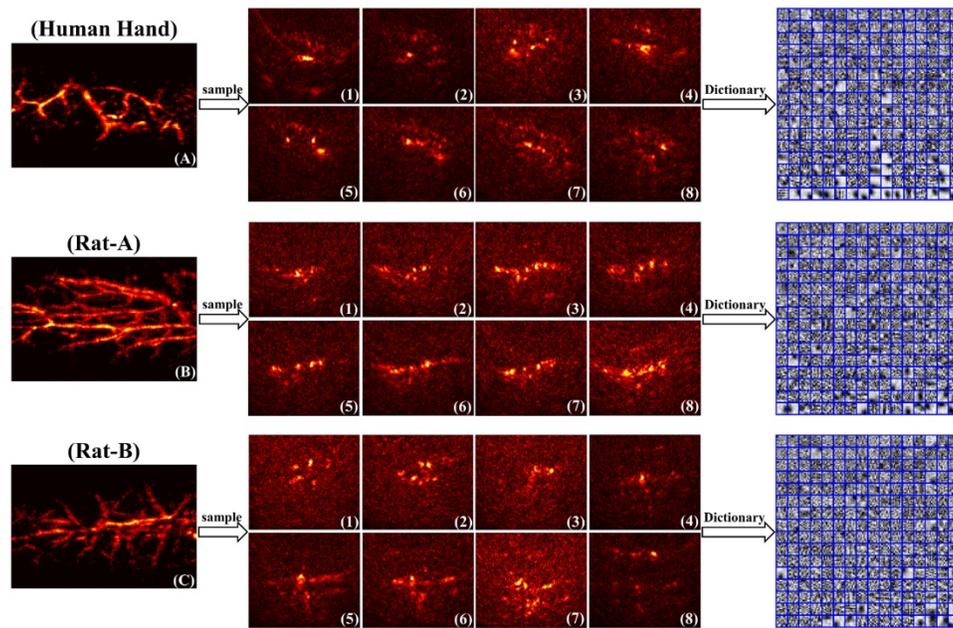

Fig. 2. Training samples and the corresponding dictionaries. (A-C), MAP images of a human hand and two rats, reconstructed by BP with the data from all transducer elements; (1) – (8), 8 full-sampling photoacoustic images reconstructed by BP.

The results of the human-hand reconstructed by the traditional BP, Wavelet-TV, DL, and DL-TV with the different sampling rate (SR) are shown in Fig. 3. The error images between the reconstructed results and the reference image (Fig. 2(A)) are also shown in this figure and are placed adjacent to their MAP images. While computing the errors, all images are normalized by their mean values. In our Wavelet-TV method, a four-level Daubechies wavelet is used as the sparse transform. Figures 3(A1) – (A4) are MAP images reconstructed by BP, Wavelet-TV, DL and DL-TV with 2/3 SR, respectively. Under this 2/3 sampling rate, the artifacts in BP image are very apparent and are suppressed effectively by the Wavelet-TV, although few streak artifacts remain. Our proposed DL and DL-TV can recover higher-quality photoacoustic images while removing almost all sparse-sampling artifacts. The error images showed that all the methods discussed above have very small reconstruction errors and

demonstrated their reconstruction accuracy at the corresponding sampling rate. When the sampling rate decreased to 1/2, the artifacts in the photoacoustic image reconstructed by BP increased (Figs. 3(B1) – (B4) and their errors). Although Wavelet-TV suppressed artifacts to some extent, many artifacts remained in the reconstructed results (Fig. 3(B2)). Using our proposed DL method, most of the artifacts are removed, and acceptable results are obtained (Fig. 3(B3)) even for 1/2 sampling rate. When TV is incorporated, artifacts are suppressed further while maintaining the details at the same time (Fig. 3(B4)). The error images show that higher fidelity was achieved by our proposed method at this SR. When the sampling rate was decreased further, i.e., to 1/3 of sampling rate, the sparse-sampling artifacts become distinct in the reconstructed BP image (Fig. 3(C1)). In this case, only few artifacts are suppressed by Wavelet-TV, and this method fails to recover photoacoustic images at this sampling rate (Fig. 3(C2)). When our proposed DL and DL-TV methods are used, acceptable results are achieved with most of the artifacts removed except for the small amount of signals blurring. Although errors increase with the sampling rate, the majority of the signals are recovered completely by our proposed method.

