Contents lists available at ScienceDirect

Photoacoustics

journal homepage: www.elsevier.com/locate/pacs

A jointed feature fusion framework for photoacoustic image reconstruction[☆]

Hengrong Lan[a,b,1], Changchun Yang[a,1], Fei Gao[a,c,*]

[a] Hybrid Imaging System Laboratory, Shanghai Engineering Research Center of Intelligent Vision and Imaging, School of Information Science and Technology, ShanghaiTech University, Shanghai 201210, China
[b] Department of Biomedical Engineering, School of Medicine, Tsinghua University, Beijing 100084, China
[c] Shanghai Clinical Research and Trial Center, Shanghai 201210, China

ARTICLE INFO

Keywords:
Photoacoustic tomography
Deep learning
Reconstruction
Convolutional neural network

ABSTRACT

The standard reconstruction of Photoacoustic (PA) computed tomography (PACT) image could cause the artifacts due to interferences or ill-posed setup. Recently, deep learning has been used to reconstruct the PA image with ill-posed conditions. In this paper, we propose a jointed feature fusion framework (JEFF-Net) based on deep learning to reconstruct the PA image using limited-view data. The cross-domain features from limited-view position-wise data and the reconstructed image are fused by a backtracked supervision. A quarter position-wise data (32 channels) is fed into model, which outputs another 3-quarters-view data (96 channels). Moreover, two novel losses are designed to restrain the artifacts by sufficiently manipulating superposed data. The experimental results have demonstrated the superior performance and quantitative evaluations show that our proposed method outperformed the ground-truth in some metrics by 135% (SSIM for simulation) and 40% (gCNR for in-vivo) improvement.

1. Introduction

As a hybrid imaging modality, photoacoustic tomography (PAT) has emerged to visualize the chromophores in biological tissue by converting absorbed optical energy into acoustic energy. It has high spatial resolution at deep penetration in tissues. Photoacoustic computed tomography (PACT) is one of the major implementations of PAT, which explores a higher penetrability with large view. Many potential applications have been explored in biomedical imaging areas, such as blood oxygen saturation (sO2) quantification for cancer diagnostics. PACT possesses high temporal resolution by reconstructing a photoacoustic (PA) image with single-shot pulsed laser light and provides potential preclinical and clinical prospects in thyroid cancer, breast cancer diagnostics, and small animal whole-body imaging [1–9]. Standard reconstruction algorithm, e.g. delay-and-sum (DAS), is widely used to rebuild PA image with high frame rate. However, ill-posed conditions, e.g. limited-view and limited elements, could cause poor quality with blurry image and artifacts. Many studies have improved the reconstruction methods to address these issues to some extent [10,11]. These methods improved the quality of PA image paid by increasing the computational complexity of reconstruction.

Recently, deep learning (DL) has emerged to reconstruct the PA image [12,13]. Specially, convolutional neural networks (CNN) have been a great success in the computer vision area. DL enables PACT reconstruction in both image and signal domains. In image domain, a straightforward way of applying DL is to reduce image artifacts as a post-processing step [14–16]. For instance, Austin Reiter identified the point source locations from pre-beamformed PA data using a CNN [17]. Neda Davoudi et al. used a U-net for efficient recovery of image quality from sparse data [18]. Also, DL directly learns the map from PA signals to PA image, which could contain a complex physical procedure [19,20]. However, DL has difficulty in learning the process from signal to image for complex objects, since the model is hard to properly fit the cross-modality mapping between signal and image for complex objects.

[☆] This research was funded by National Natural Science Foundation of China (61805139), United Imaging Intelligence (2019×0203-501-02), Double First-Class Initiative Fund of ShanghaiTech University (2022×0203-904-04), and Shanghai Clinical Research and Trial Center (2022A0305-418-02).

[*] Corresponding author at: Hybrid Imaging System Laboratory, Shanghai Engineering Research Center of Intelligent Vision and Imaging, School of Information Science and Technology, ShanghaiTech University, Shanghai 201210, China.

E-mail addresses: lanhr@shanghaitech.edu.cn (H. Lan), yangchch@shanghaitech.edu.cn (C. Yang), gaofei@shanghaitech.edu.cn (F. Gao).

[1] Hengrong Lan and Changchun Yang contributed equally to this work.

https://doi.org/10.1016/j.pacs.2022.100442
Received 16 November 2022; Accepted 19 December 2022
Available online 20 December 2022

In signal domain, DL is used to recover the bandwidth of PA signals and improve the signal-to-noise ratio (SNR) of PA signals [21,22]. Then, a higher quality image can be rebuilt by standard reconstructed method. Some frameworks extract different features by combined signal and image instead of single domain. In [23,24], the authors proposed a multi-input reconstruction framework by combining signal and image inputs. Steven Guan et al. used pixel-wise delayed data as input of CNN, which includes more positional information [25]. Meanwhile, MinWoo Kim et al. converted raw data into a 3-D array, where additional positional information is considered [26]. In addition, DL inspires iterative reconstruction methods to simplify the adjustment and repeating optimization for the inverse problem by learning the regularization and some parts of the optimization procedure [27,28].

For most artifacts removal and limited-view compensation problems, the artifacts are identified by taking PA image as input, which can be treated as an image denoising task. The missing view can be also compensated by building the relationship of input and output images. Namely, it is feasible that the neural network model can restore the lost information from large training data in image inpainting and enhancement task. Inspired by the input format of [25,26], we make use of the position-wise data as input data and propose a jointed feature network (JEFF-Net) to reconstruct the limited-view PA image and eliminate the artifacts of reconstruction. We define the position-wise data of delay-and-sum (DAS) as a sub-image, which can be superimposed as a PA image. Meanwhile, the superimposed image provides the shape of the object, which can be fused with output position-wise data. Therefore, the common parts of the sub-images, i.e. the object, are extracted from the fused features. In this paper, we demonstrate JEFF-Net using limited-view (a quarter view) PACT data, which are fed into the model and generates delayed data of another 3 quarters positions as Fig. 1(a) shows. Furthermore, an image feature path transforms the output of the 3 quarters' positions and obtains the full-view image in every channel. Compared with the limit-view image, the output result shows superior performance. The ground-truth coming from DAS reconstruction with full-view data, which still contains distorted ingredients in the background. By this detached data arrangement, two novel losses are designed to restrain the artifacts in superposed position-wise data. Specifically, we could surpass the ground-truth in PACT by simple operation, which indicates fewer artifacts and higher contrast results than ground-truth. While JEFF-Net completes the compensation task through usual supervised learning, the image is split into a state where multiple sub-images are superimposed. Meanwhile, the common parts of the sub-images, i.e. the object, are extracted through a novel residual structure. In this paper, we demonstrate JEFF-Net using limited-view (a quarter view) photoacoustic computed tomography (PACT) data. Inspired by input format of data procured by isolated probes [25,26], a quarter position-wise delayed raw data (32-channels) is fed into model and generates delayed data of another 3 quarters positions as Fig. 1(a) shows. The ground-truth coming from DAS reconstruction with full-view but still sparse data could contain distorted ingredients in background due to the limited number of detectors. Specially, by virtue of this detached data arrangement, two novel losses are designed to restrain the artifacts in superposed position-wise data. We first surpass the ground-truth in PACT by designing special supervision and loss, which is embodied in less artifacts, higher contrast than ground-truth.

The numerical experiments are demonstrated to compare the standard full-view DAS reconstruction (ground-truth) with the proposed JEFF-Net. Meanwhile, we also compare the compensated view result with other deep learning models. Moreover, we perform in-vivo imaging experiments of mice abdomen to illustrate the superiority of our method. In quantitative evaluations, the results show better performance compared with ground-truth (0.667 >0.283 SSIM value).

Our contributions can be summarized as follows:

- For the first time, we introduce a DL solution to resolve the limited-view problem in PACT by feeding a quarter position-wise data.
- We try to remove the artifacts in full-view reconstructed result by two steps:
- Two joint feature data are used to fuse the reconstructed results, which caused the transformation of the output, i.e. 3 quarters' data, from position-wise data to reconstruct image by the backtracked supervision.
- Two novel losses are designed to mitigate interference in PA position-wise data.
- We validate our method on both synthetic and in-vivo PACT dataset and compare our method with other models. We further compare our result with ground-truth in quantitative analysis.
- Finally, we share and release the code of our model and the mice dataset, using which other researchers can reproduce and train their DL model.

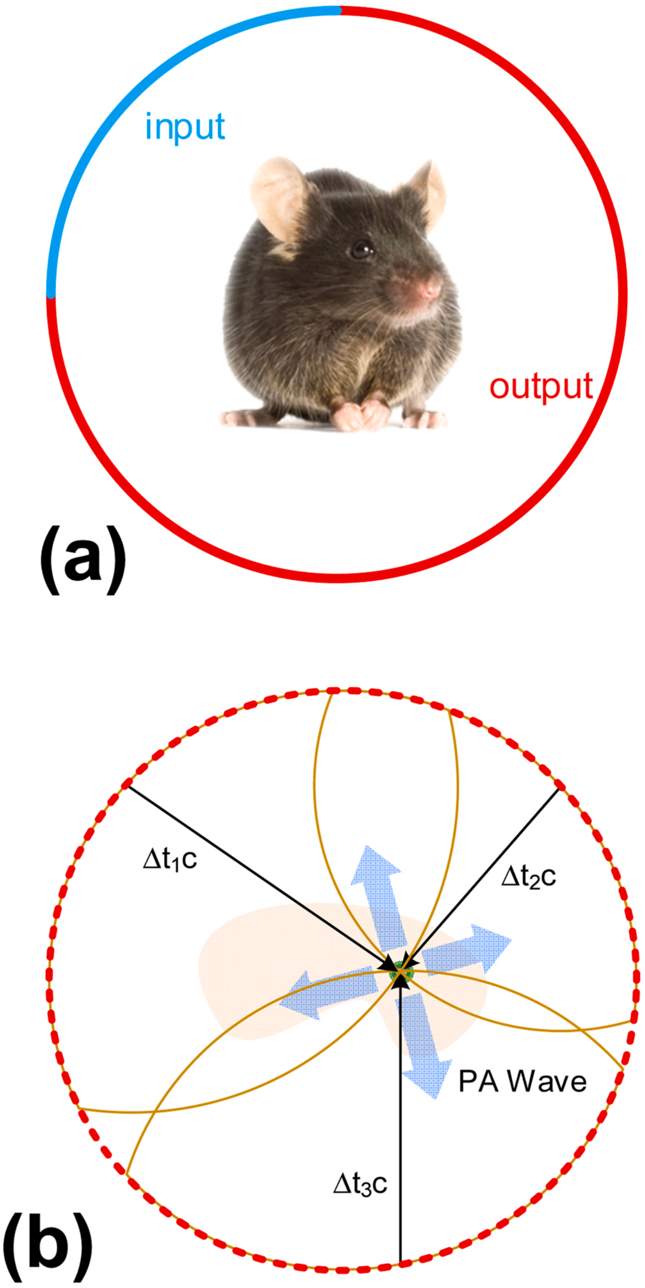


Fig. 1. (a) Illustration of the scanning setup and the input/output views of our task. (b) The propagation of PA signals and the image reconstruction principle; Δt_i indicates the PA wave propagation time from source to detector.

