

Sparse-view Signal-domain Photoacoustic Tomography Reconstruction Method Based on Neural Representation

Bowei Yao, Yi Zeng, Haizhao Dai, Qing Wu, Youshen Xiao, Fei Gao, Yuyao Zhang, Jingyi Yu, and Xiran Cai

Abstract—Photoacoustic tomography is a hybrid biomedical technology, which combines the advantages of acoustic and optical imaging. However, for the conventional image reconstruction method, the image quality is affected obviously by artifacts under the condition of sparse sampling. In this paper, a novel model-based sparse reconstruction method via implicit neural representation was proposed for improving the image quality reconstructed from sparse data. Specially, the initial acoustic pressure distribution was modeled as a continuous function of spatial coordinates, and parameterized by a multi-layer perceptron. The weights of multi-layer perceptron were determined by training the network in self-supervised manner. And the total variation regularization term was used to offer the prior knowledge. We compared our result with some ablation studies, and the results show that our method outperforms existing methods on simulation and experimental data. Under the sparse sampling condition, our method can suppress the artifacts and avoid the ill-posed problem effectively, which reconstruct images with higher signal-to-noise ratio and contrast-to-noise ratio than traditional methods. The high-quality results for sparse data make the proposed method hold the potential for further decreasing the hardware cost of photoacoustic tomography system.

Index Terms—Photoacoustic tomography, Image reconstruction, Implicit neural representation, Multi-layer perceptron.

I. INTRODUCTION

PHOTOACOUSTIC compute tomography (PACT) is a medical imaging technique which combines the contrast of optical imaging and the imaging depth of ultrasound

Yi Zeng, School of Information Science and Technology, ShanghaiTech University, Shanghai, China

Haizhao Dai, School of Information Science and Technology, ShanghaiTech University, Shanghai, China

Qing Wu, School of Information Science and Technology, ShanghaiTech University, Shanghai, China

Youshen Xiao, School of Information Science and Technology, ShanghaiTech University, Shanghai, China

Fei Gao, School of Information Science and Technology, ShanghaiTech University, Shanghai, China

Yuyao Zhang, School of Information Science and Technology, ShanghaiTech University, Shanghai, China

Jingyi Yu, School of Information Science and Technology, ShanghaiTech University, Shanghai, China

Xiran Cai, School of Information Science and Technology, ShanghaiTech University, Shanghai, China

imaging [1]–[3]. This technique uses short-pulse (nanosecond duration) laser to illuminate biological tissue and employs ultrasonic transducer to acquire ultrasonic wave generated by instantaneous thermal variation from light-absorbing structures, namely photoacoustic (PA) signals. PACT enables functional and molecular imaging capabilities, which is widely applied in preclinical studies in the fields of tissue imaging, cancer detection and neuroscience research [4]–[10]. For in-vivo imaging, PACT has been used to acquire cross-section images and whole-body dynamics of small-animal [11]–[14], periphery blood vessel structure of human limbs such as lower leg, finger and breast [15]–[17], as well as human cerebral vasculature [18].

High quality PACT imaging usually requires a dense spatial sampling around the object to satisfy the Nyquist sampling theorem and to avoid spatial aliasing [19]. Thus, a high-channel count acquisition system and a transducer array of large number of transducer elements are usually deployed for imaging, resulting increased cost and complexity of the PACT system. To eliminate the artifacts in the images caused by sparse data, such as the streak-type artifact, various approaches have been adopted. Sacrificing spatial resolution of the image, the PA signals may be low-pass filtered so that the required spatial sampling frequency may be lowered as the wavelength used by the universal back-projection (UBP) method [20] for image reconstruction shall be increased [12], [19]. Another approach to improve the image reconstruction quality in PACT is using the model-based (MB) inversion schemes [21], i.e., minimizing the loss function between the measured PA signals and the theoretical ones predicted by a certain PA forward model. Applying proper regularization terms and optimizing method, MB methods yield better image quality than the UBP method, even for a low number of sensors [22]. The trade-off is, however, the associated high computational cost. And the sparsity of sampling will yield the ill-posed problem, which impacts the solution accuracy of inverse problem. For relieving ill-posed problem, proper regularization terms and parameters are selected as constraint to accelerate the convergence of solution [23]. Deep learning based methods have been also introduced to improve image reconstruction quality from sparse data, using the networks based on U-net structure [24]–[26] or attention-steered network [27] in a supervised manner. Thus, the training process requires a large

dataset which is generally difficult to obtain in PACT and the generalization capability of the method heavily depends on the quality of the dataset. Integrating a diffusion model to the inversion of MB method, the optimization problem for image reconstruction with sparse data can be better constrained [28]. However, the diffusion model essentially learns the prior information of the data distribution to better constrain data consistency. The method may only applicable to the learnt data distribution while not to other scenarios. For learning the data distribution, the training process of diffusion model takes for hours, which reducing the timeliness of image reconstruction.

A common feature in all the currently adopted strategies to overcome the ill-posed problem of the inverse problem in PACT, particularly under sparse-view, is that the inversion is formulated in a discrete framework, i.e., the images to be optimized are represented as discrete matrix. Thus, the reconstruction is prone to discretization errors at low resolution. At high-resolution, the increased number of unknowns imposes high time and space complexity and makes the inverse problem more ill-posed.

In recent years, a new paradigm to formulate the inverse problem has emerged in the computer vision and graphics communities [29]. Based on optimizing a multi-layer perceptron (MLP), implicit neural representation (INR) models and represents 3-D scenes from a sparse set of 2-D views in a self-supervised fashion. Benefiting from the image continuity prior imposed by the neural network, INR has achieved superior performance for various computer vision tasks [29]–[31]. Specially, by adding prior information provided by the regularization term, INR has achieved novel view synthesis from sparse-view data [32], [33].

Initiated by the computer vision and graphics communities, INR has been widely applied in other image reconstruction tasks. In computed tomography (CT) image reconstruction, INR methods were used to recover high-quality artifact-free images from a sparse-view sinogram data [34]–[36]. In non-line-of-sight (NLOS) imaging, INR method achieved state-of-the-art reconstruction performance under both confocal and non-confocal settings [37].

Enlightened by the aforementioned work, in this work, we proposed a framework based on neural representation(NR) for PACT image reconstruction with sparse data. Specifically, the initial heat distribution is represented implicitly with a neural network whose parameters are determined by the radio frequency (RF) PA signals after training. The training process essentially involves minimizing the errors between the measured PA signals and the signals predicted by a forward model relating the neural representation of the initial heat distribution to PA signals in self-supervised manner. After the training, the PA images can then be mapped by feeding the coordinates to the network. In the following, we present, evaluate and validate the proposed framework with the comparison of the reference image reconstruction methods (UBP and MB) using both simulation and experimental data for different spatial sampling conditions. The results show that the proposed method performed best in terms of image fidelity and artifacts suppression than the reference methods for the same sparse view sampling conditions.