The reconstructed photoacoustic images of one typical B-scan (the 92nd frame, indicated by the yellow line in Fig. 3 (A1)) are displayed in Fig. 4. Figures 4(a1) – (a4) are the reconstructed results by BP, Wavelet-TV, DL and DL-TV, respectively, with 2/3 SR, Figs. 4(b1) – (b4) are the corresponding reconstructed results with 1/2 SR, and Figs. 4(c1) – (c4) are those obtained with 1/3 SR. Error images are placed next to the reconstructed results. These recovered B-scan images exhibited similar results with the MAPs given in Fig. 3. With the decrease in sampling rate, the artifacts become more and more prominent in the photoacoustic images recovered by BP, and they can be suppressed with varying degree of success using Wavelet-TV for of 2/3 and 1/2 SR, except for 1/3 SR where artifacts were significant. However, when our proposed DL-TV method is used, the artifacts in the photoacoustic images can be removed dramatically independent of the 2/3 or 1/3 SR. The error images in this figure clearly show the reconstruction fidelity of our method compared to different reconstruction methods.
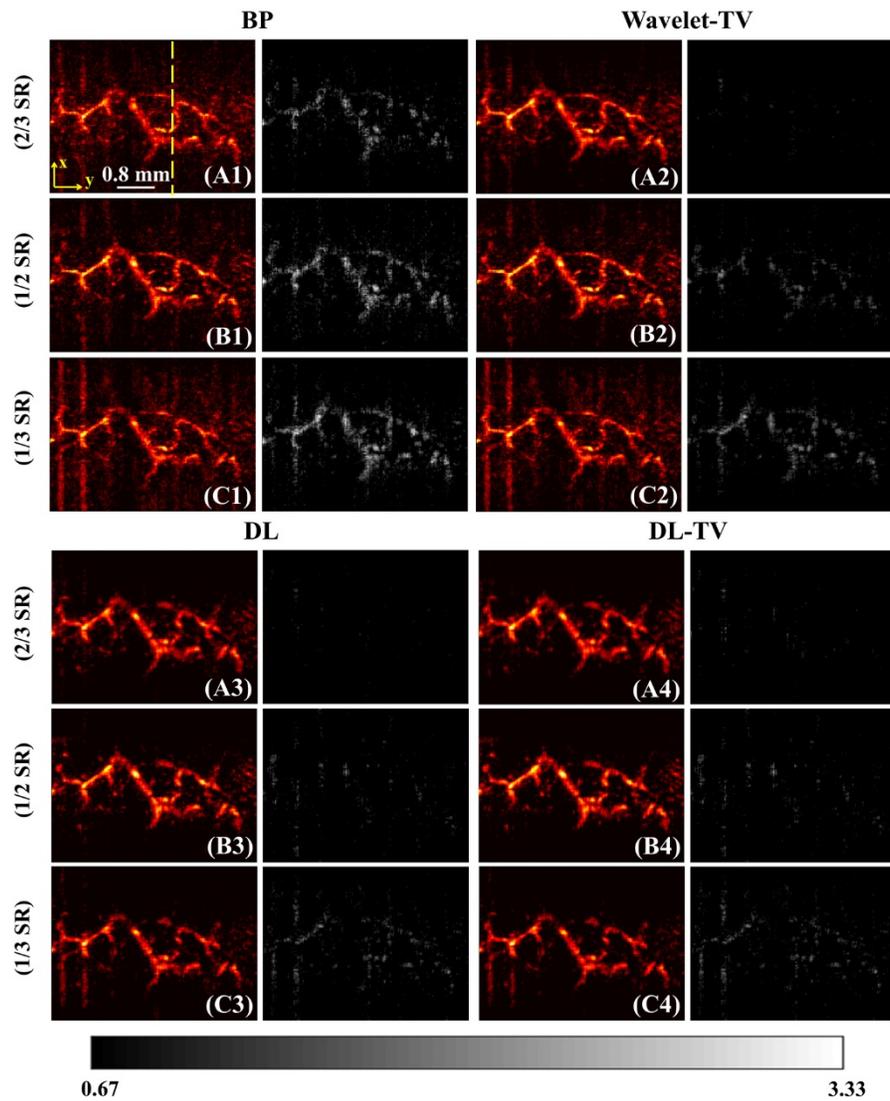
Fig. 3. Reconstructed photoacoustic images and the errors of a human hand. (A1) – (A4), MAP images reconstructed by BP, Wavelet-TV, DL and DL-TV with 2/3 SR; (B1) – (B4), Results reconstructed by the four methods with 1/2 SR; (C1) – (C4), Results reconstructed by the four methods with 1/3 SR. x is the lateral direction of the transducer array, y is the mechanical scanning direction, colorbar represents the grayscale range of the error images.
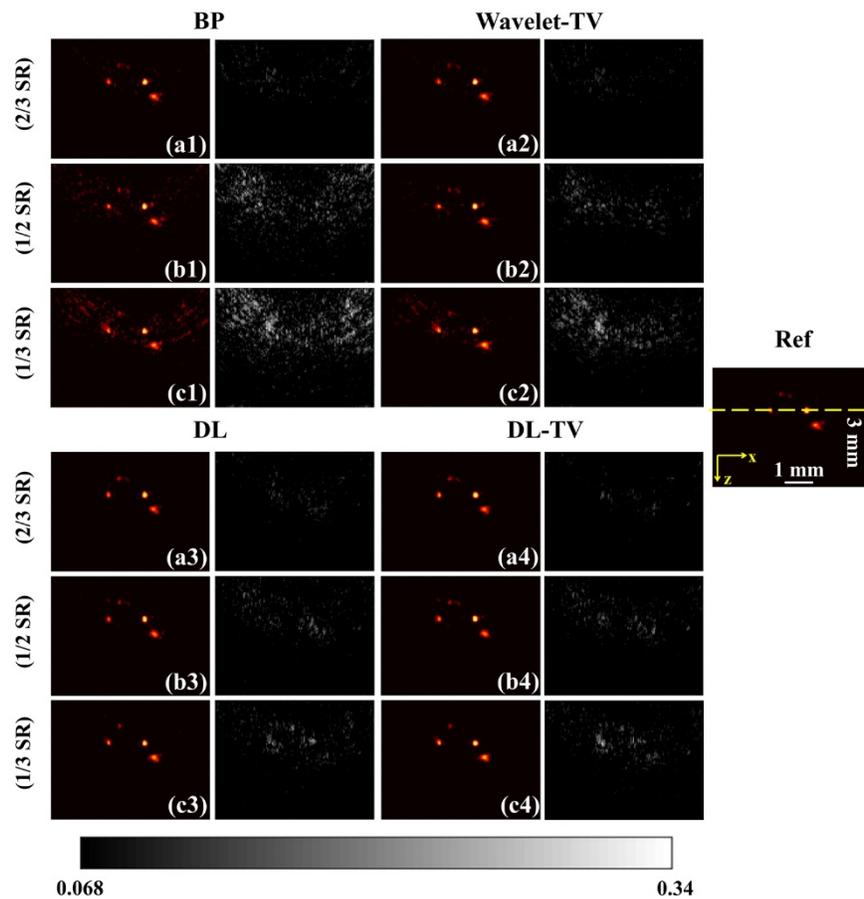
Fig. 4. Reconstructed results and errors of one B-scan (indicated by the yellow line in Fig. 3 (A1)) of the human hand. (a1) – (a4), B-scan images reconstructed by BP, Wavelet-TV, DL, and DL-TV with 2/3 SR; (b1) – (b4), Results reconstructed with 1/2 SR; (c1) – (c4), Results reconstructed with 1/3 SR; z represents the depth direction.