2. Backgrounds

2.1. Photoacoustic computed tomography

In PACT, the initial pressure is excited by the single short laser pulse, which can be expressed as [7]:

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

where Γ_0 is the Gruneisen coefficient, η_{th} is the conversion efficiency from light to heat, μ_a is the optical absorption coefficient, and F is the optical fluence. We use p and b to indicate the initial pressure and the received PA signals. The forward operation can be modeled as a linear operator A:

$$b = Ap, \tag{2}$$

which contains propagation of PA wave in the medium. The PA signals are detected by transducers as shown in Fig. 1(b). The basic idea of reconstruction is to recover p from b. For PACT, the light uniformly illuminates the whole target, which excites the PA signals simultaneously. The transducer array is used to receive the PA data at different positions. In general, the transducer with a large detection angle is desirable to receive PA signals from different directions. Several algorithms are used for PA image reconstruction, within which universal back-projection (UPB) is widely used due to less computational cost and easy implementation. In short, the basic principle of DAS can be depicted in Fig. 1(b), where PA signals are delayed to the region of interest for every channel's data based on the distance between detector and PA source.

2.2. The physical fundamentals of limited-view and artifacts in PACT

An accurate reconstruction could be maintained if the transducer fully encloses the target, and the number of elements has enough spatial density. The transducer often only accesses the PA signals from partial coverage of the tissue due to geometric restrictions. The ill-posed situation could be caused by incomplete enclosed-angle or sparse elements, which degrades image quality or lose important information. Here, we simulated different enclosed views and spatial density of transducer in Fig. 2 [29] (The simulated setup is an ideal case with a homogeneous medium and 20 PML in grid points, which only use a simple case to show as a demonstration). A full circular transducer with enough elements (e.g. 256 elements) produces a superior result as shown in Fig. 2(b). Fig. 2(c) shows that some minor artifacts will be generated if the sensors' number decreases from 256 to 128. Once the enclosed view decreases to half view, more artifacts have emerged, and the target is becoming blurry as Fig. 2(d) shows. A severe situation could happen if we further decrease the angle to a quarter-view (32 channels) as shown in Fig. 2(e). Only part of the object that is close to the sensor array can be reconstructed. Fig. 2(f) shows the result using a 128-elements linear transducer, which is polluted severely by many artifacts.

Similarly, a more serious issue occurs as a result of DAS reconstruction. In standard DAS, the received PA signals are simply delayed to every pixel based on the distance from detector to pixel's position as Fig. 3(a) shows (e.g., the 5th channel with clockwise from left one.), i.e., position-wise data. Since we cannot judge the accurate orientation of PA pressure, pixels of the same radius distance are usually assigned to the same value arbitrarily. We will obtain the reconstructed PA image if we superpose all channels' delayed data as Fig. 3(b) shows. We can divide the delayed data into the data of object d_{oj} and the data of artifact d_{ar}. Fig. 3(c) shows d_{oj} of 5th-channel position-wise data, which is necessary for reconstruction. Fig. 3(d) shows d_{ar} of 5th-channel position-wise data, which is the components of the artifacts, although they are small and irrelevant for all channels. d_{ar} could cause severe interference when the system has a sparse number of detectors. Likewise, the DAS result

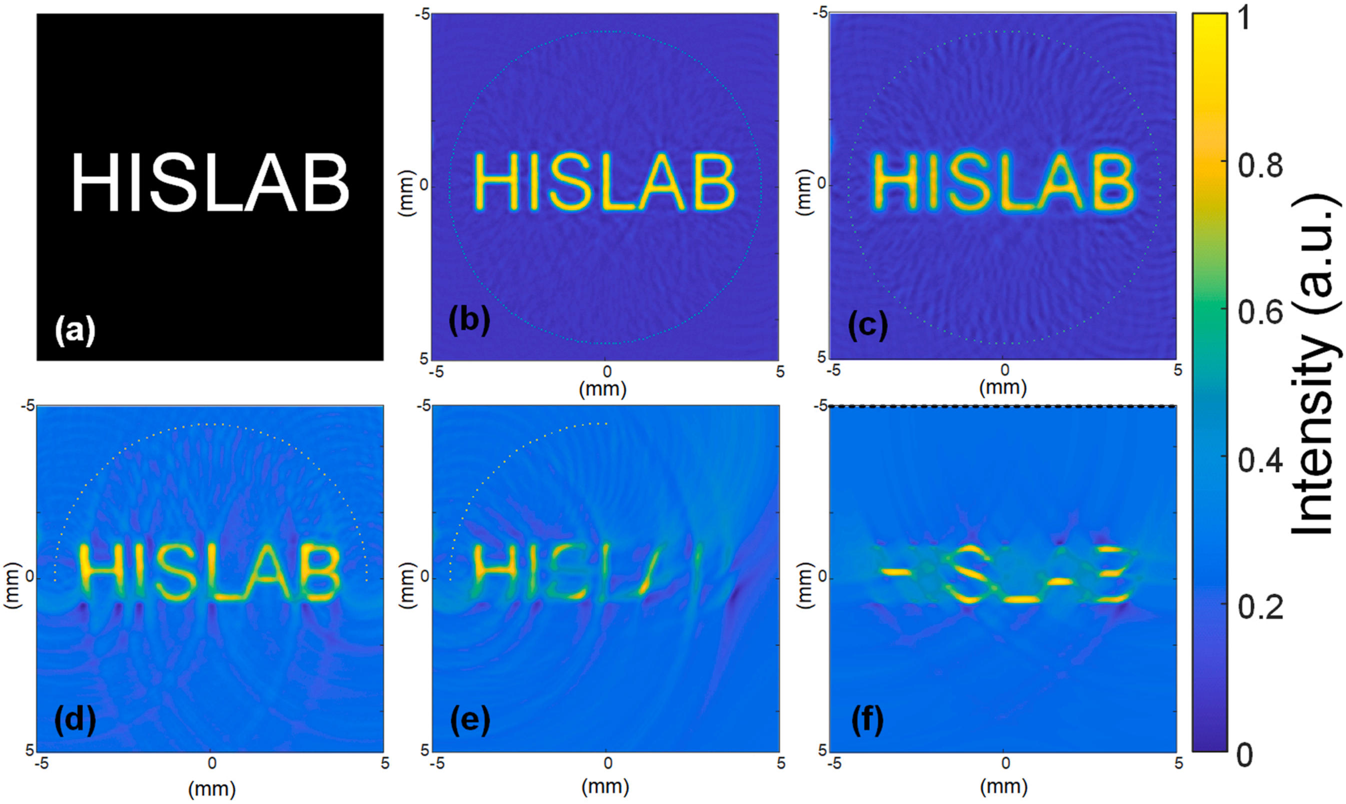


Fig. 2. Simulation results of different limited-view situations. (a) The example object. (b) reconstructed result with 256 enclosed sensors. (c) reconstructed result with 128 enclosed sensors. (d) reconstructed result with 64 half-view sensors. (e) reconstructed result with 32 quarter-view sensors. (f) reconstructed result with 128 line sensors.