II. MATERIALS AND METHODS

A. Photoacoustic Wave Equation

In photoacoustic tomography, nanosecond laser pulses satisfying the condition of thermal confinement and of neglecting heat conductance are used. In this way, the thermal expansion of the irradiated region is not affected by its neighboring regions [1]. The temporal profile of the light source can be then appreciated as a Dirac delta function and the pressure of the generated ultrasonic waves in a homogeneous medium is given by

$$\frac{\partial^2 p(\mathbf{r}, t)}{\partial t^2} - c^2 \nabla^2 p(\mathbf{r}, t) = \Gamma H(\mathbf{r}) \frac{\partial \delta(t)}{\partial t} \quad (1)$$

where c is the speed of sound (SoS) of the medium, Γ is the Grüneisen parameter and $H(\mathbf{r}) = \mu_a(\mathbf{r})U(\mathbf{r})$ is the amount of energy absorbed per unit volume at position $\mathbf{r} = (x, y)$ with $\mu_a(\mathbf{r})$ the optical absorption coefficient and $U(\mathbf{r})$ the light fluence, and $p(\mathbf{r}, t)$ represents the pressure field at \mathbf{r} and time t . Considering the initial conditions

$$p(\mathbf{r}, t)|_{t=0} = \Gamma H(\mathbf{r}) \quad (2)$$

and

$$\left. \frac{\partial p(\mathbf{r}, t)}{\partial t} \right|_{t=0} = 0 \quad (3)$$

Eq.(1) can be solved by Green's function method. $p(\mathbf{r}, t)$ can be expressed as:

$$p(\mathbf{r}, t) = \frac{1}{4\pi c} \frac{\partial}{\partial t} \int_{S(t)} \frac{p(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} dS(t) \quad (4)$$

where $S(t)$ represents a spherical wavefront originated from \mathbf{r}' with $|\mathbf{r} - \mathbf{r}'| = ct$.

B. Universal Back-projection Method

In 2D PACT with a ring transducer array of radius R (Fig. 1), image reconstruction can be treated as solving the inverse problem of Eq. (4), i.e. reconstruct the initial pressure $p(\mathbf{r}')$ from the RF data $p(\mathbf{r}, t)$ collected by the ring array. The universal back-projection (UBP) [20] method to solve the inverse problem is expressed as:

$$p(\mathbf{r}') = \frac{1}{\Omega_0} \int_{S'(t)} \left[2p(\mathbf{r}, t) - \frac{2t \partial p(\mathbf{r}, t)}{\partial t} \right] \frac{\cos \theta_0}{|\mathbf{r} - \mathbf{r}'|^2} dS'(t) \quad (5)$$

where θ_0 is the angle between the vector pointing to the source position \mathbf{r}' and transducer surface $S'(t)$ (Fig. 1(a)). Ω_0 is a solid angle of the $S'(t)$ with respect to the wave source. For the planar geometry, $\Omega_0 = 2\pi$, and $\Omega_0 = 4\pi$ for ring array with the cylindrical geometry used in this work.

C. Model-based Method

Under the condition of full and dense view sampling, Eq. (5) can be approximately implemented by adding the projection data simply. However, the artifact occur for the condition of limited-view or sparse-view. For enhancing the image quality under the non-ideal conditions, compared to the UBP method, model-based (MB) solves inverse problem for iteration. And

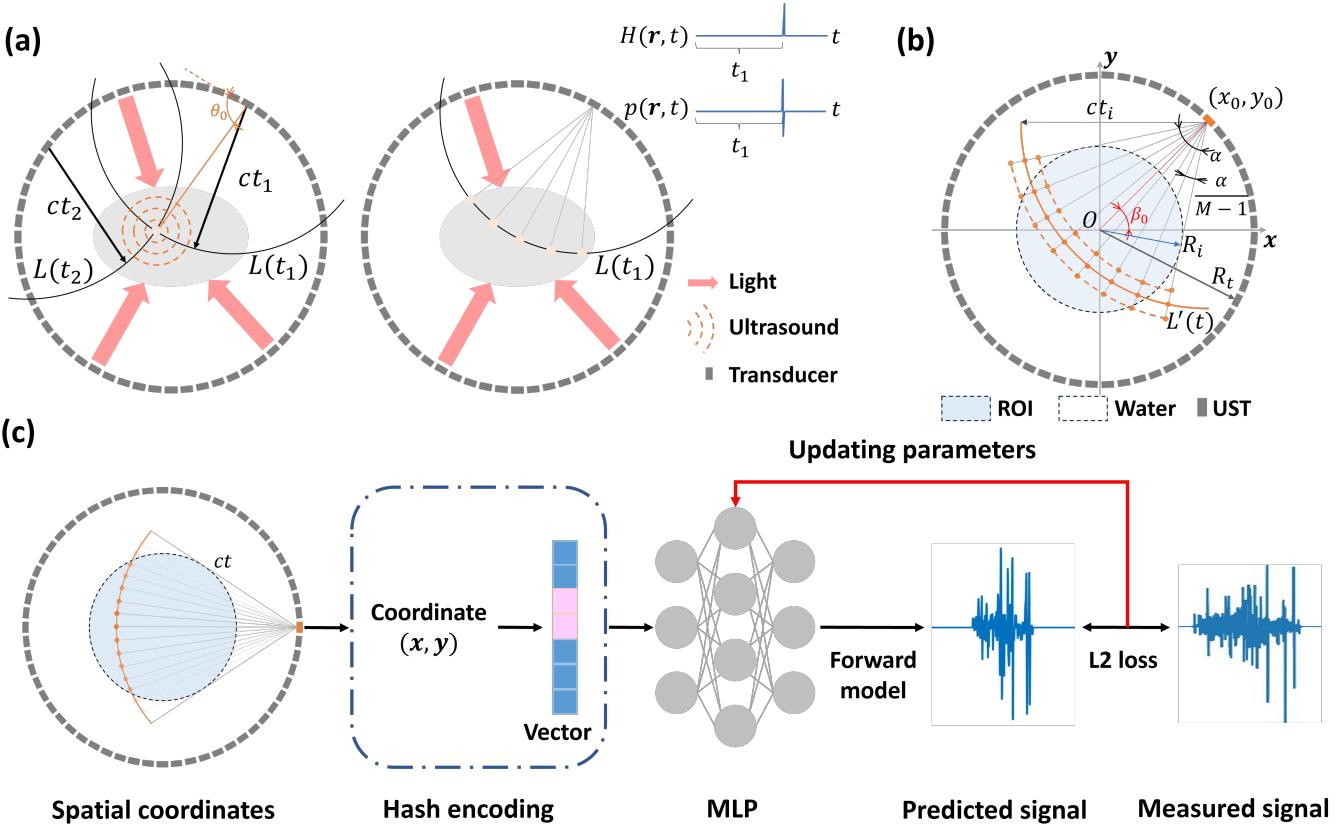


Fig. 1. Overview of neural representation for PACT. (a) Schematic demonstration of propagation of sound wave excited by laser. Under the circumstance of uniform speed of sound, the location of sound source (flesh color point) in body (grey ellipse) can be calculated from transducer element position and time-of-flight. (b) Temporal and spatial discretization of PA forward model. Points on orange solid line represent the sampled coordinates, and points on dash line represent the calculated coordinates. (c) The flowchart of PACT-NeRF proposed in this work. Plane coordinates inside ROI are encoded and put into network, which map the coordinates into heat. Subsequently, forward model maps the heat into ultrasound signal, which is calculated loss from real signal and the loss is back propagated to optimize network. ROI, region of interest; UST, ultrasound transducer.