To further validate our proposed DL-TV reconstruction method, in vivo experiments were performed on 2 rats, namely Rat-A and Rat-B. Figure 5 shows the reconstructed MAP images of two rats with different reconstruction methods and sampling rates. The first and fourth row in this figure shows reconstructed MAP images and the corresponding errors by BP, Wavelet-TV, and DL-TV with 2/3 SR, the second row and fifth row shows the corresponding results and errors with 1/2 SR, and the third row and sixth row shows the results and errors with 1/3 SR. With the decreasing sampling rate, the quality of the reconstructed images by BP decreases rapidly. The Wavelet-TV method could recover satisfying images with 2/3 SR, but cannot recover effectively for 1/2 and 1/3 SR. Especially for the 1/3 SR, many weak signals emerged in the artifacts. When our proposed DL-TV method is employed, high-quality photoacoustic images are obtained for both 2/3 and 1/2 SR, and only few artifacts appear with 1/3 SR. Compared to the reference image, our proposed method captures almost all signals while simultaneously removing the artifacts resulted from the sparse sampling effectively . Even though some weak details are suppressed in the reconstruction process, our proposed method provides a clear profile of the imaging object and higher-accuracy photoacoustic images. The error images also demonstrate the superiorities of the proposed method over Wavelet-TV and BP methods.
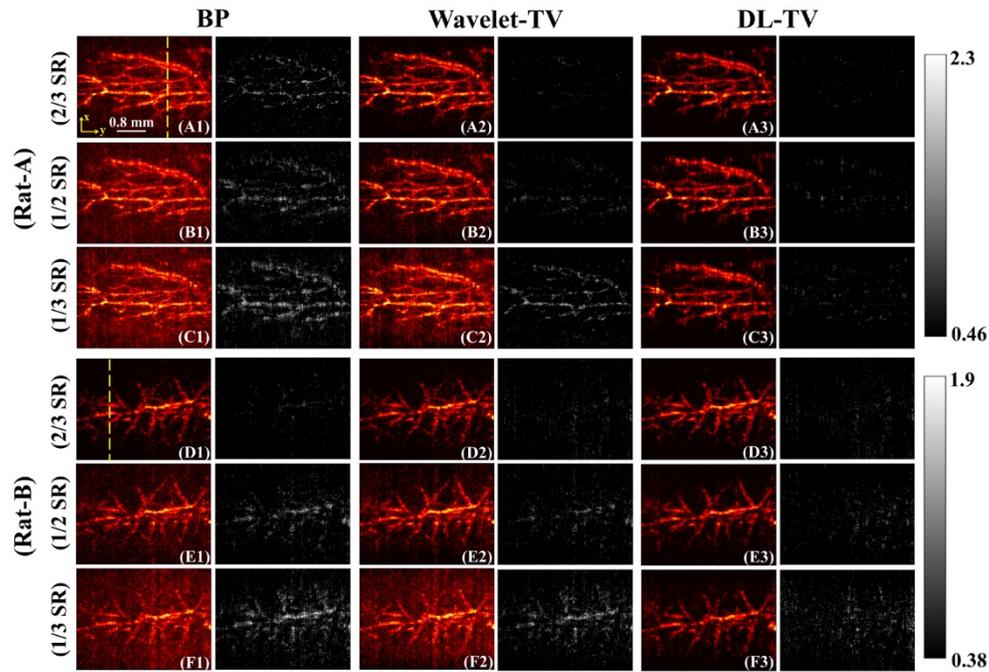
Fig. 5. MAP photoacoustic images and the corresponding errors of Rat-A and Rat-B. (A1) – (A3) and (D1) – (D3), Results of Rat-A and Rat-B reconstructed by BP, Wavelet-TV and DL-TV with 2/3 SR, respectively; (B1) – (B3) and (E1) – (E3), Results of Rat-A and Rat-B reconstructed by the three methods with 1/2 SR, respectively; (C1) – (C3) and (F1) – (F3), Results of Rat-A and Rat-B reconstructed by the three methods with 1/3 SR, respectively.