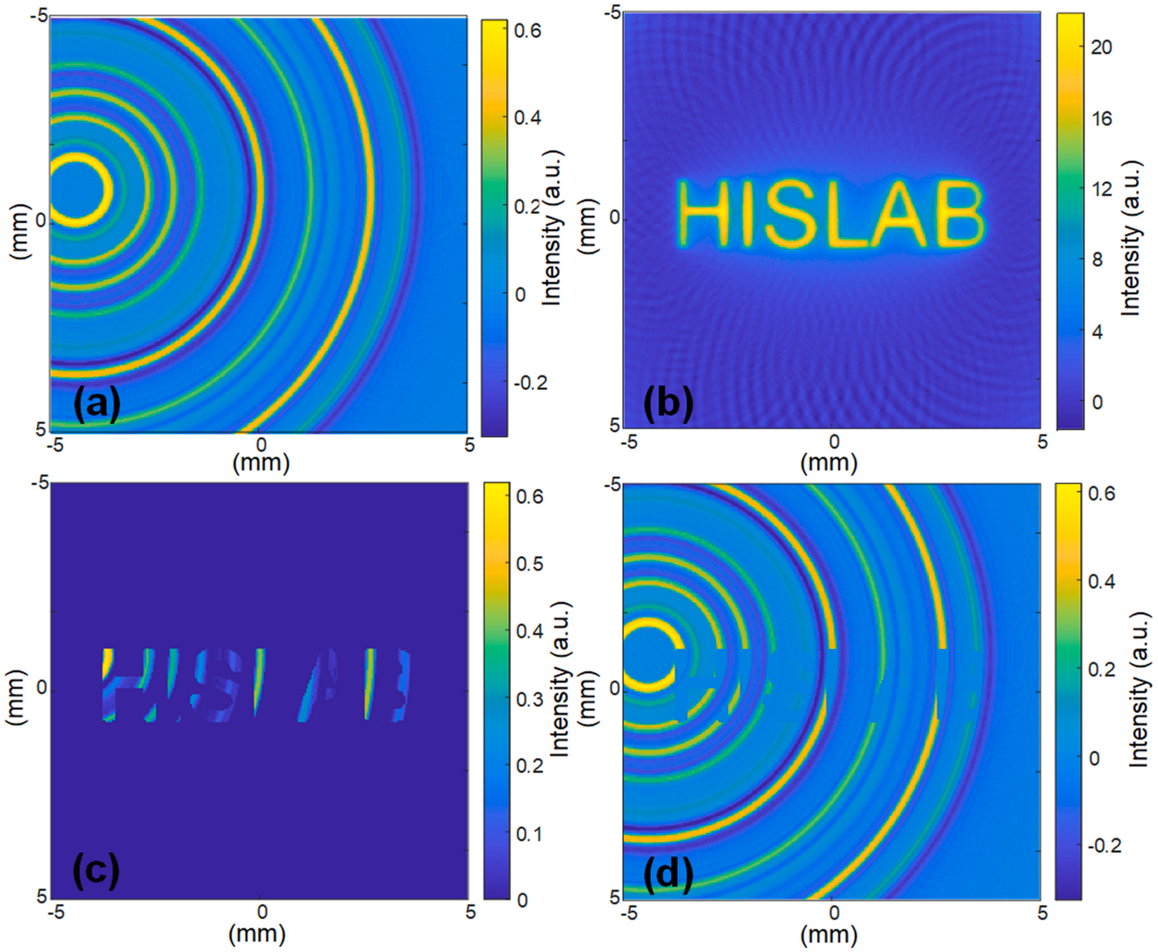


Fig. 3. Simulation results of DAS reconstruction. (a) The 5th-channel position-wise data. (b) The superposed results of 128-channel position-wise data. (c) d_{oj} of 5th-channel. (d) d_{ar} of 5th-channel.

fundamentally consists of p_{oj} and p_{ar}. Generally, d_{oj} and d_{ar} have a similar scale of value, and p_{oj} usually has a larger scale of value than p_{ar}. Considering the same scale of value for position-wise data, it could be more equal to both objects and artifacts if we use position-wise data as the label to supervise model, so that we can strip out an unblemished object.

3. Method

To go beyond supervision, we propose a novel framework to surpass the ground-truth (reconstructed image with a full circular transducer). In this section, we use a CNN model to compensate for the limited-view of PA data and propose space-based calibration and transition module to calibrate the position-wise data. And then, to restrain the output position-wise data, we introduce an image feature path, which provides an image feature (the shape of the object) with backtracked supervision. Finally, we will propose the complete JEFF-Net with two novel losses to achieve a superior result than ground-truth.

3.1. PA Position-wise data for reconstruction

To effectively resolve the limited-view problem, the plain method expects to train an end-to-end network by feeding a limited-view result. This scheme is regarded as image repair to learn the lost information from the data with the powerful deep learning approach, which could ignore the underlying physical meaning of each channel data. Inspired by previous literature [25,26], for the first time to our best knowledge, we introduce a limited-view compensation for position-wise data, which simply delays the raw PA signals to every pixel. In this way, the position-wise data, which contain a location relation between each channel and the respective channel, are treated as inputs of the CNN.

In our work, we demonstrate a quarter view data x with 32 channels as input of our model, which can be denoted by:

$$x = [d_1(h,w), d_2(h,w), \ldots, d_{32}(h,w)] \tag{3}$$

where h and w indicate height and width. The model should generate another 96 channels' data by feeding x. We express this procedure as:

$$G(x) = [d_{33}(h,w), d_{34}(h,w), \ldots, d_{128}(h,w)], \tag{4}$$

where $G(\cdot)$ denotes the DL model, and these 96 channels' data are distributed covering the remaining 270° angles. Finally, we superpose these 128 channels' data, i.e., $\Sigma G(x) + \Sigma x$, and obtain a full-view result of DAS.

3.2. Image feature path for jointed feature fusion

In previous works, most DL-based methods used limited-view PA image to train their end-to-end model, which crudely treats the input as an incomplete image. Considering the weights of the artifacts and the object have significant overlap in single DAS reconstructed image, we cannot simply normalize artifact and object, respectively. These models could not be sensitive to small values due to differences of weight in DAS result. However, the main structure of the object can be boosted in DAS result since p_{oj} is a common part for each channel. On the other hand, position-wise data distribute p_{oj} and p_{ar} in every channel (p_{oj}=Σd_{oj},

p_{ar}=Σd_{ar}), which have similar scale. We use position-wise data to equilibrate the weight between objects and artifacts of every channel.

In general, the post-processing scheme of deep learning enhances the limited-view image, which is restrained by full-view image. Practically, it is difficult to acquire the ground-truth image, so we reconstruct the full-view image with artifacts by the above operation. To improve the quality of output image, we further introduce the image feature path using the 32 channels' superimposed image as input, which guides the transformation of 96 channels' data with the backtracked supervision. We consider this branch separately, the limited-view reconstructed PA image is fed into a CNN model, obtaining the output of full-view image. This scheme is a commonly used post-processing solution to enhance the quality of PA image in ill-condition. We use the full-view DAS image as the ground-truth of this path. Moreover, we combine these two paths and add some additional losses to achieve the feature fusion (these losses will be introduced in the next section). Finally, two different features are used to reconstruct the PA image, which comes from: (1). Same weight between object and artifact of position-wise data; (2). The object with a high weight of reconstructed image.

3.3. Jointed feature fusion framework

As mentioned above, we integrate these structures and introduce a novel JEFF-Net to surpass the quality of ground-truth, which consists of two components as Fig. 4 shows. To fully leverage the benefit of complementary information from highly similar tasks, we proposed space-based calibration and transition module (SCTM) and two novel losses (response loss and overlay loss) to fuse the features and reconstruct the image. We design a backtracked point (BTP) in that two additional losses are used to besiege the output at the positions before and after the output, which is also called backtracked supervision. We desire that these losses can restrain d_{oj} of every channel position-wise data and obtain the artifacts and the negative object.

3.3.1. Space-based calibration and transition module

The above subnetwork generates 96-channel compensation data by feeding x, and SCTM is used to replace the final layer of U-Net. We use SCTM to transfer the angle from 90° to 270°, and calibrates the relation of position-wise. SCTM has been shown in Fig. 4, which connects the encoder features to the decoder.

We design SCTM with a spatially fully-connected layer inspired by Ref. [30]. For the given feature map from the encoder, we first conduct two transformations with max pooling and average pooling. These two feature maps are further fused with grouped fully-connected layer, intended to propagate information of the corresponding position. If the input has m feature maps of size n × n, the grouped fully-connected layer can decrease the number of parameters from mn^4 to $2mn^2$ compared to fully-connected layer. Finally, the feature map should be reshaped to $n \times n \times m$.

3.3.2. JEFF-Net

We further introduce JEFF-Net to integrate the above modules. The overall architecture has been shown in Fig. 4, which can be divided into two pipelines: position-wise data compensation and limited-view image inpainting.

A U-Net [31] model is used for position-wise data compensation, which takes 32-channel position-wise data with a 90° limited view. Given input data with size 128 × 128 × 32, the first four convolutional layers and the following pooling layer (fourth layer without pooling) are used to encode the position-wise data, followed by the final layer with a convolutional layer. SCTM can connect these feature maps to decoder with 8 × 8 × 128 size. SCTM is followed by a series of five up-convolutional layers, generating 96-channel position-wise data with other 270° views. To leverage the image features from limited-view, we use a small network to extract the feature of PA image, named Residual Global Context subnetwork (RGC-Net), which consists of five residual global context layers [32]. Furthermore, we introduce a backtracked supervision before and after BTP, two losses restrain the 96 channels' data output indirectly.

To fully leverage the benefit of complementary information from highly correlated data, we bridge these two results using residual separation. By combining response loss and overlay loss for G(x), the artifacts in the reconstructed result can be learned, which will be described in detail below.

3.3.3. Novel loss for position-wise data path

We train our JEFF-Net by regressing to the ground-truth content of full-view PA image. However, ΣG(x) + Σx has multiple equally plausible

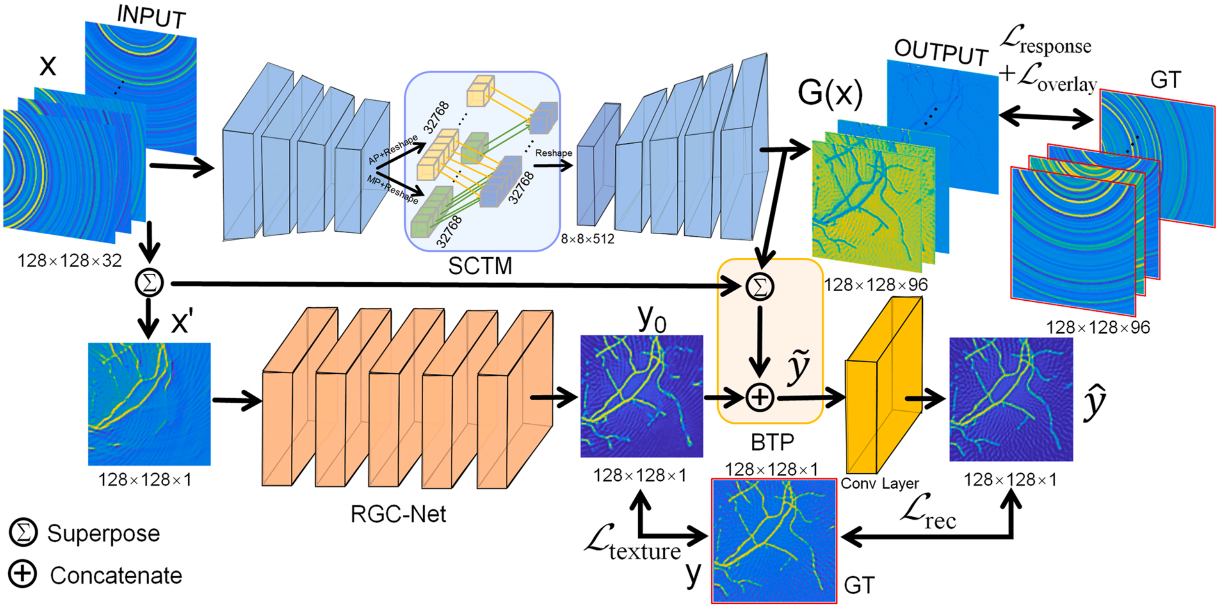


Fig. 4. The overall of proposed JEFF-Net architecture. SCTM: Space-based Calibration and Transition Module. RGC-Net: Residual Global Context subnetwork. BTP: backtracked point. GT: ground-truth. Raw data are pre-reconstructed to 32-channel position-wise data x, and superposing these 32-channel data x' as inputs of network.

ways to satisfy the residual relation. For conventional L1 or L2 loss, they only consider the difference of single position-wise data. However, they ignore the relation of different channel data. We propose response loss and overlay loss to handle both the artifacts and the opposite object in the output.

