


Image restoration for ring-array photoacoustic tomography based on an attention mechanism driven conditional generative adversarial network

Wende Dong^{a,1} , Yanli Zhang^{a,1}, Luqi Hu^a, Songde Liu^{b,d,*}, Chao Tian^{b,c,d,e,*}

^a College of Automation Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China

^b School of Engineering Science, University of Science and Technology of China, Hefei, Anhui 230026, China

^c Institute of Artificial Intelligence, Hefei Comprehensive National Science Center, Hefei, Anhui 230088, China

^d Department of Anesthesiology, the First Affiliated Hospital of USTC, Division of Life Sciences and Medicine, University of Science and Technology of China, Hefei, Anhui 230001, China

^e Anhui Province Key Laboratory of Biomedical Imaging and Intelligent Processing, Institute of Artificial Intelligence, Hefei Comprehensive National Science Center, Hefei 230088, China

ARTICLE INFO

Keywords:

Photoacoustic tomography
Image restoration
Attention mechanism

ABSTRACT

Ring-Array photoacoustic tomography (PAT) systems have shown great promise in non-invasive biomedical imaging. However, images produced by these systems often suffer from quality degradation due to non-ideal imaging conditions, with common issues including blurring and streak artifacts. To address these challenges, we propose an image restoration method based on a conditional generative adversarial network (CGAN) framework. Our approach integrates a hybrid spatial and channel attention mechanism within a Residual Shifted Window Transformer Module (RSTM) to enhance the generator's performance. Additionally, we have developed a comprehensive loss function to balance pixel-level accuracy, detail preservation, and perceptual quality. We further incorporate a gamma correction module to enhance the contrast of the network's output. Experimental results on both simulated and in vivo data demonstrate that our method significantly improves resolution and restores overall image quality.

1. Introduction

Photoacoustic tomography (PAT) is a rapidly advancing imaging modality that has seen increasing applications in biomedical and clinical practice [1,2]. A typical PAT system operates on the principle of the photoacoustic effect, where nanosecond laser pulses illuminate biological tissues, causing rapid local temperature rises and subsequent tissue expansion. This expansion compresses surrounding tissues and generates ultrasound emissions. The emitted acoustic signals are recorded by an ultrasonic transducer array, which is then used to reconstruct the optical absorption map using various types of algorithms [3], among which the time reversal (TR) [4,5], filtered back projection (FBP) [6,7], and model-based iterative methods [8,9] are the most classic. PAT offers significant advantages, including greater imaging depth and higher resolution compared to other acoustic or optical imaging technologies, making it a highly promising modality for biomedical and clinical

applications [10,11].

The ultrasonic transducer array is a critical component of the PAT system, responsible for receiving and converting ultrasonic signals. These arrays come in various geometries, such as linear, circular, cylindrical, and spherical. However, most reconstruction algorithms are designed based on ideal imaging conditions, ignoring practical limitations, which often results in degraded image quality. For instance, images obtained from a ring-array PAT system may suffer from rotational blur due to the shape of the ultrasonic transducer aperture, specifically the spatial impulse response (SIR). This leads to progressively severe resolution degradation along the tangential direction as the distance from the center of the transducer array increases [12,13]. Additionally, the limited electrical bandwidth of the transducer array, or electrical impulse response (EIR), introduces spatially invariant blur. Besides, for the real PAT system with insufficient number of transducers, the traditional reconstruction methods usually result in streak artifacts in the

* Corresponding authors at: School of Engineering Science, University of Science and Technology of China, Hefei, Anhui 230026, China.

E-mail addresses: liusde@ustc.edu.cn (S. Liu), ctian@ustc.edu.cn (C. Tian).

¹ The authors contributed equally to this work

reconstructed images [13]. Hence, developing new image restoration methods to mitigate the blur and streak artifacts in reconstructed PAT images is of significant importance.

Spatially invariant blur caused by EIR is a common degradation effect, and various deconvolution methods have been developed to address it [14,15]. In contrast, spatially variant blur is difficult to be eliminated, although some spatially variant deconvolution methods have also been proposed, the quality of the restored images remains unstable, and the deconvolution process is time-consuming [16–19]. For streak artifacts suppression, the most common methods utilize various penalty constraints to remove them, e.g., the model based reconstruction methods in [20–23] adopts regularizations such as the L_1 -norm of image gradients and total variation (TV). While the methods in [24,25] design special weight functions to reduce them.