The results of two typical B-scans from the two rats, indicated by the vertical dashed lines in Figs. 5(A1) and (D1), are shown in Fig. 6. The recovered photoacoustic images using different methods and their reconstruction errors are all shown in this figure. In these images, it can be seen that high quality and high accuracy photoacoustic images are obtained by our proposed method for all 2/3, 1/2 and 1/3 SRs. Although few weak signals are suppressed while removing artifacts for the 1/3 SR reconstruction, the major signals of the images are well preserved.
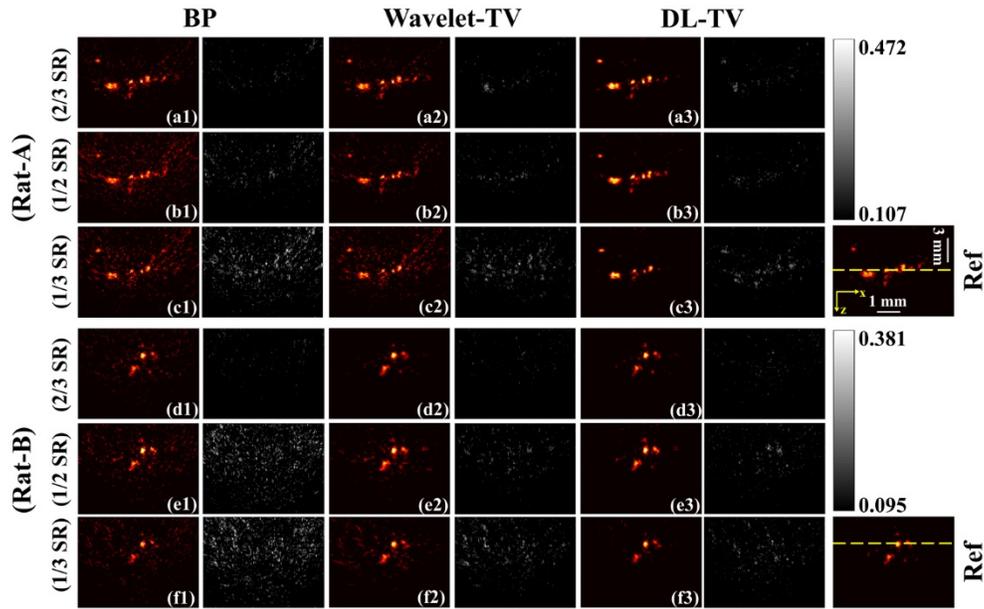
Fig. 6. Reconstructed results and errors of two B-scans from Rat-A and Rat-B (indicated by the vertical dashed lines shown in Figs. 5(A1) and (D1)). (a1) – (a3) and (d1) – (d3), B-scan images reconstructed by BP, Wavelet-TV, and DL-TV with 2/3 SR, respectively; (b1) – (b4) and (e1) – (e3), Results reconstructed with 1/2 SR of the two B-scans, respectively; (c1) – (c4) and (f1) – (f3), Results reconstructed with 1/3 SR of the two B-scans, respectively.

3.2 Quantitative analysis

In this section, quantitative analyses are presented to compare the results between our DL based model against other models.

The first quantitative parameter is the contrast-to-noise ratio (CNR) calculation. Three B-scans reconstructed at 1/2 SR from the data sets of human-hand, Rat-A and Rat-B, were selected to compute the CNRs, and are shown in Fig. 4 (B-scan from the human hand) and Fig. 6 (B-scans from Rat-A and Rat-B). The plots of relative optical absorption amplitudes of the three B-scans, along the dashed lines in the reference images, are shown in Fig. 7. For each B-scan, two representative peak signals (indicated by the arrows) are selected to compute the CNRs, and the results are listed in the tables for different methods. From the table, the images recovered by the DL-TV method achieves highest CNR, and it is about 1.9x compared to the reference image. When compared to BP and Wavelet-TV methods, DL-TV method achieved higher CNR and were 2.5x and 1.8x, respectively.

The second quantitative parameter used for comparison between the reconstructed images and the reference images is the mean square errors (MSE). The definition of MSE is:

$$MSE(\hat{\mathbf{X}}) = \left\| \hat{\mathbf{X}} - \mathbf{X} \right\|_2^2 \Big/ N \qquad (8)$$