The compensated delay data we expect to obtain is G(x) in Equation (4), which consists of additional*Nl* (96 in our paper) channels' PA signals. That means this layer *l* (*l* indicates that our loss can be used to the feature maps of arbitrary layer, and it is the final layer G(x) in our work) has *Nl* feature maps, each is with size *Ml*, where *Ml* (128 ×128 in our paper) is the height times the width of the feature map. There are a lot of similarities between the received signals at two adjacent positions. We use overly response to describe the correlation of adjacent view. Although all channels' ultrasonic sensors are spatially and independently placed around the imaging target, their delayed data should have a dependent response relationship at the same position. Hence, we built a response representation that computes the correlations between the different channel's responses, which is the relationship between the PA signal detection channels of each sensor. For given feature maps *D*, the responses between any two channels are given by the Gram matrix [33] $G^l \in \mathcal{R}^{Nl\times Nl}$, where G^l_{ij} is the inner product between the vectorized delay data *i* and *j* in layer *l*:

$$G^l_{ij} = \sum_k D^l_{ik} D^l_{jk}. \tag{5}$$

The Gram matrix is used to measure the dependent response relationship of position-wise data at adjacent positions. Therefore, the response loss is by minimizing the mean-squared distance between the entries of the Gram matrix from the original delay data and generated delay data (A^l and G^l denote their respective response representations):

$$\mathscr{L}_{response} = \frac{1}{4N_l^2 M_l^2}\sum_{i,j}\left(G^l_{ij} - A^l_{ij}\right)^2. \tag{6}$$

Although when acquiring data, a single channel records the signal sequence from its own view, the resultant image is superimposed by the delayed data of all channels. The superposition of arbitrary channels still has a certain dependence. Considering the plainest case, the contribution of the superposition of the arbitrary two views to the full views is measured here. Hence, we built an overlay representation that computes the correlations between the different channels' overlays. We propose an Overlay matrix $O^l \in \mathcal{R}^{Nl\times Nl\times M_l}$ to describe the overlays between vectorized delay data of the two probes *n* and *n*′:

$$O^l_{nn'm} = \sum_{n,n'} D^l_{nm} + D^l_{n'm}. \tag{7}$$

The overlay matrix builds the relation among different detectors, which indicate different view in physics. The superposition of arbitrary channels still has a certain dependence. Similarly, the arbitrary two views to the full views are considered here. So, it causes a 3-D overlay matrix related detector *n* and *n*'. But it is worth noting that our overlay can be extended to 3, 4 or *n* detectors. The overlay loss is by minimizing the mean-squared distance between the entries of the Overlay matrix from the original delay data and generated delay data (P^l and O^l denote their respective overlay representations):

$$\mathscr{L}_{overlay} = \frac{1}{4N_l^2 N_l^2 M_l^2}\sum_{n,n',m}\left(O^l_{nn'm} - P^l_{nn'm}\right)^2. \tag{8}$$

Namely, overlay loss constrains the contribution of each channel to the final image.

And then, we use a texture loss to supervise the limited-view image enhancement. We apply commonly used mean square error (MSE) loss as our texture loss:

$$\mathscr{L}_{texture}(y_0) = \|y - y_0\|_F^2, \tag{9}$$

where *F* denotes the Frobenius norm. Furthermore, a reconstruction loss is used to optimize the residual result by minimizing the mean pixel-wise error:

$$\mathscr{L}_{rec}(\hat{y}) = \|y - \hat{y}\|_F^2. \tag{10}$$

Finally, we define the overall loss function as follow:

$$\begin{aligned}\mathscr{L}_{overall} = {} & \lambda_{re}\mathscr{L}_{\text{response}} + \lambda_{ov}\mathscr{L}_{\text{overlay}} \\ & + \lambda_{tex}\mathscr{L}_{texture} + \lambda_{rec}\mathscr{L}_{rec},\end{aligned} \tag{11}$$

where λ_{re}, λ_{ov}, λ_{tex}, λ_{rec} are hyper-parameters that decide the proportion of every loss, which have different values in different experiments.

4. Experiments

In this section, we validate our method using both simulation and experimental data. Furthermore, some ablation studies are demonstrated. All deep learning methods are implemented on Pytorch [34], which is an open-source framework. The high-speed graphics computing workstation is used to train our model, which consists of four NVIDIA RTX Titan graphics cards. The batch size is set as 16, and iteration is set as 600 epochs. Adam optimization is used with the initial learning rate of 0.005. We also computed that the amount of multiply-add operations (MACs) are 44.72 G, and the number of parameters are 20.39 M. In testing procedure, the computational time of single data is about 0.01 s using a RTX Titan. For single patch data, the forward path should cost about 1266.7 MB GPU memory. Therefore, it can be computed with a proper batch size on most GPUs. For U-Net methods, the MACs and parameters are similar about 13 G and 20.55 M respectively. All experiments are described in detail below. Furthermore, the source code is available at https://github.com/chenyilan/BSR-Net.

4.1. Training on synthetic vessels data

We use the MATLAB toolbox k-Wave [35] to generate the synthetic dataset. The detectors surround the object evenly with 18 mm radius. The center frequency of all sensors is set as 2.5 MHz with 110% fractional bandwidth. The sound speed is 1480 m/s, and the reconstructed region is set as 26 mm × 26 mm.

We use the public fundus oculi vessel DRIVE [36] as initial pressure distribution, which should be expanded by segmentation and combination to increase the data size. We individually compute the position-wise data and concatenate them on a new dimension, which causes a 3D data with 128 × 128 × 128 size. And then we divide this 3D data into two partitions (32 ×128 ×128 and 96 ×128 ×128) along channel dimension. (Same operation for experimental data.) Finally, we have 2800 training sets and 200 test sets. In this experiment, we use 130, 0.02, 42, 60 for λre,λov,λtex,λrec, respectively. Note that all hyper-parameters are chosen for different data empirically, and then we adjust these parameters based on the experimental results.

4.2. Ablation study for different sub-network comparison

In this section, we train an individual G(x) without RGC-Net. Namely, we compare different channel's position-wise data with JEFF-Net. RGC-Net indicates the conventional DL scheme to resolve the limited-view problem in image domain. Noting that G(x) is the final output in this work, we change the output image by combining two sub-networks with the backtracked supervision (it means we get the desired result by constraining the output after the output.). Therefore, we compare two different frameworks with our JEFF-Net: 1. G(x) (above sub-network) with response and overlay loss; 2. RGC-Net (nether sub-network) with texture loss.

4.3. Ablation study for response and overlay loss

One of the contributions in this work is to propose two novel losses. To verify them, we compare the superposed G(x) in four different cases as ablation study: remove these two losses, only use overlay loss, only use response loss, and use these two losses. Therefore, the effect of these two losses could be validated in this experiment.

4.4. Comparisons of different inputs for U-Net

Considering that models trained on the position-wise channel data have better reconstruction performance than the models trained on image-to-image or signal-to-image under limited-view or sparse sensing configuration, we also train another U-net which feeds 32 position-wise channel data with a DAS image ground-truth.

4.5. Comparisons with other methods

We also compare two different methods with our method: 1) a conventional total variation (TV) method; 2) an end-to-end U-Net with limited-view image as input. The DAS results have been processed by thresholding to improve the quality.

4.6. Training on mice data

Last but not least, we also verify the performance of our method on the in vivo data of mice abdomen. A customized PAT system (HISRing, HISLAB Inc., China) is employed to record PA signals, which is equipped with a 128-elements full-view ring-shaped transducer (2.5 MHz, Doppler Inc.). A pulsed laser (720 nm wavelength, 10 Hz repetition rate) is used to illuminate the object by a fiber optic bundle, whose output fluence is 35 mJ/cm^2, and the data sampling rate of our system is 40 MSa/s. The signal is not averaged during data acquisition due to good enough SNR. The region of image reconstruction is 20 mm × 20 mm.

We vertically scanned the mice using our system and obtained 1100 training sets and 116 test sets, which are available at https://ieee-dataport.org/documents/his-ring-abdomen. In this experiment, we use 250, 0.6, 30, 40 for λ_{re}, λ_{ov}, λ_{tex}, λ_{rec} in Equation (11), respectively.

5. Results

5.1. Simulated results

We show two examples of imaging results from the test set in Fig. 5, which compares the limited-view results, full-view results, and three results in the procedure of JEFF-Net. In addition, we simply process G(x) (thresholding for the negative G) and obtain the final result, which is called processed G(x).

We also do same processing for full-view DAS results. Namely, all full-view DAS results have been processed to eliminate negative values by thresholding. The obvious artifacts can be seen in full-view DAS from Fig. 5(c)-(d). Noting that the DAS results used to train JEEF-Net (ground-truth) are not processed by thresholding. The whole objects are recovered from the limited-view input comparing Fig. 5(e)-(f) with Fig. 5(a)-(b). Some of the details are distorted since RGC-Net is not deep enough, which are embodied in the ruptured vessels in the circle of Fig. 5(e)-(f). The vascular structure becomes more complete after the addition of the two paths as Fig. 5(g)-(h) showed. In Fig. 5(i)-(j), the superposed 96-channel data have transformed to the artifacts and the negative object, which should be the position-wise data. G(x) should be further processed to separate objects and artifacts, and here we do simple threshold processing as Fig. 5(k)-(l) showed. We find that TV cannot solve the limited-view issue only with fewer detectors from Fig. 5(m)-(n). U-Net performs a similar result with full-view image since the labeled image is full-view image as Fig. 5(o)-(p) showed.