regularization term is applied to improve the reconstructed image quality [21]. For numerical evaluation of the forward model of PA wave propagation, the derivative in Eq. (4) is approximated as:

$$p(\mathbf{r}, t) = \frac{I(t + \Delta t) - I(t - \Delta t)}{2\Delta t} \quad (6)$$

where

$$I(t) = \int_{L'(t)} \frac{H(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} dL'(t) \quad (7)$$

which is discretized as

$$I(t) \approx \frac{1}{2} \sum_{l=1}^{M-1} \left[\frac{H(\mathbf{r}'_l)}{|\mathbf{r} - \mathbf{r}'_l|} + \frac{H(\mathbf{r}'_{l+1})}{|\mathbf{r} - \mathbf{r}'_{l+1}|} \right] d_{l,l+1} \quad (8)$$

where $L'(t)$ is uniformly sampled (M points) across the angular sector covering the circular region-of-interest (ROI) of radius R_t (Fig. 1(b)). Assuming uniform SoS distribution of the medium, $L'(t)$ in Eq. (6) is an arc of a circle centered on the position of transducer element. Thus, the segment $d_{l,l+1}$ of curve $L'(t)$ can be expressed as:

$$d_{l,l+1} = \begin{cases} \frac{\alpha}{M-1} ct & 1 \leq l \leq M-1 \\ 0 & l = 0, l = M \end{cases} \quad (9)$$

where l indexes the segments, and α represents the opening angle of the sector covering the whole ROI viewing from transducer, calculated as:

$$\alpha = 2\arcsin\left(\frac{R_i}{R_t}\right) \quad (10)$$

in which R_t is the radius of ring-shape transducer. Eq. (8) can be reduced to

$$I(t) \approx \frac{1}{2} \sum_{l=1}^M \left[\frac{H(\mathbf{r}'_l)}{|\mathbf{r} - \mathbf{r}'_l|} \right] (d_{l-1,l} + d_{l,l+1}) \quad (11)$$

Combining Eq. (6) and Eq. (11), the forward model (Eq. 3) can be formulated in matrix form as:

$$\mathbf{p} = \mathbf{AH} \quad (12)$$

where \mathbf{p} is the model predicted pressure signals collected by the transducers, \mathbf{A} is the measurement matrix, and \mathbf{H} is the vectorized image to reconstruct. With the MB method, PA images \mathbf{H} are then reconstructed by solving the inverse problem by minimizing the error between \mathbf{p} and experimentally measured signal \mathbf{p}_m , i.e.

$$\mathbf{H}_{sol} = \arg \min_{\mathbf{H}} \|\mathbf{p}_m - \mathbf{AH}\|_2^2 + \lambda R(\mathbf{H}) \quad (13)$$

where $R(\mathbf{H})$ and λ are the regularization term and regularization parameter, respectively.

D. Neural representation method

We use implicit neural representation (INR) for the PACT images, i.e., the image is represented as a continuous implicit function by a neural network \mathcal{M}_Θ parameterized with Θ :

$$\mathcal{M}_\Theta : (x, y) \rightarrow \mathcal{H} \quad (14)$$

Image reconstruction with the NR method is then converted to finding the optimal Θ minimizing the loss function:

$$\mathcal{L} = \arg \min_{\Theta} \|\mathcal{F}(\mathcal{H}) - \mathbf{p}_m\|_2^2 + \eta R(\mathcal{H}) \quad (15)$$

Here, $\mathcal{F}(\cdot)$ represents the forward operator from heat distribution \mathcal{H} to PA signals (Eq. 12), \mathbf{p}_m is the measured PA signals by the ring array, and η is a hyper-parameter for the regularization term. Thus, the network is trained in a self-supervised manner.

1) Coordinates Selection: To train the neural network, the PA signal must be related to the spatial coordinate of the PA source. At a specific sampling moment t_i , the PA signal amplitude received by the transducer element at (x_0, y_0) is contributed by the point sources located on the curve $L'(t_i)$ (Fig. 1b). After discretization, their coordinate (x, y) can be calculated as:

$$\begin{cases} x = ct_i \cos(\beta_0 + j \frac{\alpha}{M-1}) + x_0 \\ y = ct_i \sin(\beta_0 + j \frac{\alpha}{M-1}) + y_0 \end{cases} \quad (16)$$

where $j \in [0, 1, 2, \dots, M-1]$ indexes the points located on $L'(t_i)$, with β_0 defined as

$$\beta_0 = \arctan\left(\frac{y_0}{x_0}\right) \quad (17)$$

2) Position Encoding and Network Structure: After determining the coordinates of the point sources, the coordinates are hash encoded by multi-resolution mapping [31], and the encoded hash vectors are fed to a MLP, which has two hidden-layers with 128 neurons (Fig. 1(c)). ReLU is selected as the activation function of the neurons and the output activation function is the Sigmoid function to normalize the output value.

3) Training: During the training process, the initial learning rate was set to 0.001 and decreased for every 20 epochs with a momentum of 0.5, and Adam optimizer [38] was used to minimize the loss function. The network was implemented with the tiny-cuda-nn [31] framework in Python.

E. Experiments

1) In silico: A vessel-like structure sized $2.5 \times 2.5 \text{ cm}^2$ was placed at the center of a ring-shaped transducer array (40 mm radius, 256 elements) in a medium of uniform SoS (1500 m.s^{-1}). The sampling frequency was set at 20 MHz and the size of the computational grids was 512×512 with a pixel size of 0.05 mm. Given the the initial thermal distribution of the vessel-like structure (Fig. 3e) and the aforementioned settings, the PA signals were calculated by the forward model (Eq. 12) stated in Sec. II .

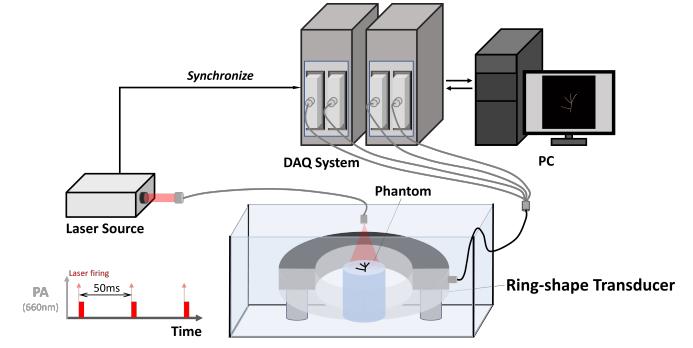


Fig. 2. Phantom experiment setup. An optical fiber illuminates from top-view of the phantom and light spot covers the whole ROI. The laser source fires 20 times per second and delivers laser with 660 nm wavelength, and triggers vantage systems to acquire data synchronously at the same time.

2) In vitro: Five agar phantoms (1% agar w/v, 6 cm diameter) of different embedding materials forming various shapes were prepared for experimental validation of the proposed method. Phantom 1 (1% agar w/v) had three embedded black plastic spheres (3 mm diameter) to mimick the scenario of imaging the cross-section of blood vessels in the body. We also embedded black tungsten wires (0.1 mm diameter) forming leaf branch (Phantom 2), delta (Phantom 3), heart (Phantom 4) and star (Phantom 5) shapes, to mimick the scenario of imaging the longitudinal view of blood vessels in the body.

A 512-element ring transducer array of 80 mm diameter (center frequency: 5 MHz, Guangzhou Doppler Eletronic Technologies Co., Ltd, Guangzhou, China) was mated with two ultrasound research systems (Vantage 256, Verasonics, Kirkland, USA) for receiving PA signals in the experiments. The PA signals were initiated by a laser source (PHOCUS MOBILE, OPOTek Inc. USA) (660 nm wavelength) with 20 Hz repetition rate, synchronizing the data acquisition with 20 MSPS sampling rate (resulting 1024 time samples). For all the experiments, the light was emitted from the top-view of the imaging phantom in the de-ionized and degassed water, and the light spot covered the whole ROI (Fig. 2). The SoS was obtained from the SoS-temperature relationship in pure water [39]. During the experiments, the embedded material of the all the phantoms were place in the imaging plane of the ring array.