In recent years, deep learning has gained significant attention for its potential to solve complex inverse imaging problems [26–31]. Applications of deep neural networks in PAT can be broadly categorized into three types based on input-output relationships, i.e., signal-to-image (direct mapping), signal-to-signal (signal preprocessing), and image-to-image (image post-processing). However, both of the first two mean modeling complicated mapping between data matrices of large sizes, making it challenging to construct robust neural networks for this purpose. Conversely, end-to-end image processing is more practical for enhancing image quality. Therefore, we propose an end-to-end image post-processing approach to improve image quality, where the original photoacoustic image is reconstructed using FBP or TR algorithms and subsequently processed with the proposed network.

In this study, we design a neural network driven by an attention mechanism, based on the framework of conditional generative adversarial network (CGAN). Specifically, we construct the generator of the CGAN with the Residual Shifted Window Transformer Module (RSTM) [32,33], and further enhance it with hybrid modules of spatial attention, channel attention, and gamma correction. For the discriminator, we employ adversarial training to refine the PatchGAN discriminator [34]. In terms of loss function design, we develop a comprehensive loss model incorporating adversarial training loss, pixel-level content loss, and feature-level content loss. A simulated training dataset is created using the k-Wave toolbox [35], containing large amounts of paired clear and corresponding degraded PAT images contaminated by blur and streak artifacts. Finally, we conduct evaluations of the proposed network and compare its performance with two state-of-the-art image restoration networks, demonstrating that our method significantly enhances the resolution of reconstructed images and effectively removes streak artifacts, resulting in noticeably improved image quality.

The remaining of this paper is organized as follows: Section 2 reviews related work. Section 3 details the proposed method, including the network design, optimization approach, and the construction of the dataset for training and testing. Section 4 presents simulation and experimental results. Section 5 presents some necessary discussions. Finally, a conclusion is drawn in Section 6.

2. Related Work

Recently, deep learning has been extensively applied to PAT image processing, with applications broadly categorized into three types, as mentioned in Section 1, i.e., direct mapping, signal preprocessing, and image post-processing.

For direct mapping, various methods have been explored, e.g., in [36], a U-Net model is used to reconstruct images from 128-channel raw data of a linear transducer array, with its feasibility validated through simulations. Similarly, in [37], a dense convolutional neural network (CNN) architecture is proposed for beamforming photoacoustic data, consisting of five dense blocks with different dilated convolutions. The work in [38] introduces the feature projection network (FPNet), which directly maps lightly preprocessed photoacoustic data to final reconstructed images, enhancing image quality in sparse-view and

limited-view scenarios.

Signal preprocessing has also been a focus of recent research. For instance, in [39], a five-layer fully connected deep neural network is utilized to broaden the bandwidth of channel data, leading to substantial improvements in both image quality and reconstruction speed when compared to the least squares based method. In [28], deep neural networks are employed to not only expand the bandwidth of transducers but also to denoise and augment channel data, achieving higher reconstruction quality compared to traditional methods. Additionally, deep neural networks incorporating attention mechanisms have been applied to PAT. For example, researchers have designed the AS-Net [40], which preprocesses sparse raw photoacoustic signals before reconstruction and utilizes an attention mechanism to guide the network's focus on different regions of the signals, thereby enhancing the quality of the reconstructed images.

Neural networks have shown great effectiveness in image post-processing, further improving image quality. For example, various U-shaped neural networks (U-Net) have been used to remove artifacts caused by sparse-view or limited-view reconstruction [41–43]. In [44], a U-Net with residual connections is utilized to process photoacoustic images reconstructed using the FBP algorithm, and numerical simulation results demonstrate its effectiveness in removing streak artifacts caused by sparse-view angles while enhancing overall image quality. An improved fully dense UNet (FD-Unet) is proposed in [45] to enhance information exchange between layers to address the issue of streak artifacts in sparse-view images, while [41] employs an enhanced U-Net to restore sparse channel data into full channel images, showing marked improvement over the original images. Conditional generative adversarial networks (CGANs) have also been successfully applied to improve PAT image quality, e.g., in [46], the authors use the Wasserstein GAN with gradient penalty to mitigate the artifacts caused by the limited aperture and bandwidth of transducer elements. Additionally, two GAN variants, i.e., ResGAN and LV-GAN, are developed in [47] and [48], respectively, to enhance the quality of reconstructed images.