where $\hat{\mathbf{X}}$ denotes the reconstructed B-scan image using different methods, $\mathbf{X}$ denotes the reference B-scan image, and N is the number of total pixels in one B-scan.
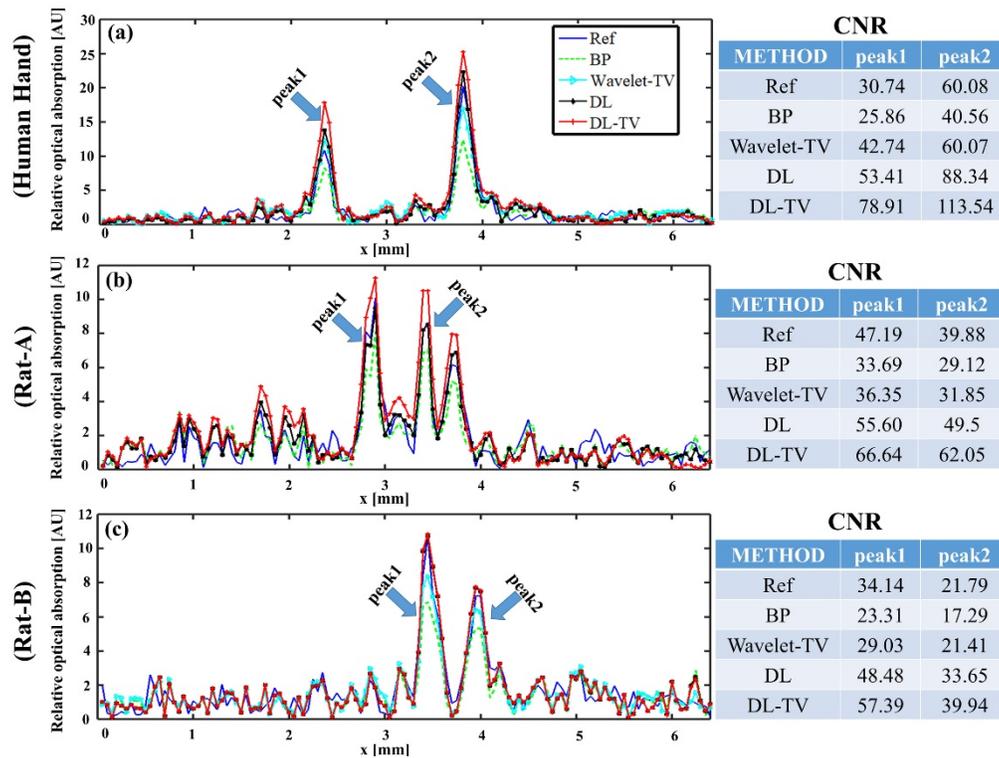
Fig. 7. Photoacoustic amplitudes (relative optical absorption) along the chosen dashed lines in B-scan images of (a) Human hand (Ref. in Fig. 4), (b) Rat-A and (c) Rat-B (Refs. in Fig. 6).

The MSEs for 166-frame photoacoustic images reconstructed at 1/2 SR by different methods is shown in Fig. 8. Figures 8(a) – (c) are the MSEs corresponding to the data sets of the human hand, Rat-A, and Rat-B, respectively. From these curves, we can see that the MSEs of our proposed DL/DL-TV are much smaller than that of other traditional reconstruction methods. In several frames, the MSEs of our proposed method is larger than the Wavelet-TV method. We analyzed these frames and found that these frames contained very little signal, and most of the regions in the reference B-scan images contain background noises. Our proposed method has stronger noise-suppressing abilities than other methods. Thus, for this kind of frames, the errors between the reconstructed results from our proposed method and the reference images are larger than that from the wavelet-TV method. In order to have a better quantitative comparison, the maximum MSE, minimum MSE and average MSE of all frames are analyzed, and the computed values are shown in the bar graphs. Compared with the Wavelet-TV method, the MSE of our proposed method is decreased by about 3.7x demonstrating the superiorities of our proposed method in reconstructing PACT with sparse sampling.
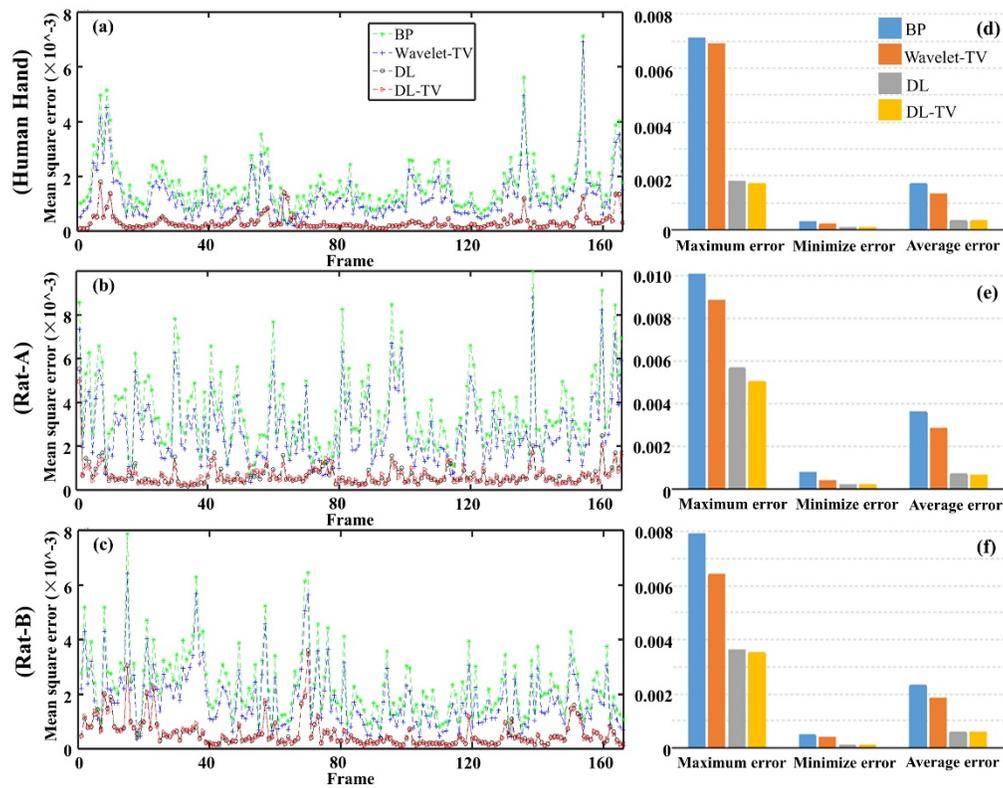
Fig. 8. MSE Curves of the 166-frame photoacoustic images reconstructed by different methods with 1/2 SR. (a), Curves of the human hand data set; (b), Curves of the Rat-A data set; (c), Curves of the Rat-B data set; (d) – (f), the bar graphs of maximum MSE, minimum MSE and average MSE of the three data sets.