In our work, the G(x) is final output as Fig. 4 showed. In Fig. 4, we use full-view image y to supervise the y_0, $\hat{y}$, x' and G(x) also are added at BTP. It causes a situation that x'+G(x)= ε ($\varepsilon \approx 0$),so G(x)= -x'+ ε without additional losses as Fig. 8(a) showed. G(x) could be the common parts of 96 channels data (ground-truth) and 32 channels data (x') if we add response and overlay losses. The common parts of 96 channels and x' are the object, which causes G(x) can learn the object only using 96 channels. Therefore, the target could be separated better if we explore a more advanced processing method for G(x), which will be developed in future work. For synthetic data, we have initial pressure distribution even though we do not use it in this task. Namely, we could calculate structural similarity index (SSIM), peak signal-to-noise ratio (PSNR) and root mean square error (RMSE) to quantitatively compare these results. We list all results in Table 1, and we only calculate processed G(x) since superposed G(x) is an opposite image. For average of SSIM, the processed G(x) has 0.667 and the full-view DAS (the ground-truth in our work) only has 0.283. Similarly, for the average of PSNR, the processed

Table 1
The quantitative evaluations of simulated test data. G(x) indicates processed G (x); FV DAS: full-view DAS.

	SSIM	PSNR	RMSE
x'	0.051 ± 0.008	8.233 ± 0.579	0.397 ± 0.037
y_0	0.294 ± 0.031	14.848 ± 1.195	0.182 ± 0.044
$\hat{y}$	0.318 ± 0.035	15.418 ± 1.044	0.178 ± 0.025
G(x)	0.667 ± 0.057	14.269 ± 0.966	0.067 ± 0.053
FV DAS	0.263 ± 0.078	12.484 ± 1.303	0.170 ± 0.020
TV	0.489 ± 0.046	13.127 ± 1.015	0.125 ± 0.032
U-Net	0.327 ± 0.030	14.348 ± 1.052	0.180 ± 0.021

* Small RMSE value indicates high performance

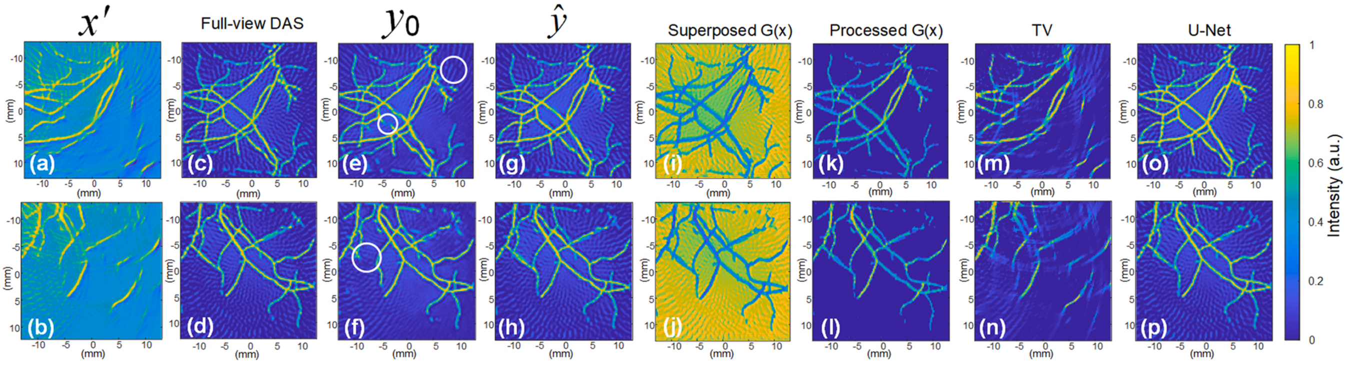


Fig. 5. Two examples of performance comparison of different results. (a,b) Limited-view DAS results. (c,d) Full-view DAS results. (e,f) y_0 results in JEEF-Net. (g,h) $\hat{y}$ results in JEEF-Net. (i,j) G(x) superposed along channel dimension in JEEF-Net. (k,l) G(x) superposed along channel dimension with the thresholding operation. (m,n) TV results with limited-view data (100 iterations). (o,p) end-to-end U-Net results.

G(x) has 15.469 dB and the full-view DAS only has 14.284 dB. It shows a significant superiority of our method from Table 1, which outperforms ground-truth (full-view DAS) in this work. Noting that a small RMSE value also indicates a high performance from Table 1.

5.2. Evaluation of sub-networks

Firstly, we validate the original idea of compensating limited-view position-wise data. We only use a G(·) in Fig. 4 to compensate the view of PA data and plot different channel's data. In Fig. 6, we show different channel of output data and the superposed final image. Every channel of the outputs is the sensor's position-wise data at different position as shown in Fig. 6(a)-(c), and the final image (Fig. 6(d)) can be reconstructed by summing input and output 128 channels' data.

Moreover, one key component of our proposed method is the different feature fusion. Fig. 7 shows the output of RGC-Net, which indicates the image post-processing scheme. Noting that some prevailing end-to-end deep learning solutions for reconstruction are often implemented by arbitrarily changing this backbone [15,17,37,38], which is also a comparative experiment. Fig. 7(a) shows an obvious texture of the object, which decreases the weight of the object in the output position-wise data.

5.3. Ablation study results

The superposed G(x) of four ablation studies have been shown in Fig. 8. G(x) contains both objects and artifacts with small overall value if we do not supervise G(x) as Fig. 8(a) shows, which can be regarded as a supplement to y_0. G(x) will be closer to y with only one loss shown in Fig. 8(b) and (c). These two losses are designed to focus on different characteristics of position-wise data (overlay loss focuses on the relation of each channel data; response loss focuses on the common area d_{oj} of each channel data). We further compare the quantitative results of these methods in Table 2. The overlay loss is proposed to constrain the contribution of each channel to the final image, which could emphasize the common parts of different channel (the object). On the other hands, the response loss can surpass the small values after overlaying, and the big value will become bigger. Therefore, the result of only using response loss has a higher contrast compared with that only using overlay loss. Obviously, the object and artifact can be separated into different ranges only when we use two losses simultaneously (negative objects and positive artifacts).

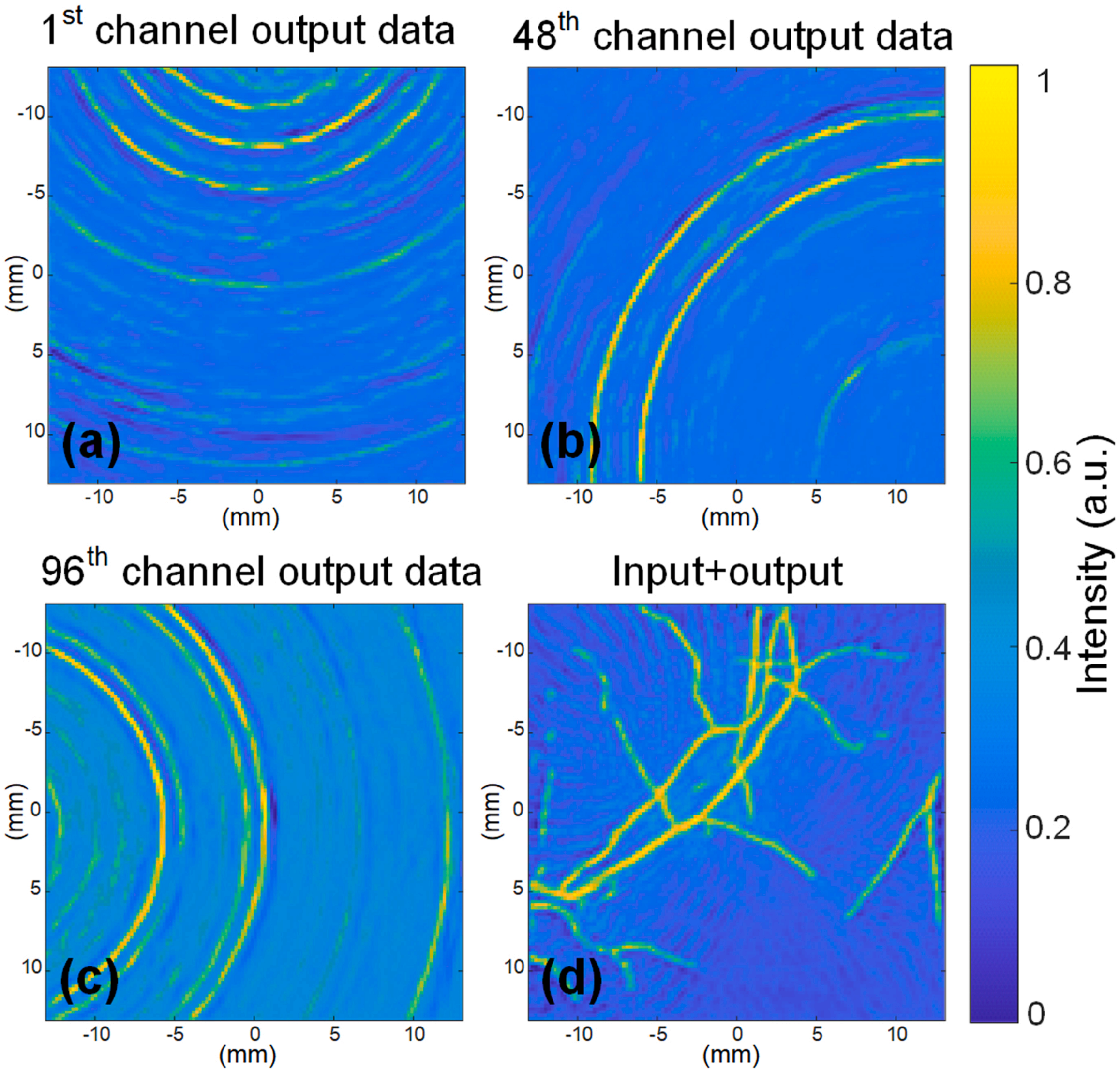


Fig. 6. The different channel of output data and the superposed data of G(·) without residual structure. (a) The 1st channel output data. (b) The 48th channel output data. (c) The 96th channel output data. (d) The sum of superposed input and superposed output data.