F. Image reconstruction and performance evaluation

For all the methods, the image reconstruction area was set as $2.5\text{cm} \times 2.5\text{cm}$ consisting of 512×512 pixels. The signals consisting of 32, 64, 128, 256 and 512 projections were used to reconstruct image by different methods, respectively. For UBP method, the meaningless negative values on the image were set to zero for making a fair comparison. MB method was implemented following [40] using non-negative least square (NNLS) optimizing method with and the iteration number for solving the inverse problem was set to 50. For the NR method, network training was stopped when the loss function is less than 0.0001. Total Variation (TV) was used in the loss function of both the MB and NR methods, in which the regularization

parameters λ and η were selected between 0 and 0.2 after trials. All the above algorithms were implemented in Matlab 2020a on a workstation (Intel core i7, 14 cores at 2.90 GHz and 16GB memory) interfaced with a graphical processing unit (GPU, Nvidia Geforce RTX 4090 Ti).

For the simulation data, we utilized structural similarity index (SSIM) [41] and peak signal-to-noise ratio (PSNR) as the metrics to evaluate the performance of different methods. For the experimental data, we used signal-to-noise ratio (SNR) and contrast-to-noise ratio (CNR) as the evaluating metrics.

SSIM is defined as:

$$SSIM(f, gt) = \frac{(2\mu_f\mu_{gt} + C_1)(2\sigma_{cov} + C_2)}{(\mu_f^2 + \mu_{gt}^2 + C_1)(\sigma_f^2 + \sigma_{gt}^2 + C_2)} \quad (18)$$

in which f and gt represent the reconstruction image and ground truth image, respectively. μ_f (σ_f) and μ_{gt} (σ_{gt}) are the mean (standard derivation) of reconstruction image f and ground truth image, respectively, and σ_{cov} is cross-covariance of f and gt . The default parameter values of C_1 is 0.01, and C_2 is 0.03 [42]. The dynamic range of two parameters is 1.

PSNR is defined as:

$$PSNR(f, gt) = 10\log_{10}\left(\frac{I_{max}^2}{MSE}\right) \quad (19)$$

where I_{max} represents the max value of f and gt , and MSE is calculated by

$$MSE = \frac{1}{n^2} \|f - gt\|_F^2 \quad (20)$$

in which $\|\cdot\|_F$ is Frobenius norm, and n is the size of image.

SNR (unit in dB) is defined as:

$$SNR = 20\log_{10}\left(\frac{\bar{I}_{signal}}{\sigma_{background}}\right) \quad (21)$$

where \bar{I}_{signal} and $\bar{I}_{background}$ are the average amplitude of the selected imaging object region (green dash rectangles in Fig.4&5) and background region (yellow dash rectangles in Fig.4&5), respectively, σ_{signal} and $\sigma_{background}$ represent the standard deviations of selected imaging object region and background region, respectively.

CNR (unit in dB) is defined:

$$CNR = 20\log_{10}\left(\frac{abs(\bar{I}_{signal} - \bar{I}_{background})}{\sigma_{background}}\right) \quad (22)$$

We computed the evaluation metrics with the images reconstructed by different methods from every projection number.

III. RESULTS

A. Simulation Results

For the vessel-like structure, the reconstructed images by the UBP, MB and NR methods using 32, 64, 128 and 256 projections are compared in Fig.3. The parameters for the TV regularization term in the loss functions for MB and NR methods were selected as 0.01 and 0.02, respectively. For 32, 64, 128 and 256 projections, the network converged after 100, 60, 40, and 20 training epochs, respectively. The streak-type artifacts are noticeable in the images reconstructed by the UBP method, which became less visible as the projection

number increased from 32 to 256 (SSIM/PSNR increased from 0.16/16.05 dB to 0.37/24.41 dB) (Fig.3(a-e)).

Compared to UBP, MB and NR methods remarkably removed the artifacts and improved the image quality. As the the projection number increased from 32 to 256, the SSIM/PSNR of MB vs NeRF increased from 0.44/20.81 dB vs 0.92/26.59 dB to 0.97/32.05 dB vs 0.99/36.34 dB (Fig.3(f-i, k-o)). A detailed comparison of two branches in the vessel-like structure (Fig.3 e, p, q) and their line-profiles (Fig.3 (r)-(s)) clearly demonstrated that NR had the best match with the ground-truth data.

B. Phantom Experiments

For phantom experiments, the images were reconstructed using 64, 128, 256 and 512 projections by different methods, respectively. For MB method, λ was set as 0.05, and the iteration number was set as 50. For NR method, The signal amplitude was normalized for training the network, and η was set as 0.02. The network was convergent after 50 training epochs.

1) Plastic Sphere: In phantom 1, multiple rings were observed for each sphere in the images reconstructed by UBP, MB and NeRF (Fig. 4). Consistent with previous observations in the simulated data, image quality of the images reconstructed by the MB and NeRF were better than that by the UBP method for the same projection number and adding more projection data for image reconstruction also improved the image quality for all different methods (Fig.4). Specifically, the SNR/CNR of the images reconstructed by the UBP, MB and NeRF were 1.29 dB/5.92 dB, 12.07 dB/14.57 dB and 14.00 dB/27.33 dB (for 64 projections) and were consistently improved to 21.07 dB/30.10 dB, 26.10 dB/35.03 dB and 49.03 dB/37.62 dB, as the projection number increased to 512 projections. Noticeably,in the images reconstructed by MB and NeRF, the two plastic spheres at the lower part in the images each had a dot recovered at the center of the rings, which was not observed in the images by UBP (Fig.4(a-l)). When comparing the line profiles crossing the two spheres, it was observed that the profile lines in the UBP images had more fluctuations in the background area, while the MB and NR methods had more flat profile lines (Fig.4(m)). Compare to the MB images, the profile lines in the NR images were closer to zero intensity in the background area between the two spheres.

2) Tungsten Wires: Consistent with the observation in phantom 1, in phantom 2 image reconstruction quality of the MB and NR methods were better than that by UBP method for the same projection number. For all the methods, the reconstructed image quality was improved with adding more projection data(Fig. 5). Severe streak-type artifacts in the images reconstructed by UBP method when PA signals were very sparsely acquired (64 and 128 projections) (Fig. 4(a-b)). These streak-type artifacts were largely mitigated by the MB and NR methods for the same projection number (Fig. 4(e-f, i-j)). For 256 and 512 projections, the streak-type artifacts were not obviously observed. Comparing the line profiles along the middle of the object, it was observed that the profile lines of MB and NR methods had less fluctuations than that of UBP