3.3 Convergence analysis

The above quantitative analyses have demonstrated the superiorities of the DL based PACT in improving the CNR and accuracy of the photoacoustic images reconstructed with sparse sampling. In this section, we will analyze the convergence of our proposed method.

The difference between two successive photoacoustic images reconstructed from iterations was used as the iterative stopping criterion, as illustrated in section 2.3. In our experiments, when the difference between the reconstructed images from two successive iterations becomes very small, i.e., less than the set threshold, the iterative optimization process will stop. To verify the convergence of our proposed method, the different curves for three B-scans, which are randomly selected from the three data sets, are plotted against iterations and are shown in Figs. 9(a) – (c). The differences decrease rapidly within the first 30 iterations and they converge after about 80-100 iterations for all cases of 2/3, 1/2 and 1/3 SRs. For our proposed method, the differences between images from successive iterations are near zero after 60 iterations, demonstrating the fast convergence of our proposed method.
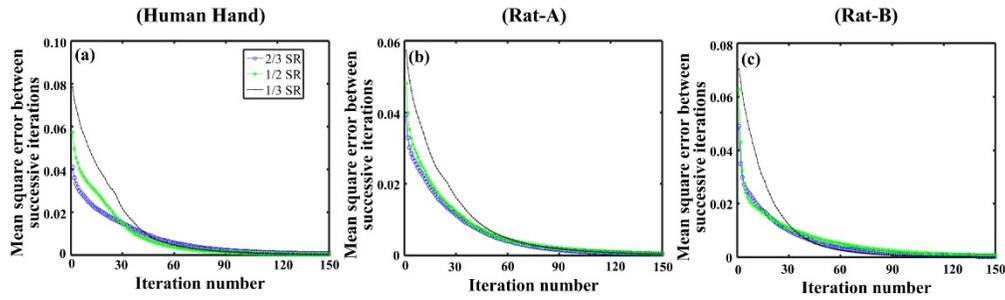
Fig. 9. Convergence curves with iteration number of the B-scans (a) one B-scan from the human hand, (b) one B-scan from the Rat-A, and (c) one B-scan from the Rat-B.

To compare the computing performances of our proposed method, the imaging speed of BP, Wavelet-TV, and DL-TV are provided in Table 1 for a 10Hz laser repetition rate and 8-channel DAQ. The presented imaging speed, which includes data acquisition time as well as reconstruction time, is for one frame and is determined on a PC with an AMD A8-4500 APU of 1.9GHz. Following observations can be made from the Table: (1) The DL-TV with 1/3 SR has ~60% smaller DAQ time than that of the methods with full SR. For DL-TV and Wavelet-TV with the same SR, DAQ time are almost same; (2) The reconstruction time of DL-TV and Wavelet-TV is much longer than the conventional BP method due to the intrinsic iterative process involved in these techniques. Although DL-TV image-reconstruction time is 2x more than Wavelet-TV time, when using the same SR, the DL-TV reconstruct higher-quality photoacoustic images. To accelerate the off-line iterative reconstruction process further, graphical processing unit (GPU) based parallel computation method can be employed in the future work.

Table 1. Comparisons of the imaging speed of the different reconstruction methods

Image size	Reconstruction time (s) / frame			DAQ time (s) / frame		
128*128 pixels	BP (Full SR)	Wavelet-TV (1/3 SR)	DL-TV (1/3 SR)	Full SR	1/3 SR (Wavelet-TV)	1/3 SR (DL-TV)
	5.33	78.46	164.69	0.6	0.2	0.22

4. Discussion and conclusions

Compared to the traditional BP and Wavelet-TV methods in reconstructing the sparse-sampling PACT, the proposed DL based method is superior in reconstruction quality and accuracy. We demonstrated the improved quality and accuracy through *in vivo* experiments. In this section, we discuss several novel aspects of our reconstruct process and also present important issues to be considered for the final reconstructed results.

4.1 Optimization of reconstruction parameters

We used several heuristic parameters while implementing our algorithm and all these parameters are listed in Table 2. First, choosing the right parameters for dictionary learning is very important. The patch size N in the dictionary training should be determined optimally. Smaller patch size will increase the computation cost, and on the contrary a larger patch size cannot capture the image features adequately. In our work, we set the patch size N to 64, which is much smaller compared to column size 256 of the dictioanry. This patch size guarantees sufficient dictionary redundancy for the perfect representation of signals, i.e., the column number K of a dictionary is much larger than the patch size N. Since too much redundancy will result in large computation and excessive smoothing on signals, we limited the redundancy to 4-fold, which proved to be optimal for our cases. The sparsity level L_0 is another important factor affecting the imaging quality, and it is usually set to 5~10. The

details on how to set the sparsity level can be found in Ref. 18. The sparsity level L_0^D in dictionary training and L_0^S in image reconstruction used in our work are given in Table 2.