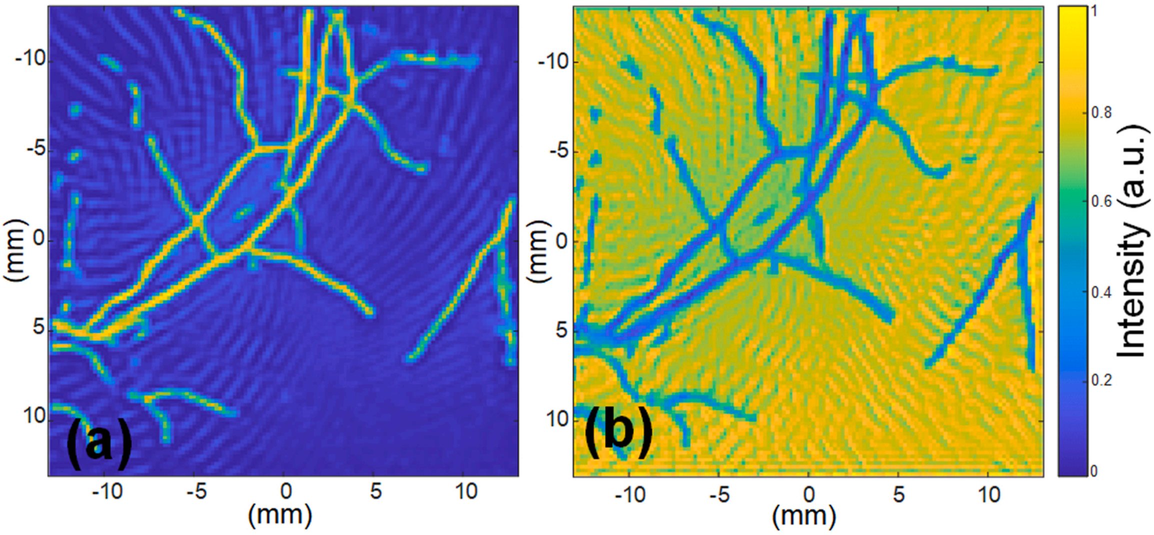


Fig. 7. Results of different sub-network. (a) The result of RGC-Net, the input is a limited-view PA image; (b). The result of superposed G(x).

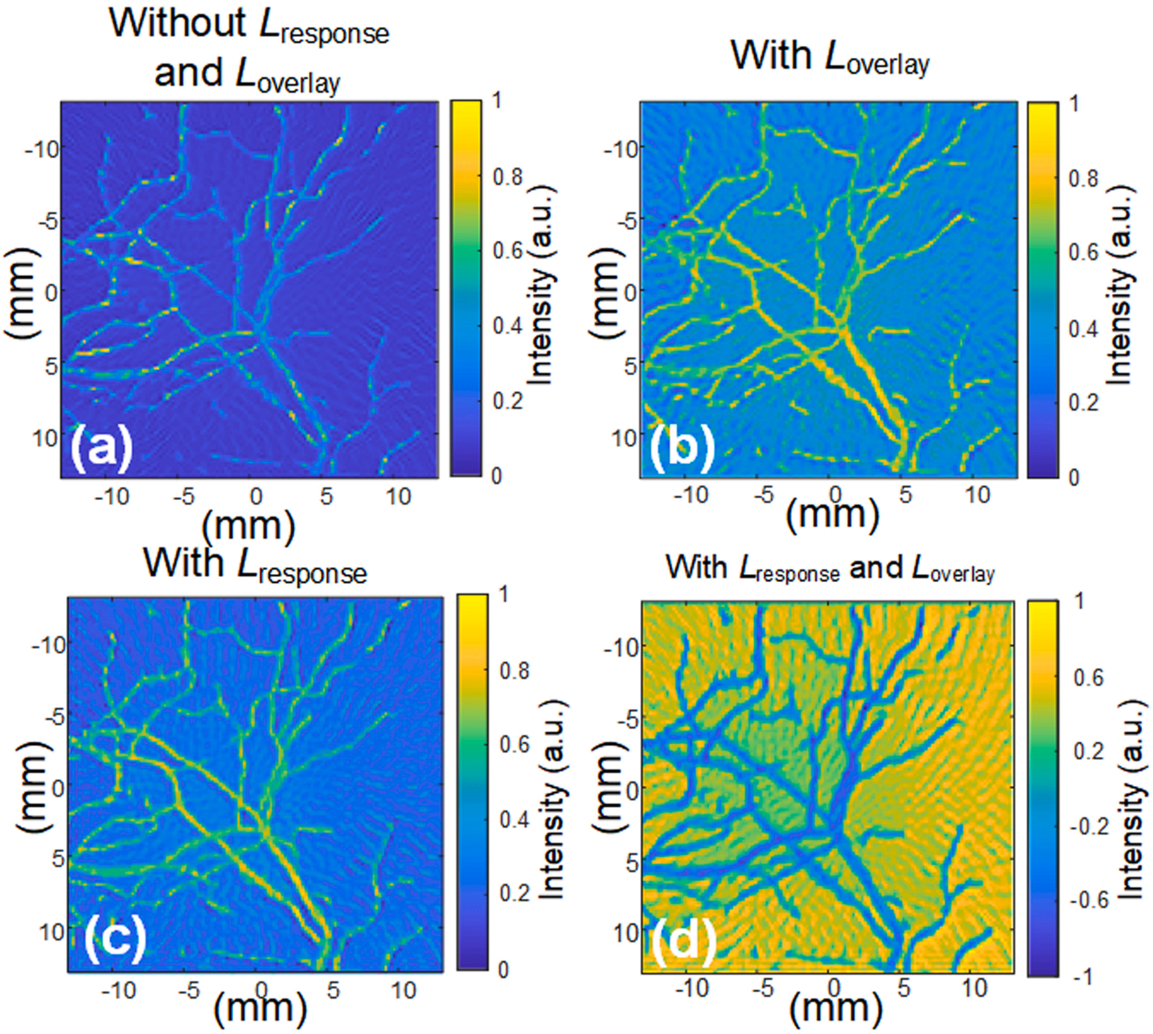


Fig. 8. The G(x) results of ablation study for two novel losses. (a) The result without response loss and overlay loss. (b) The result with overlay loss. (c) The result with response loss. (d) The result with both response loss and overlay loss.

Table 2
The quantitative evaluations of different losses.

	SSIM	PSNR	RMSE
w/o L_{re} , L_{ov}	0.203 ± 0.016	12.070 ± 1.169	0.251 ± 0.033
L_{ov}	0.245 ± 0.019	12.630 ± 1.053	0.235 ± 0.028
L_{re}	0.258 ± 0.038	12.243 ± 1.002	0.246 ± 0.028

* Small RMSE value indicates high performance

5.4. Different inputs for U-Net

We further compare two U-Nets trained on different inputs. Fig. 9 shows a typical result of three different methods, U-Net^1 denotes the input is 32 position-wise data and U-Net^2 denotes the input is a limited-view image with 32 channels. Our approach shows a higher quality after processing. Two different input types do not show significant difference from Fig. 9. Therefore, we also compute the quantitative results of U-Net^1 for test set: 0.331 ± 0.027 (SSIM), 14.738 ± 1.113 (PSNR), 0.184 ± 0.022 (RMSE). By comparing with Table 1, it maintains a slight difference no matter what input types are for U-Net. Therefore, we only compare U-Net^2 in other comparison studies.

5.5. In-vivo results

Lastly, we demonstrated an in-vivo imaging result in Fig. 10, which is the abdomen of a mouse. We also show the limited-view result with 32 detectors, the full-view result with 128 detectors, three results of JEFF-Net and the processed G(x).

Only a minor structure can be found in the limited-view result, and we even cannot recognize the outline of the object from Fig. 10(a). However, an explicit outline can be recognized from the full-view DAS result as Fig. 10(b) shows. All results from JEFF-Net show higher contrast in Fig. 10(c)-(f). The complexity of the in-vivo experimental condition leads to the interference of imaging objects and artifacts compared with synthetic data before, which could cause a little difference between y_0 and$\hat{y}$as shown in Fig. 10(c) and (d). It is interesting that the artifacts of G(x) are relieved due to the residual structure in Fig. 10 (e). Similarly, we further enhance Fig. 10(e) by threshold processing shown in Fig. 10 (f), which clearly shows the profile of the mouse abdomen with a high contrast. Fig. 10(g) shows TV can remove most artifacts and remain similar structure as Fig. 10 (a). U-Net also shows a similar structure in Fig. 10(h).

We can compare all learning-based results with contrast-to-noise ratio (CNR) in Fig. 11, and it shows the best performance for our method. Finally, we select three regions where foreground and background can be clearly distinguished from outline to inside. We further compute generalized contrast-to-noise ratio (gCNR) of Fig. 11 in region 1 and 3 as Table 3 shown. The expressions of CNR and gCNR can be found in [39]. G(x) has a higher value of gCNR, although y and y_0 could be better in some regions, which could be caused by incomplete artifacts reconstruction in y and y_0.

6. Discussions

For some clinical applications, full-view detection is impossible, and limited view issue is inevitable, such as imaging of thyroid nodule or sentinel lymph node detection. So, it is a major significance from the application perspective. Although, more sensors can provide more information of image in limited-view condition, we still think high-quality image reconstruction with fewer sensors is meaningful for such limited-view application.

U-Net is an appropriate framework for most image-to-image tasks, which also shows a superior performance compared with other models. Therefore, we chose U-Net to generate the other views' data by feeding 90° view' position-wise data, then modified it more suitable for our task.