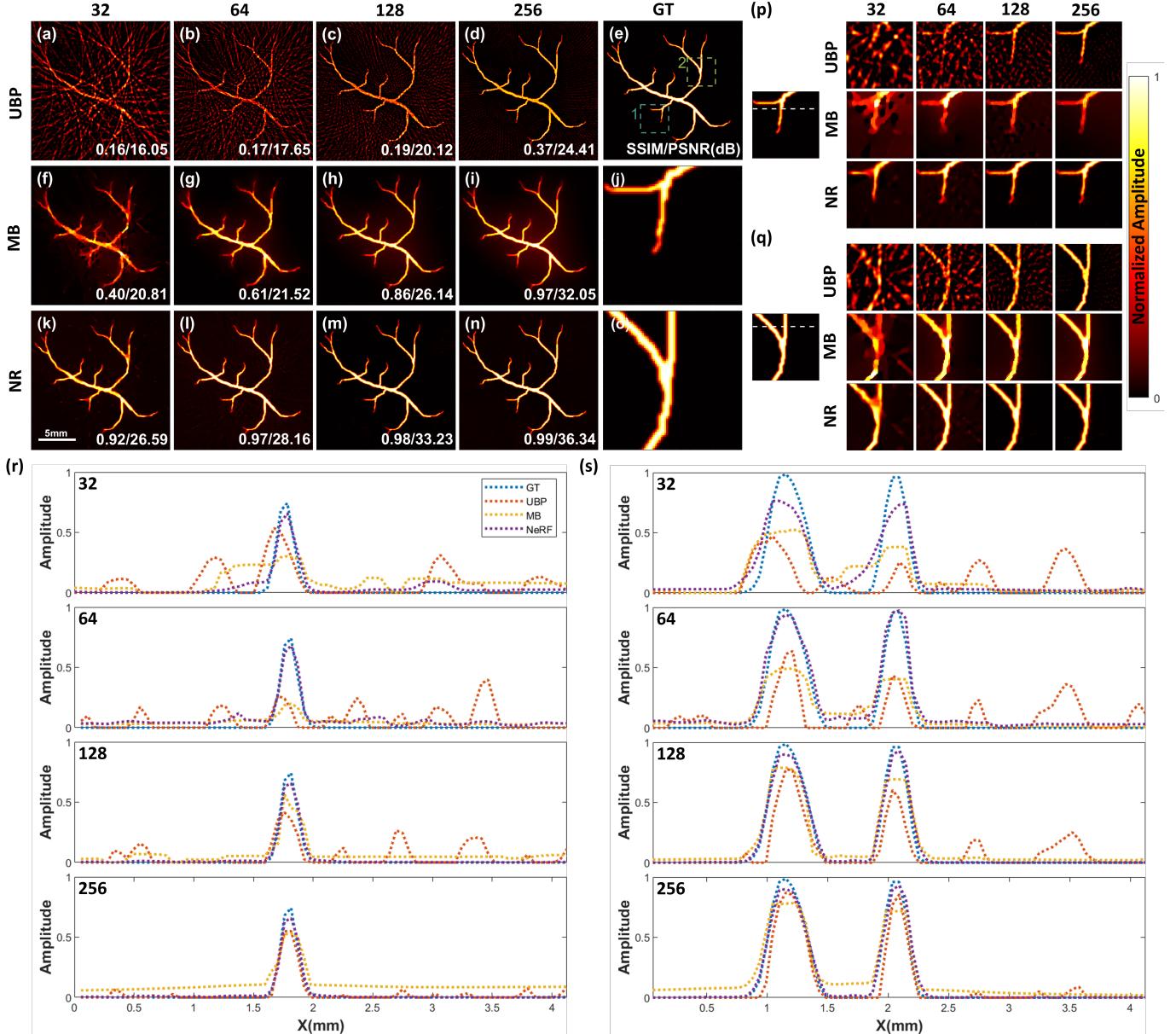


Fig. 3. Results of simulation experiment. (a)-(d) Results of images reconstructed by UBP method from 32, 64, 128 and 256 projections respectively. (e), (j) and (o) Ground truth. (f)-(i) Reconstructed results by MB method, the numbers of projection are the same as UBP method. (k)-(n) Reconstructed images by neural representation from different projection number respectively. (p) and (q) Zoom in specific regions of images. (r) and (s) the profile lines of images reconstructed by different methods from different projection number. NR, neural representation.

for all projection number (Fig. 5(m)). Halo artifacts were consistently observed in the middle of the images reconstructed by the MB method, while this artifact was removed in that of the NR method. Thus, in the background area of images, the profile lines in the NR images were closer to zero intensity than that in the MB images.

Quantitatively, the SNR(CNR) of the images reconstructed by the UBP, MB and NR were 1.94 dB(-3.22 dB), 15.69 dB(13.13 dB) and 19.11 dB(19.08 dB) (for 64 projections) and were consistently improved as the projection number increased to 18.37 dB(17.76 dB), 19.80 dB(17.77 dB) and 48.02 dB(48.06 dB) (for 512 projections). For the same number of projections, SNR(CNR) in the images reconstructed by the NR was consistently better than that of the UBP and MB method,

with 17.2-29.6 dB (22.3-29.8 dB) improvement over UBP and 3.4-28.2 dB (6.0-30.3 dB) improvements over MB for 64-512 projections.

To further demonstrate the performance of NR for sparse data, we compared the images reconstructed by UBP, MB and NR for Phantom 3-5 (Delta, Heart, Star) with 128 projections data(Fig. III-B.2). For the imaging object with a radius of 1cm, as observed, image reconstructed by UBP method remained streak-type artifacts in the background. With TV regularization term, MB method reconstructed images with less artifacts than UBP method. But the SNR and CNR of images were affected by the ill-posed problem. NR method removed most of the artifacts, and reconstructed the most distinct object structure.

Quantitatively, the SNR(CNR) of the images reconstructed

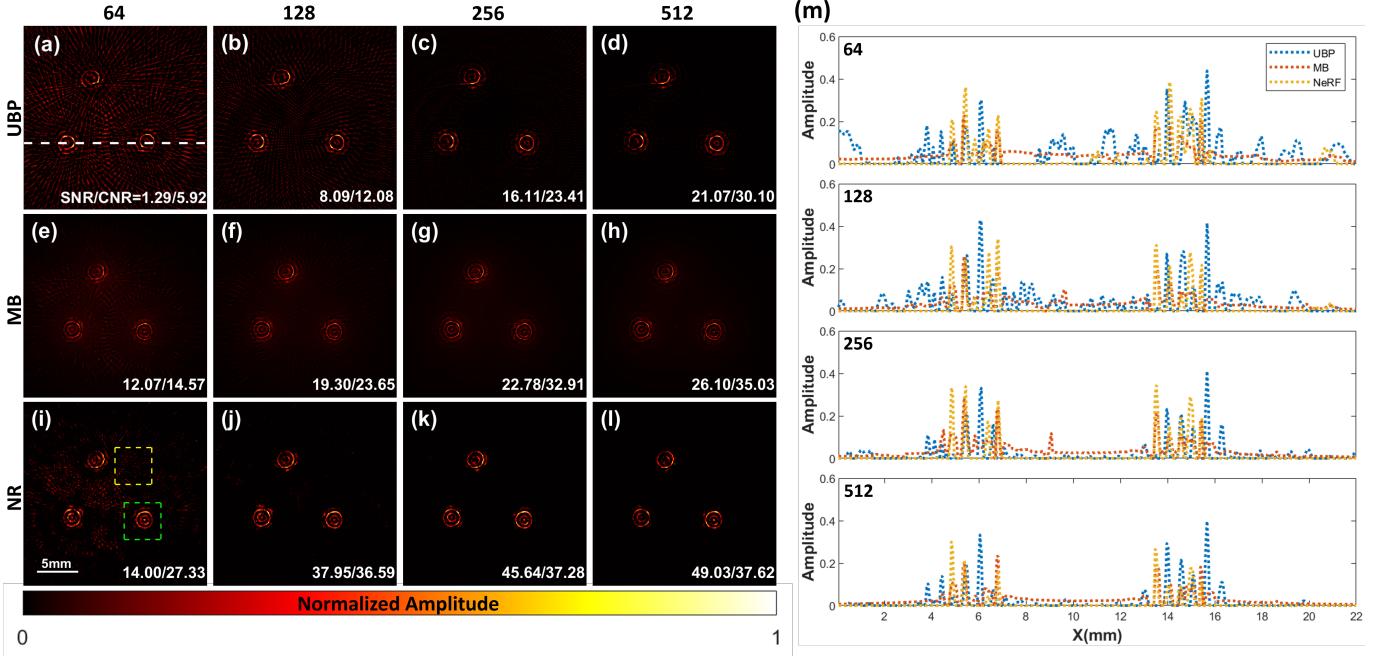


Fig. 4. Results of phantom 1 experiment. (a)-(d) Results of images reconstructed by UBP method from 64, 128, 256 and 512 projections respectively. White dash line in (a) was selected to show the profile of pixels for each image. (e)-(h) Reconstructed results by MB method, the numbers of projection are the same as UBP method. (i)-(l) Reconstructed images by neural representation method from different projection number respectively. Green and yellow dash rectangles in (i) represent the signal region and background region which are selected to calculate CNR and SNR. (m) The profile line of images reconstructed by different method from all projections. The CNR/SNR of image is set at the lower right corner of each image. NR, neural representation