Second, the regularization parameters λ_1 and λ_2 used in the proposed reconstruction model should be determined appropriately. Overweighed values will result in distortion of reconstructed photoacoustic images and on the contrary, if they are set too small the prior constraints of the two parameters will be ineffective. In addition, the weight distribution between DL and TV, i.e., relative magnitude of λ_1 and λ_2, is also important. Currently, there is no effective method to determine the optimal value for these parameters, and they potentially vary with different data sets. In our experiments, they were determined after several trials, and their values are listed in Table 2 for various sampling rates and image data sets. Other factors influencing the image quality in the reconstruction process include, number of iterations and step size for signal updating per iteration . Their values used in our experiments are also listed in Table 2 for different data sets. In near future, we will investigate methods to determine these parameters optimally.

Table 2. Values of some key parameters used in our experiments

Parameter	Human hand			Rat-A			Rat-B		
	2/3 SR	1/2 SR	1/3 SR	2/3 SR	1/2 SR	1/3 SR	2/3 SR	1/2 SR	1/3 SR
λ_1	0.1	0.2	0.2	0.1	0.15	0.2	0.1	0.15	0.2
λ_2	0.1	0.15	0.15	0.1	0.15	0.15	0.1	0.15	0.15
step	0.1	0.2	0.2	0.1	0.2	0.2	0.1	0.2	0.2
others	Iteration number=90 ; $L_0^S = 10$; $L_0^D = 7$; $N = 64$; $K = 256$								

4.2 Dictionary learning method

In this paper we used the conventional dictionary learning method to reconstruct 3D PACT. However, the field of dictionary learning is continuously evolving, and many novel dictionary-learning methods, such as sparsifying transform learning, structured dictionary learning method, and convolutional dictionary learning have been proposed in the literature to meet higher processing requirements or to meet requirements of the specific application. All these methods are advanced versions of the traditional dictionary learning method and have exhibited superiorities in many signal-processing fields. Compared to these new dictionary-learning methods, the conventional dictionary learning method is simple and easy to use. Since we obtained good results using a conventional technique, we expect new methods to enhance our results further. Other researchers could extend the use of new DL learning methods in photoacoustic imaging based on our results. We also plan to use these new methods in photoacoustic imaging in near future.

4.3 Design of low-cost PACT system

Since our method results in higher image quality compared to other techniques, as presented in the results section, it provides an opportunity to use sparse ultrasonic array with less transducers in practical PACT systems, thus potentially reducing the cost of the system. When the same number of channels in a data acquisition card and the laser repetition rate are used in a system, our method is faster compared to other systems. Thus, our system could potentially reduce the exam time of the patients clinically, benefitting the patients' clinical care.

In many PACT applications, the number of data acquisition (DAQ) system used are usually only a fraction (e.g., 1/2 or 1/4) of the number of elements of the ultrasonic transducer array due to the high cost of the DAQ systems. Typically, multiple laser pulses are required to acquire one B-scan cross-section image. Since the laser repetition rate used in PACT is low (typically 10 – 20 Hz), the image acquisition time is usually large. Therefore, by multi-

plexing the small number of fast DAQ channels, using the large number of DAQ channels can be avoided. Since our proposed method provides an opportunity to use sparse ultrasonic array with less transducers in practical PACT systems without compromising the image quality, a sparse-PACT system with a much smaller number of DAQ channels can be implemented, thus reducing the system cost further.

4.4 Summary

In summary, we presented an advanced reconstruction model with sparse dictionary representation for the 3D sparse-sampling PACT. By incorporating the prior information extracted from the high-quality photoacoustic images via DL, the higher-accuracy results with higher CNR were obtained by our proposed method. Our in vivo results show that using high-speed sparse-sampling PACT system for routine clinical use is feasible.

Disclosures

The authors declare that there are no conflicts of interest related to this article.