For a 2D PA data, we individually compute the position-wise data and concatenate them on a new dimension, which causes a 3D position-wise data with 128 × 128 × 128 size. And then we divide this 3D position-wise data into two partitions (32 ×128 ×128 and 96 ×128 ×128) along channel dimension. For the 3D input (32 ×128 ×128) from 2D PA data, input data need 32 × 128 × 128 × 4 = 2097152 Byte = 2 MB memory (float32 type). The 3D output (96 ×128 ×128) need 96 × 128 × 128 × 4 = 6291456 Byte = 6 MB memory. For single patch data, the forward path should cost about 1266.7 MB GPU memory. We further consider a 3D PA data, and we do same operation and obtain a 4D data, the consumption is linearly increasing with the number of elevation detectors for cylindrical transducer. Therefore, it could be a problem for GPU memory if we have a dense 3D data.

In numerical experiment, we used k-Wave to generate dataset, which indicate the initial PA pressure distribution can be obtained. However, we assumed a practical circumstance that we cannot obtain better image than full-view image. Therefore, in numerical experiment, we used full-view image as ground-truth to simulate the experimental circumstance.

7. Conclusion

The limited-view is a non-negligible issue in PACT system for many applications (e.g. carotid artery, breast), which usually causes artifacts in the PA image reconstruction. In this paper, we introduce a novel

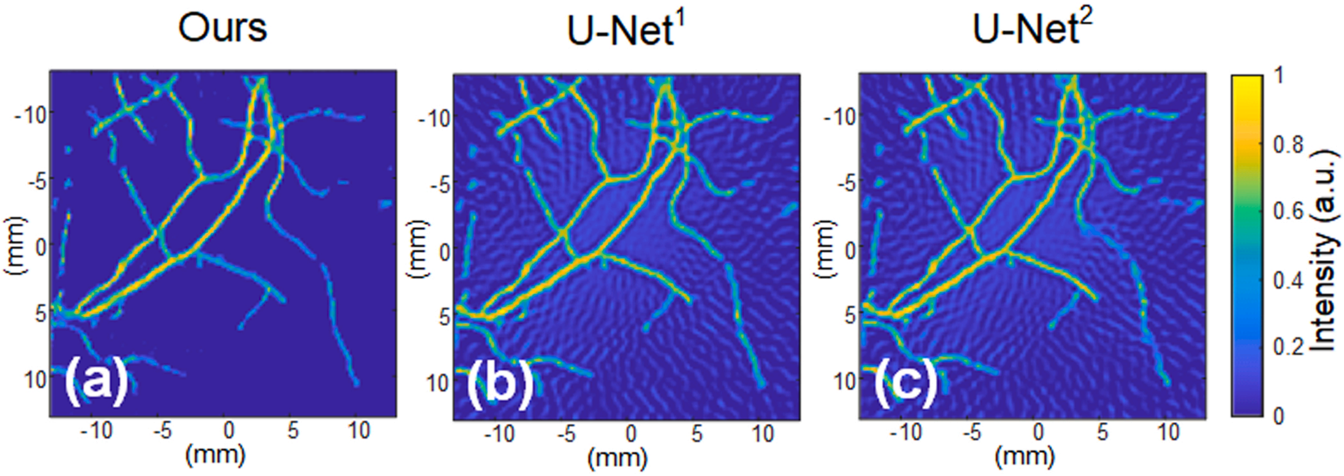


Fig. 9. Imaging results for U-Net with different inputs. (a) Our processed G(x). (b) U-Net with 32 position-wise data as input. (c) U-Net with 32 channels image as input.

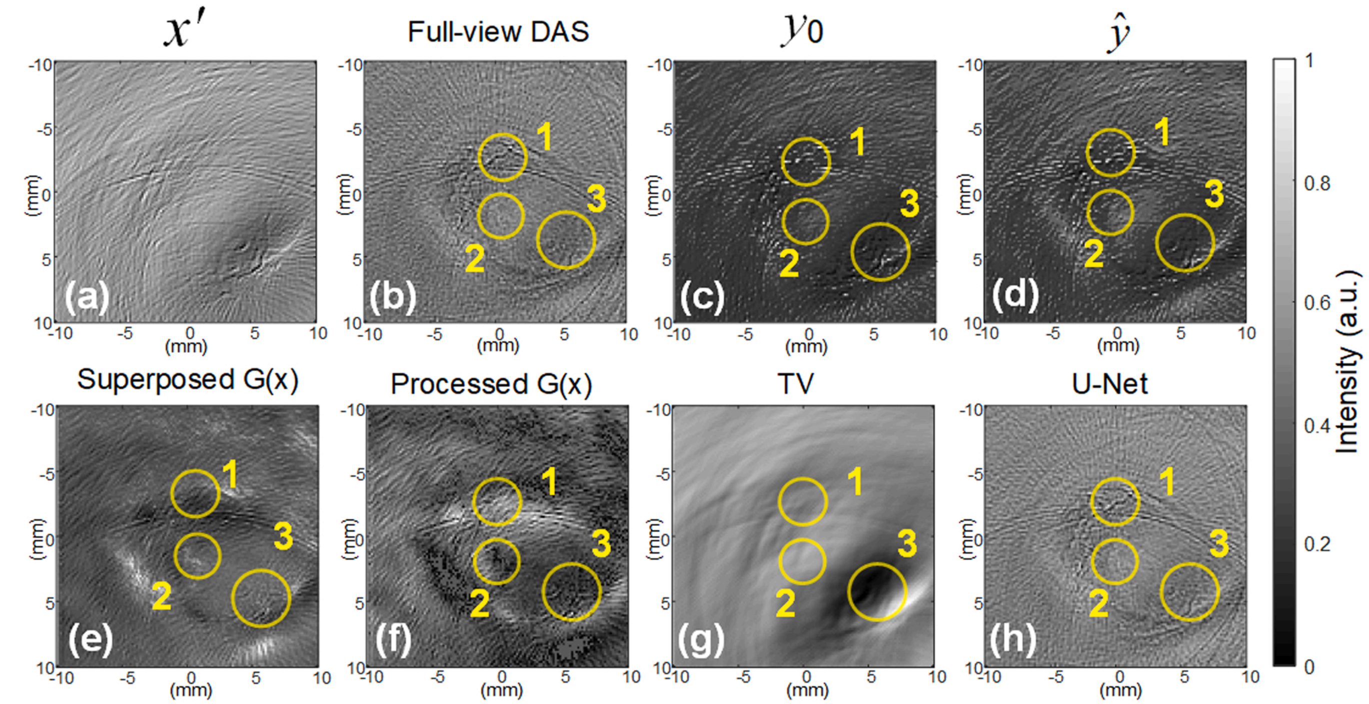


Fig. 10. The in-vivo results, (a) Limited-view DAS result. (b) Full-view DAS result. (c) y_0 result in JEFF-Net. (d) $\hat{y}$ result in JEFF-Net. (e) G(x) superposed along channel dimension in JEFF-Net. (f) G(x) superposed along channel dimension with thresholding operation. (g) TV result with limited-view data (100 iterations). (h) end-to-end U-Net result.

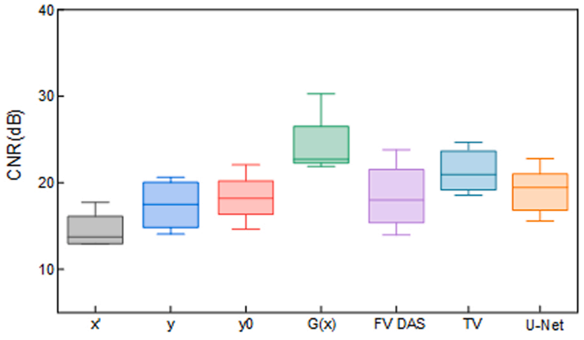


Fig. 11. The CNR of Fig. 10 in-vivo data. G(x) indicates processed G(x); FV DAS: full-view DAS.

Table 3
The gCNR of region 1 and 3 of Fig. 10.

	x`	y_0	$\hat{y}$	G (x)	FV-DAS	TV	U-Net
1	0.351	0.351	0.431	0.477	0.340	0.454	0.362
3	0.357	0.409	0.403	0.380	0.374	0.413	0.391

framework to reconstruct the limited-view PA image: limited position-wise data are used as input of deep learning model, and generate the position-wise data of lost view. We use two different data to fuse the jointed feature for object and artifacts. To further fuse these features, a backtracked supervision is proposed, which adds redundant supervisions before and after G(x). This method may inspire more research fields such as image de-noising, and foreground separation. Furthermore, we proposed two novel losses to constrain the position-wise output. Therefore, we can remove the artifacts by a simple threshold processing. In our work, we propose JEFF-Net implement the proposed framework. A quarter view data is fed into the model, which outputs a group of full-view data. The numerical and in-vivo imaging results show that our methods have good performance compared with other models, even to ground-truth. Finally, we have also published our data and codes to facilitate other researchers for further research. It is worth noting that G(x) can be further used to extract more information, although we only use a threshold processing in this paper, which will be explored in future work.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Fei Gao reports financial support was provided by United Imaging Intelligence. Fei Gao reports financial support was provided by National Natural Science Foundation of China. Fei Gao reports financial support was provided by Shanghai Clinical Research and Trial Center.

References

[1] S. Mallidi, G.P. Luke, S. Emelianov, Photoacoustic imaging in cancer detection, diagnosis, and treatment guidance, Trends Biotechnol. vol. 29 (5) (2011) 213–221.

[2] J.G. Laufer, E.Z. Zhang, B.E. Treeby, B.T. Cox, P.C. Beard, P. Johnson, B. Pedley, In vivo preclinical photoacoustic imaging of tumor vasculature development and therapy, J. Biomed. Opt. vol. 17 (5) (2012), 056016.

[3] I. Steinberg, D.M. Huland, O. Vermesh, H.E. Frostig, W.S. Tummers, S.S. Gambhir, Photoacoustic clinical imaging, Photoacoustics vol. 14 (2019) 77–98.

[4] C. Bench, A. Hauptmann, B.T. Cox, Toward accurate quantitative photoacoustic imaging: learning vascular blood oxygen saturation in three dimensions, J. Biomed. Opt. vol. 25 (8) (2020), 085003.

[5] J. Lv, Y. Xu, L. Xu, L. Nie, Quantitative functional evaluation of liver fibrosis in mice with dynamic contrast-enhanced photoacoustic imaging, Radiology no. 1 (2021) 89–97.

[6] L.V. Wang, S. Hu, Photoacoustic tomography: in vivo imaging from organelles to organs, 1458-62, Mar 23, Science vol. 335 (6075) (2012), https://doi.org/10.1126/science.1216210.