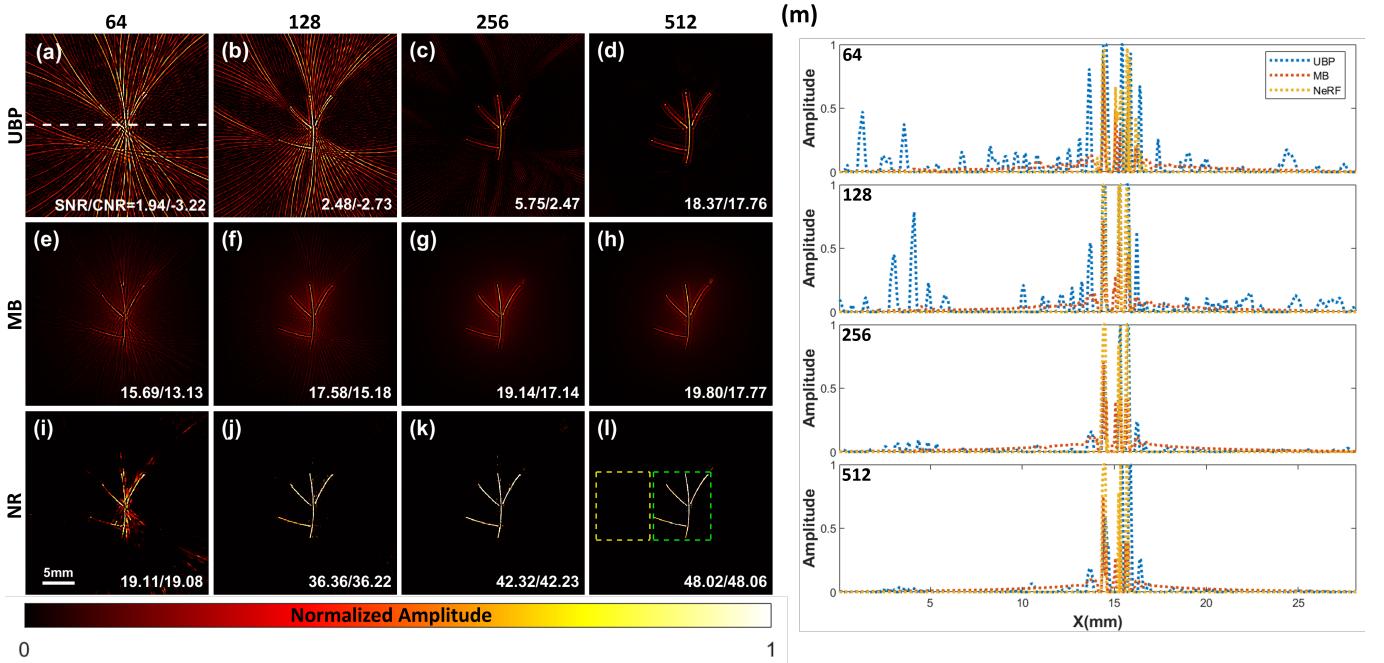


Fig. 5. Results of phantom 2 experiment. (a)-(d) Results of images reconstructed by UBP method from 64, 128, 256 and 512 projections respectively. White dash line in (a) is the location of profile line. (e)-(h) Reconstructed results by MB method, the numbers of projection are the same as UBP method. (i)-(l) Reconstructed images by neural representation from different projection number respectively. Green and yellow dash rectangles in (l) represent the signal region and background region which are selected to calculate CNR and SNR. (m) The profile line of images reconstructed by different method from 64 and 128 projections. The CNR/SNR of image is set at the lower right corner of each image. NR, neural representation

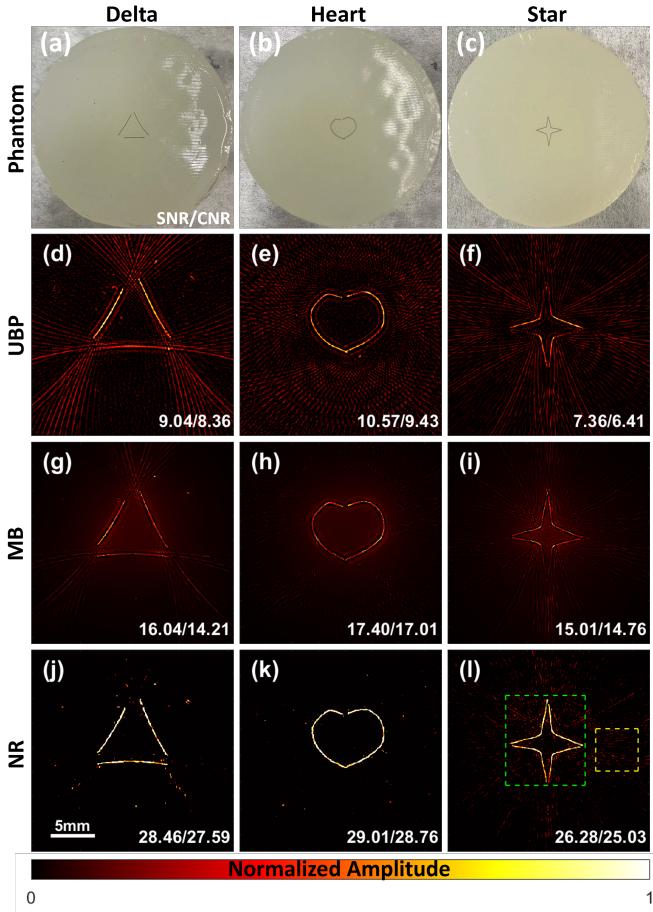


Fig. 6. Experiment results for phantom 3, phantom 4 and phantom 5. (a)-(c) Phantom in different shapes. (d)-(f) Reconstruction images by UBP method from 128 projections. (g)-(i) Reconstruction images by MB method from 128 projections. (j)-(l) Reconstruction images by neural representation method from 128 projections. Green and yellow dash rectangles in (l) represent the signal region and background region which are selected to calculate CNR and SNR. NR, neural representation

by the UBP were 9.04 dB(8.36 dB), 10.57 dB(9.43 dB) and 7.36 dB(6.41 dB) for phantoms in different shape respectively. For the MB method, the SNR(CNR) of the images were 16.04 dB(14.21 dB), 17.40 dB(17.01 dB) and 15.01 dB(14.76 dB) for the same phantoms. And the NR method improved SNR(CNR) to 28.64 dB(27.59 dB), 29.01 dB(28.76 dB) and 26.28 dB(25.03 dB) respectively.

IV. DISCUSSION

PACT combines the imaging depth of ultrasound image and contrast of medical optical image and has a great potential of being widely applied clinically [1]. However, existing reconstruction methods for PACT suffer a low image quality under the condition of sparse-view sampling. Developing imaging reconstruction method with high image quality and low imaging cost for PACT is therefore significant for the application of PACT in clinical practice. In this work, we proposed a neural representation framework to overcome the disadvantages of conventional PACT reconstruction methods under the circumstance of sparse PA acquiring condition,

and reconstructs image from raw PA signals directly by learning the continuous representation of PACT image. We validated our method on simulation and experimental data with various sampling projections. The results show that the proposed method can effectively suppress artifacts under the condition of sparse sampling, and outperform UBP and MB methods [20], [21] under the same sampling condition. These results demonstrate the ability of our method for reducing the hardware cost and improving the image quality of PACT system.