References

1. 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 (2018).
2. J. Xia and L. V. Wang, "Small-animal whole-body photoacoustic tomography: a review," IEEE Trans. Biomed. Eng. **61**(5), 1380–1389 (2014).
3. M. U. Arabul, H. M. Heres, M. C. M. Rutten, M. R. H. M. van Sambeek, F. N. van de Vosse, and R. G. P. Lopata, "Investigation on the effect of spatial compounding on photoacoustic images of carotid plaques in the in vivo available rotational range," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **65**(3), 440–447 (2018).
4. P. Suwannasom, Y. Sotomi, Y. Miyazaki, E. Tenekecioglu, Y. Onuma, and P. W. Serruys, "Multimodality imaging to detect vulnerable plaque in coronary arteries and its clinical application," Eur. Heart J. Cardiovasc. Imaging **18**(6), 613–620 (2018).
5. L. Song, K. Maslov, R. Bitton, K. K. Shung, and L. V. Wang, "Fast 3-D dark-field reflection-mode photoacoustic microscopy in vivo with a 30-MHz ultrasound linear array," J. Biomed. Opt. **13**(5), 054028 (2008).
6. B. Wang, L. Xiang, M. S. Jiang, J. Yang, Q. Zhang, P. R. Carney, and H. Jiang, "Photoacoustic tomography system for noninvasive real-time three-dimensional imaging of epilepsy," Biomed. Opt. Express **3**(6), 1427–1432 (2012).
7. G. Paltauf and R. Nuster, "Artifact removal in photoacoustic section imaging by combining an integrating cylindrical detector with model-based reconstruction," J. Biomed. Opt. **19**(2), 026014 (2014).
8. 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 3D photoacoustic tomography," IEEE Trans. Med. Imaging **37**(6), 1382–1393 (2018).
9. L. Nie, Z. Guo, and L. V. Wang, "Photoacoustic tomography of monkey brain using virtual point ultrasonic transducers," J. Biomed. Opt. **16**(7), 076005 (2011).
10. Z. Guo, C. Li, L. Song, and L. V. Wang, "Compressed sensing in photoacoustic tomography in vivo," J. Biomed. Opt. **15**(2), 021311 (2010).
11. J. Meng, C. Liu, J. Zheng, R. Lin, and L. Song, "Compressed sensing based virtual-detector photoacoustic microscopy in vivo," J. Biomed. Opt. **19**(3), 036003 (2014).
12. J. Meng, L. V. Wang, D. Liang, and L. Song, "In vivo optical-resolution photoacoustic computed tomography with compressed sensing," Opt. Lett. **37**(22), 4573–4575 (2012).
13. J. Meng, L. V. Wang, L. Ying, D. Liang, and L. Song, "Compressed-sensing photoacoustic computed tomography in vivo with partially known support," Opt. Express **20**(15), 16510–16523 (2012).
14. J. Meng, Z. Jiang, L. V. Wang, J. Park, C. Kim, M. Sun, Y. Zhang, and L. Song, "High-speed, sparse-sampling three-dimensional photoacoustic computed tomography in vivo based on principal component analysis," J. Biomed. Opt. **21**(7), 076007 (2016).

15. Z. Liu, L. Yu, and H. Sun, "Image restoration via Bayesian dictionary learning with nonlocal structured beta process," J. Vis. Commun. Image Represent. **52**, 159–169 (2018).
16. S. M. Xiang, G. F. Meng, Y. Wang, C. H. Pan, and C. S. Zhang, "Image deblurring with coupled dictionary learning," Int. J. Comput. Vis. **114**(2–3), 248–271 (2015).
17. K. Kreutz-Delgado, J. F. Murray, B. D. Rao, K. Engan, T. W. Lee, and T. J. Sejnowski, "Dictionary learning algorithms for sparse representation," Neural Comput. **15**(2), 349–396 (2003).
18. T. Bai, H. Yan, X. Jia, S. Jiang, G. Wang, and X. Mou, "Z-Index parameterization for volumetric CT image reconstruction via 3-D dictionary learning," IEEE Trans. Med. Imaging **36**(12), 2466–2478 (2017).
19. Y. Wang, N. Cao, Z. J. Liu, and Y. D. Zhang, "Real-time dynamic MRI using parallel dictionary learning and dynamic total variation," Neurocomputing **238**, 410–419 (2017).
20. S. Chen, H. Liu, P. Shi, and Y. Chen, "Sparse representation and dictionary learning penalized image reconstruction for positron emission tomography," Phys. Med. Biol. **60**(2), 807–823 (2015).
21. D. Karimi and R. K. Ward, " A novel structured dictionary for fast processing of 3D medical images, with application to computed tomography restoration and denoising," in Medical Imaging 2016: Image Processing (SPIE., 2016), **9784**: 97840N.
22. C. Jiang, Q. Zhang, R. Fan, and Z. Hu, "Super-resolution CT Image Reconstruction Based on Dictionary Learning and Sparse Representation," Sci. Rep. **8**(1), 8799 (2018).
23. T. M. Quantm and W. K. Jeong, "Compressed sensing reconstruction of dynamic contrast enhanced MRI using GPU-accelerated convolutional sparse coding," in 2016 IEEE 13th International Symposium on Biomedical Imaging (IEEE, 2016), pp. 518–521.
24. S. Ravishankar and Y. Bresler, "Sparsifying transform learning for compressed sensing MRI," in 2013 IEEE 10th International Symposium on Biomedical Imaging (IEEE, 2013), pp. 17–20.
25. L. Zhou, J. Wang, and D. Hu, "The application of dictionary based compressed sensing for photoacoustic image," in International Conference on Machine Learning and Cybernetics (IEEE, 2015), pp. 98–102.
26. S. Govinahallisathyanarayana, B. Ning, R. Cao, S. Hu, and J. A. Hossack, "Dictionary learning-based reverberation removal enables depth-resolved photoacoustic microscopy of cortical microvasculature in the mouse brain," Sci. Rep. **8**(1), 985 (2018).
27. Y. Wang and L. Ying, "Compressed sensing dynamic cardiac cine MRI using learned spatiotemporal dictionary," IEEE Trans. Biomed. Eng. **61**(4), 1109–1120 (2014).
28. M. Lustig, D. Donoho, and J. M. Pauly, "Sparse MRI: The application of compressed sensing for rapid MR imaging," Magn. Reson. Med. **58**(6), 1182–1195 (2007).
29. The Laser Institute of America, American National Standard for Safe Use of Lasers, (ANSI Z136.1–2000), The Laser Institute of America, (2000)