[7] L.V. Wang, Tutorial on photoacoustic microscopy and computed tomography, IEEE J. Sel. Top. Quantum Electron. vol. 14 (1) (2008) 171–179, https://doi.org/10.1109/jstqe.2007.913398.

[8] Y. Zhou, J. Yao, L.V. Wang, Tutorial on photoacoustic tomography, J. Biomed. Opt. vol. 21 (6) (2016) 61007, https://doi.org/10.1117/1.JBO.21.6.061007.

[9] H. Zhong, T. Duan, H. Lan, M. Zhou, F. Gao, Review of low-cost photoacoustic sensing and imaging based on laser diode and light-emitting diode, Sensors vol. 18 (7) (2018), https://doi.org/10.3390/s18072264.

[10] X. Ma, C. Peng, J. Yuan, Q. Cheng, G. Xu, X. Wang, et al., Multiple delay and sum with enveloping beamforming algorithm for photoacoustic imaging, 1-1, IEEE Trans. Med. Imaging (2019), https://doi.org/10.1109/tmi.2019.2958838.

[11] I. Steinberg, J. Kim, M.K. Schneider, D. Hyun, A. Zlitni, S.M. Hooper, et al., Superiorized photo-acoustic non-negative reconstruction (SPANNER) for clinical photoacoustic imaging, IEEE Trans. Med Imaging (2021), https://doi.org/10.1109/TMI.2021.3068181.

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

[13] C. Yang, H. Lan, F. Gao, F. Gao, Review of deep learning for photoacoustic imaging, Photoacoustics vol. 21 (2021), 100215.

[14] D. Waibel, J. Gröhl, F. Isensee, T. Kirchner, K. Maier-Hein, and L. Maier-Hein, "Reconstruction of initial pressure from limited view photoacoustic images using deep learning," in Photons Plus Ultrasound: Imaging and Sensing 2018, 2018, vol. 10494: International Society for Optics and Photonics, p. 104942S.

[15] T. Lu, T. Chen, F. Gao, B. Sun, V. Ntziachristos, J. Li, LV-GAN: a deep learning approach for limited-view optoacoustic imaging based on hybrid datasets, J. Biophoton. (2020), https://doi.org/10.1002/jbio.202000325.

[16] S. Guan, A.A. Khan, S. Sikdar, P.V. Chitnis, Fully dense UNet for 2-D sparse photoacoustic tomography artifact removal, IEEE J. Biomed. Health Inform. vol. 24 (2) (2019) 568–576.

[17] A. Reiter and M.A.L. Bell, A machine learning approach to identifying point source locations in photoacoustic data, in Photons Plus Ultrasound: Imaging and Sensing 2017, 2017, vol. 10064: International Society for Optics and Photonics, p. 100643J.

[18] N. Davoudi, X.L. Deán-Ben, D. Razansky, Deep learning optoacoustic tomography with sparse data, Nat. Mach. Intell. vol. 1 (10) (2019) 453–460.

[19] T. Tong, W. Huang, K. Wang, Z. He, L. Yin, X. Yang, et al., Domain transform network for photoacoustic tomography from limited-view and sparsely sampled data, Photoacoustics (2020), 100190.

[20] H. Lan, C. Yang, D. Jiang, F. Gao, Reconstruct the Photoacoustic Image Based On Deep Learning with Multi-frequency Ring-shape Transducer Array, in 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2019: IEEE, pp. 7115–7118.

[21] N. Awasthi, R. Pardasani, S.K. Kalva, M. Pramanik, P.K. Yalavarthy, Sinogram super-resolution and denoising convolutional neural network (SRCN) for limited data photoacoustic tomography, arXiv preprint arXiv:2001.06434, 2020.

[22] N. Awasthi, G. Jain, S.K. Kalva, M. Pramanik, P.K. Yalavarthy, Deep neural network based sinogram super-resolution and bandwidth enhancement for limited-data photoacoustic tomography, IEEE Trans. Ultrason., Ferroelectr., Freq. Control (2020).

[23] H. Lan, D. Jiang, C. Yang, F. Gao, F. Gao, Y-Net: hybrid deep learning image reconstruction for photoacoustic tomography in vivo, Photoacoustics vol. 20 (2020), 100197.

[24] H. Lan, K. Zhou, C. Yang, J. Cheng, J. Liu, S. Gao et al., "Ki-GAN: Knowledge Infusion Generative Adversarial Network for Photoacoustic Image Reconstruction In Vivo," in Medical Image Computing and Computer Assisted Intervention – MICCAI 2019, (Lecture Notes in Computer Science, 2019, ch. Chapter 31, pp. 273–281.

[25] S. Guan, A.A. Khan, S. Sikdar, P.V. Chitnis, Limited-view and sparse photoacoustic tomography for neuroimaging with deep learning, Sci. Rep. vol. 10 (1) (2020) 1–12.

[26] M.W. Kim, G.-S. Jeng, I. Pelivanov, M. O'Donnell, Deep-learning image reconstruction for real-time photoacoustic system, IEEE Trans. Med. Imaging (2020).

[27] Y.E. Boink, S. Manohar, C. Brune, A partially-learned algorithm for joint photo-acoustic reconstruction and segmentation, IEEE Trans. Med. Imaging vol. 39 (1) (2019) 129–139.

[28] A. Hauptmann, F. Lucka, M. Betcke, N. Huynh, J. Adler, B. Cox, et al., Model-based learning for accelerated, limited-view 3-D photoacoustic tomography, IEEE Trans. Med Imaging vol. 37 (6) (2018) 1382–1393, https://doi.org/10.1109/TMI.2018.2820382.

[29] C. Tian, M. Pei, K. Shen, S. Liu, Z. Hu, T. Feng, Impact of system factors on the performance of photoacoustic tomography scanners, Phys. Rev. Appl. vol. 13 (1) (2020), 014001.

[30] D. Pathak, P. Krahenbuhl, J. Donahue, T. Darrell, and A.A. Efros, Context encoders: Feature learning by inpainting, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 2536–2544.

[31] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in: International Conference on Medical image computing and computer-assisted intervention, 2015: Springer, pp. 234–241.

[32] Y. Cao, J. Xu, S. Lin, F. Wei, H. Hu, Gcnet: Non-local networks meet squeeze-excitation networks and beyond, in: Proceedings of the IEEE International Conference on Computer Vision Workshops, 2019, pp. 0–0.

[33] L.A. Gatys, A.S. Ecker, and M. Bethge, A neural algorithm of artistic style, arXiv preprint arXiv:1508.06576, 2015.

[34] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, et al., Pytorch: an imperative style, high-performance deep learning library, Adv. Neural Inf. Process. Syst. (2019) 8026–8037.

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

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

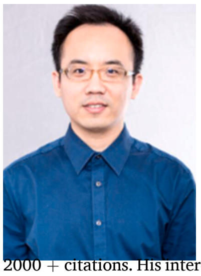
[37] S. Antholzer, M. Haltmeier, R. Nuster, J. Schwab, Photoacoustic image reconstruction via deep learning, in: Photons Plus Ultrasound: Imaging and Sensing 2018, 2018, vol. 10494: International Society for Optics and Photonics, p. 104944U.

[38] T. Vu, M. Li, H. Humayun, Y. Zhou, J. Yao, A generative adversarial network for artifact removal in photoacoustic computed tomography with a linear-array transducer, Exp. Biol. Med. vol. 245 (7) (2020) 597–605.

[39] K.M. Kempski, M.T. Graham, M.R. Gubbi, T. Palmer, M.A.L. Bell, Application of the generalized contrast-to-noise ratio to assess photoacoustic image quality, Biomed. Opt. Express 11 (7) (2020) 3684–3698.

Hengrong Lan received his bachelor degree in Electrical Engineering from Fujian Agriculture and Forestry University in 2017, and PhD degree in Microelectronics and Solid-State Electronics from University of Chinese Academy of Sciences in 2022. He is currently a post-doctor in School of Medicine, Tsinghua University, Beijing, China. His research topic is mainly focused on applying deep learning to photoacoustic/ultrasound imaging. During past 5 years, as first author, he has published more than 10 papers in PACS, IEEE TCI, IEEE Sensors Journal, IEEE JSTQE, BOE, OL, MICCAI, IUS, EMBC, etc. and numerous papers as co-authors, too. These researches have accumulated more than 500 citations from Google Scholar. Hengrong is also actively engaging in biomedical translational study and technology transfer. Collaborated with clinicians, he has developed small-animal photoacoustic tomography prototype, which has performed many phantom and in vivo imaging applications.

Changchun Yang received his bachelor's degree in computer science from Huazhong University of Science and Technology in 2018 and his master's degree in computer science in ShanghaiTech University in 2021. And he is pursuing his PhD in Delft University of Technology. His research interest is interpretable representation learning for medical image analysis, especially for quantitative cardiac MRI.



Fei Gao received his bachelor degree in Microelectronics from Xi'an Jiaotong University in 2009, and PhD degree in Electrical and Electronic Engineering from Nanyang Technological University, Singapore in 2015. He joined School of Information Science and Technology, ShanghaiTech University as an assistant professor in Jan. 2017, and established Hybrid Imaging System Laboratory (www.hislab.cn). During his PhD study, he has received Chinese Government Award for Outstanding Self-Financed Students Abroad (2014). His PhD thesis was selected as Springer Thesis Award 2016. He is currently serving as editorial board member of Photoacoustics. He has published about 130 journal and conference papers with 2000 + citations. His interdisciplinary research topics include photoacoustic (PA) imaging physics (proposed passive PA effect, PA resonance imaging, phase-domain PA sensing, pulsed-CW hybrid nonlinear PA imaging, TRPA-TRUE focusing inside scattering medium, etc.), biomedical circuits and systems (proposed miniaturization methods of laser source and ultrasound sensors, delay-line based DAQ system, hardware acceleration for PA imaging, etc.), algorithm and AI (proposed frameworks such as Ki-GAN, AS-Net, Y-Net, EDA-Net, DR2U-Net, etc,), as well as close collaboration with doctors to address unmet clinical needs.