Essentially, the image reconstruction for PACT is an inverse problem. Analytical method, such as UBP [21], deduces an analytical solution of inverse problem from the forward propagation of ultrasound wave directly. UBP method is conventional for the ease of implementation and low time complexity when the spatial sampling condition satisfy the Nyquist sampling theorem. However, for sparse-view sampling, the image quality of analytical method is impacted by streak-type artifact severely. Compared to the UBP method, our method expresses the reconstructed image as a 2-D continuous function, and optimizes a MLP to fit this function under the constraint of prior knowledge, which make a better performance for sparse-view sampling. Building a matrix expression of certain PA forward model, MB method [20] reconstruct PACT image by finding a numerical solution which minimize the loss function between the measured PA signals and the theoretical ones. For MB method, the interpolation method only take adjacent few pixels into consideration when calculating model matrix, which impact the accuracy of numerical solution. By applying hash encoding, our method interpolated a sampling pixel value from all vertexes of multi-resolution grids [31], which makes a better performance than MB method. With no need of calculating model matrix, our method has a better time complexity than MB method. Moreover, for MB method, the increase of unknowns yields ill-posed problem for high-resolution image reconstruction. Benefiting from the continuous representation of MLP, NR method can reconstruct high-resolution image and not be affected by ill-posed problem. Some learning-based methods, such as AS-Net and diffusion model [27], [28], apply a pre-trained network model to remove artifacts from sparse-view images, which depends on a mass of high-quality training dataset. Compared to these learning-based methods, our method obtains high-quality PA images from sparse RF signals directly with no need of any reconstructed image, which reduces the complexity of training model and the difficulty of building dataset. And the quality of reconstructed image is not affected by dataset.

Simulation and phantom experiments are designed to compare the performance difference between our method and traditional methods. For simulation data (Fig. 3), UBP method suffers streak-type artifacts severely for the sparse-view sampling, which impact the image quality critically. With adding the prior knowledge provided by TV regularization term, MB method makes a better performance than UBP method under the same sampling condition, and improves the SSIM and PSNR of images markedly. NR method make a better performance than both above methods under the same sampling condition. For experimental data (Fig. 4 and Fig. 5), SNR

and CNR are calculated as the metric to measure image quality, and profiles of images are shown to compare the details of images. Consistent with the observation in simulation, NR method suppresses the artifacts in images and recovers images with a higher quality. We further compared the methods performance for diverse shapes under the sparse-view condition (Fig. III-B.2). For the phantom size with a diameter of 1cm, spatial sampling interval is 0.25mm (128 transducer elements) and bigger than half-wavelength (0.15mm), which not satisfy Nyquist criterion [19], so the case can be seen as a sparse sampling condition. Under this condition, NR method reconstruct images with less artifacts and higher SNR than UBP and MB methods.

The superiority of NR method over the compared method can be attribute to (1) the application of PA forward model deduced from PA wave propagation process, which map the collected data into plane coordinates; (2) the synergy between the continuous representation of an image offered by MLP and the external regularization term (TV) in the loss function. Our simulation and experimental results indicate that NR method outperforms traditional PACT reconstruction methods, e.g., UBP and MB, highlighting the value of implicit representation in solving inverse problem for ring-shaped transducer PACT reconstruction, especially for the sparse sampling condition.

While we showed that the NR method has a better performance to sparse-view PACT image reconstruction, it is noteworthy that the results attained under the condition that the medium is homogeneous. We must point that this method currently cannot make an outstanding performance to in-vivo imaging because of the heterogeneous SOS distribution inside the tissue. Moreover, a trained MLP represents a specific heat distribution. For reconstructing different images, the MLP need to be trained under the supervision of collecting RF signal, which increase the time complexity. In the future work, we will further take the effect of heterogeneous medium into consideration and explore a more efficient scheme to decrease the time cost of training MLP.

V. CONCLUSION

In this paper, a framework based on neural representation was proposed, which reconstructed artifact-free PACT image from sparse-view PA data. Combing with the physical mechanism of photoacoustic forward model, a MLP network is trained in self-supervised fashion. We validated the feasibility and performance of our proposed method by comparing other method through simulation data. In *in-vitro* phantom experiments, our method shows superior performance compared with existing reconstruction methods. The neural representation method is still limited by the accuracy of forward model and time complexity of training, which may be further improved by building the propagation model of heterogeneous acoustic medium and finding more efficient ways to representing the plane coordinates. In the future work, we will improve the efficiency of our method and further generalize neural representation method to three dimensions for 3-D PA imaging.

REFERENCES

- [1] M. Xu and L. V. Wang, "Photoacoustic imaging in biomedicine," *Review of scientific instruments*, vol. 77, no. 4, 2006.
- [2] L. V. Wang, "Multiscale photoacoustic microscopy and computed tomography," *Nature photonics*, vol. 3, no. 9, pp. 503–509, 2009.
- [3] A. P. Jathoul, J. Laufer, O. Ogundade, B. Treeby, B. Cox, E. Zhang, P. Johnson, A. R. Pizzey, B. Philip, T. Marafioti *et al.*, "Deep *in vivo* photoacoustic imaging of mammalian tissues using a tyrosinase-based genetic reporter," *Nature Photonics*, vol. 9, no. 4, pp. 239–246, 2015.
- [4] B. Lafci, E. Mercep, J. L. Herraiz, X. L. Deán-Ben, and D. Razansky, "Noninvasive multiparametric characterization of mammary tumors with transmission-reflection optoacoustic ultrasound," *Neoplasia*, vol. 22, no. 12, pp. 770–777, 2020.
- [5] E. Brown, J. Brunker, and S. Bohndiek, "Photoacoustic imaging as a tool to probe the tumour microenvironment," *Disease Models & Mechanisms*, vol. 12, p. dmmm039636, 07 2019.
- [6] S. Gottschalk, O. Degtyaruk, B. Mc Larney, J. Rebling, M. A. Hutter, X. L. Deán-Ben, S. Shoham, and D. Razansky, "Rapid volumetric optoacoustic imaging of neural dynamics across the mouse brain," *Nature biomedical engineering*, vol. 3, no. 5, pp. 392–401, 2019.
- [7] M. Mehrmohammadi, S. Joon Yoon, D. Yeager, and S. Y Emelianov, "Photoacoustic imaging for cancer detection and staging," *Current Molecular Imaging (Discontinued)*, vol. 2, no. 1, pp. 89–105, 2013.
- [8] Y. Wu, J. Kang, W. G. Lesniak, A. Lisok, H. K. Zhang, R. H. Taylor, M. G. Pomper, and E. M. Boctor, "System-level optimization in spectroscopic photoacoustic imaging of prostate cancer," *Photoacoustics*, vol. 27, p. 100378, 2022.
- [9] C. Tian, Z. Xie, M. L. Fabiilli, and X. Wang, "Imaging and sensing based on dual-pulse nonlinear photoacoustic contrast: a preliminary study on fatty liver," *Optics letters*, vol. 40, no. 10, pp. 2253–2256, 2015.
- [10] C. Tian, W. Qian, X. Shao, Z. Xie, X. Cheng, S. Liu, Q. Cheng, B. Liu, and X. Wang, "Plasmonic nanoparticles with quantitatively controlled bioconjugation for photoacoustic imaging of live cancer cells," *Advanced Science*, vol. 3, no. 12, p. 1600237, 2016.
- [11] J. Xia, M. R. Chatni, K. Maslov, Z. Guo, K. Wang, M. Anastasio, and L. V. Wang, "Whole-body ring-shaped confocal photoacoustic computed tomography of small animals *in vivo*," *Journal of biomedical optics*, vol. 17, no. 5, pp. 050506–050506, 2012.
- [12] L. Li, L. Zhu, C. Ma, L. Lin, J. Yao, L. Wang, K. Maslov, R. Zhang, W. Chen, J. Shi, and L. Wang, "Single-impulse panoramic photoacoustic computed tomography of small-animal whole-body dynamics at high spatiotemporal resolution," *Nature Biomedical Engineering*, vol. 1, 05 2017.
- [13] X. Zhu, Q. Huang, L. Jiang, V.-T. Nguyen, T. Vu, G. Devlin, J. Shaima, X. Wang, Y. Chen, L. Ma *et al.*, "Longitudinal intravital imaging of mouse placenta," *Science Advances*, vol. 10, no. 12, p. eadk1278, 2024.
- [14] C. Yeh, L. Li, L. Zhu, J. Xia, C. Li, W. Chen, A. Garcia-Uribe, K. I. Maslov, and L. V. Wang, "Dry coupling for whole-body small-animal photoacoustic computed tomography," *Journal of biomedical optics*, vol. 22, no. 4, pp. 041017–041017, 2017.
- [15] T. Chen, L. Liu, X. Ma, Y. Zhang, H. Liu, R. Zheng, J. Ren, H. Zhou, Y. Ren, R. Gao *et al.*, "Dedicated photoacoustic imaging instrument for human periphery blood vessels: a new paradigm for understanding the vascular health," *IEEE Transactions on Biomedical Engineering*, vol. 69, no. 3, pp. 1093–1100, 2021.
- [16] M. Nishiyama, T. Namita, K. Kondo, M. Yamakawa, and T. Shiina, "Ring-array photoacoustic tomography for imaging human finger vasculature," *Journal of Biomedical Optics*, vol. 24, p. 1, 09 2019.
- [17] 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," *Nature communications*, vol. 9, no. 1, p. 2352, 2018.
- [18] S. Na, J. J. Russin, L. Lin, X. Yuan, P. Hu, K. B. Jann, L. Yan, K. Maslov, J. Shi, D. J. Wang *et al.*, "Massively parallel functional photoacoustic computed tomography of the human brain," *Nature Biomedical Engineering*, vol. 6, no. 5, pp. 584–592, 2022.
- [19] P. Hu, L. Li, L. Lin, and L. V. Wang, "Spatiotemporal antialiasing in photoacoustic computed tomography," *IEEE Transactions on Medical Imaging*, vol. 39, no. 11, pp. 3535–3547, 2020.
- [20] M. Xu and L. V. Wang, "Universal back-projection algorithm for photoacoustic computed tomography," *Phys. Rev. E*, vol. 71, p. 016706, Jan 2005.
- [21] A. Rosenthal, D. Razansky, and V. Ntziachristos, "Fast semi-analytical model-based acoustic inversion for quantitative optoacoustic tomography," *IEEE Transactions on Medical Imaging*, vol. 29, no. 6, pp. 1275–1285, 2010.

- [22] G. Paltauf, J. Viator, S. Prahl, and S. Jacques, "Iterative reconstruction algorithm for optoacoustic imaging," *The Journal of the Acoustical Society of America*, vol. 112, no. 4, pp. 1536–1544, 2002.
- [23] Y. Dong, T. Görner, and S. Kunis, "An algorithm for total variation regularized photoacoustic imaging," *Advances in Computational Mathematics*, vol. 41, pp. 423–438, 2015.
- [24] T. Tong, W. Huang, K. Wang, Z. He, L. Yin, X. Yang, S. Zhang, and J. Tian, "Domain transform network for photoacoustic tomography from limited-view and sparsely sampled data," *Photoacoustics*, vol. 19, p. 100190, 2020.
- [25] S. Guan, A. A. Khan, S. Sikdar, and P. V. Chitnis, "Limited-view and sparse photoacoustic tomography for neuroimaging with deep learning," *Scientific reports*, vol. 10, no. 1, p. 8510, 2020.
- [26] N. Davoudi, X. L. Deán-Ben, and D. Razansky, "Deep learning optoacoustic tomography with sparse data," *Nature Machine Intelligence*, vol. 1, no. 10, pp. 453–460, 2019.
- [27] M. Guo, H. Lan, C. Yang, J. Liu, and F. Gao, "As-net: fast photoacoustic reconstruction with multi-feature fusion from sparse data," *IEEE Transactions on Computational Imaging*, vol. 8, pp. 215–223, 2022.
- [28] X. Song, G. Wang, W. Zhong, K. Guo, Z. Li, X. Liu, J. Dong, and Q. Liu, "Sparse-view reconstruction for photoacoustic tomography combining diffusion model with model-based iteration," *Photoacoustics*, vol. 33, p. 100558, 2023.
- [29] B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng, "Nerf: Representing scenes as neural radiance fields for view synthesis," *Communications of the ACM*, vol. 65, no. 1, pp. 99–106, 2021.
- [30] K. Zhang, G. Riegler, N. Snavely, and V. Koltun, "Nerf++: Analyzing and improving neural radiance fields," *arXiv preprint arXiv:2010.07492*, 2020.
- [31] T. Müller, A. Evans, C. Schied, and A. Keller, "Instant neural graphics primitives with a multiresolution hash encoding," *ACM transactions on graphics (TOG)*, vol. 41, no. 4, pp. 1–15, 2022.
- [32] G. Wang, Z. Chen, C. C. Loy, and Z. Liu, "Sparsenerf: Distilling depth ranking for few-shot novel view synthesis," in *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2023, pp. 9031–9042.
- [33] M. Niemeyer, J. T. Barron, B. Mildenhall, M. S. M. Sajjadi, A. Geiger, and N. Radwan, "Regnerf: Regularizing neural radiance fields for view synthesis from sparse inputs," in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022, pp. 5470–5480.
- [34] L. Shen, J. Pauly, and L. Xing, "Nerp: implicit neural representation learning with prior embedding for sparsely sampled image reconstruction," *IEEE Transactions on Neural Networks and Learning Systems*, 2022.
- [35] Y. Sun, J. Liu, M. Xie, B. Wohlberg, and U. S. Kamilov, "Coil: Coordinate-based internal learning for tomographic imaging," *IEEE Transactions on Computational Imaging*, vol. 7, pp. 1400–1412, 2021.
- [36] Q. Wu, R. Feng, H. Wei, J. Yu, and Y. Zhang, "Self-supervised coordinate projection network for sparse-view computed tomography," *IEEE Transactions on Computational Imaging*, 2023.
- [37] S. Shen, Z. Wang, P. Liu, Z. Pan, R. Li, T. Gao, S. Li, and J. Yu, "Non-line-of-sight imaging via neural transient fields," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 7, pp. 2257–2268, 2021.
- [38] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.
- [39] J. Lubbers and R. Graaff, "A simple and accurate formula for the sound velocity in water," *Ultrasound in medicine & biology*, vol. 24, no. 7, pp. 1065–1068, 1998.
- [40] A. Longo, D. Jüstel, and V. Ntziachristos, "Disentangling the frequency content in optoacoustics," *IEEE Transactions on Medical Imaging*, vol. 41, no. 11, pp. 3373–3384, 2022.
- [41] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE transactions on image processing*, vol. 13, no. 4, pp. 600–612, 2004.
- [42] H. Lan, D. Jiang, C. Yang, F. Gao, and F. Gao, "Y-net: Hybrid deep learning image reconstruction for photoacoustic tomography *in vivo*," *Photoacoustics*, vol. 20, p. 100197, 2020.