The application of attention mechanisms in neural networks has gained traction due to their ability to extract critical information [49, 50]. These mechanisms are generally divided into channel attention mechanisms (CAM) and spatial attention mechanisms (SAM). CAM assigns different weights to feature map channels, as seen in [49], where squeeze-and-excitation networks (SE-Net) are introduced to calculate correlations and weights between channels using network element analysis. In [51], an efficient multi-spectrum attention network, FcaNet, is proposed, incorporating spectrum analysis of CAM. SAM, on the other hand, analyzes information across different spatial locations. For instance, [50] utilizes convolutional pooling to achieve pixel weighting at various positions within an image. In [52], a long short term memory (LSTM) recursive network module is integrated into a generative adversarial network (GAN) generator, improving the ability of network to distinguish between different image regions. Incorporating attention modules into networks enhances generalization performance, allowing them to outperform traditional networks in tasks such as object classification and recognition. Recently, the self-attention mechanism Transformer, which is based on encoder-decoder architecture, has made significant breakthroughs in image processing. It uses the Query-Key-Value triplet to describe correlations between elements, thereby capturing global information. For instance, in [53], the Vision Transformer (ViT) is proposed, where images are divided into blocks and fed into the original transformer structure to calculate relationships between image blocks. To improve the performance, the authors of [32] introduce the Shifted Window Transformer, i.e., Swin-Transformer, which uses a sliding window operation to establish relations between different regions in the feature maps.

To the best of our knowledge, there has been limited research on using deep learning to remove spatially variant blur and streak artifacts in images obtained from PAT systems. An exception is found in [54], where an unsupervised learning network is proposed to extract spatially

variant point spread functions (PSFs) from degraded PAT images for restoration. However, just like the traditional non-blind deconvolution methods [55,56], this network requires pre-measured PSFs to avoid converging to unsatisfactory results.

3. Methodology

Although many sophisticated neural networks have been designed for image restoration, e.g., DeblurGAN [57] and Uformer [58], these models often struggle with the specific challenges posed by the degraded images obtained from ring-array PAT systems. These challenges result in restored images with incorrect details and poor contrast. To address this, we propose a deep neural network within the CGAN framework. Our approach involves three key modifications: First, we enhance the Swin-transformer block [32] by integrating spatial and channel attention mechanisms, along with gamma correction, to build a robust generator. Second, we adapt the PatchGAN [34] architecture for the discriminator. Third, we introduce a novel loss function that combines the Wasserstein loss with content loss for more effective training of the network. The following subsections provide a detailed explanation of our methodology.

3.1. Generator architecture

The Swin-transformer has shown significant promise in image processing due to its ability to model long-range pixel relations through a window-shifting scheme. A basic Swin-transformer module (BSTM) includes a multi-head self-attention (MSA) layer with sliding windows, a multilayer perceptron (MLP), and two layer normalization (LN) layers. The MSA divides the input into small blocks and extracts features based on interactions among multiple queries, keys, and values, while the MLP integrates these features to capture global information. The MLP consists

of two fully-connected layers and a Gaussian error linear unit (GELU). The main role of GELU is to determine activation based on the probability distribution of the input, making it perform better than rectified Linear unit (ReLU) or other common activation functions in certain cases. LN helps mitigate gradient vanishing and exploding issues by normalizing neuron data. MSA is further divided into Window-Multihead Self-Attention (W-MSA) and Shifted Window-Multihead Self-Attention (SW-MSA). W-MSA captures local information by partitioning the image into fixed-size windows, whereas SW-MSA extends this by capturing global information through a shifting window mechanism. The results show that Swin-transformer has the ability to reduce computational costs while preserving global pixel relations, making it particularly suitable for handling large images. Two successive BSTMs are illustrated in Fig. 1.

Building upon the strengths of the Swin-transformer, we construct the generator for our CGAN using BSTMs, as shown in Fig. 2. Initially, six BSTMs are connected in series, followed by a convolution layer. A skip connection aggregates the input from the first BSTM with the output of the last BSTM, forming a new module similar to the ResNet module, which we refer to as the RSTM in this paper. To further enhance the network capability for exchanging information within images or feature maps, several RSTMs are connected to form the backbone of the entire network.

Additionally, to improve the generalization of the proposed network when processing real PAT images, we integrate a convolutional block attention module (CBAM) [50] at the end of the RSTMs. We then add two separate convolution modules at the beginning and end of these RSTMs, with a long-range skip connection to promote network stability. Finally, a pixel correction layer is inserted at the end of the network to adjust the contrast of the output. The specific structures and functions of these auxiliary components are detailed as follow.

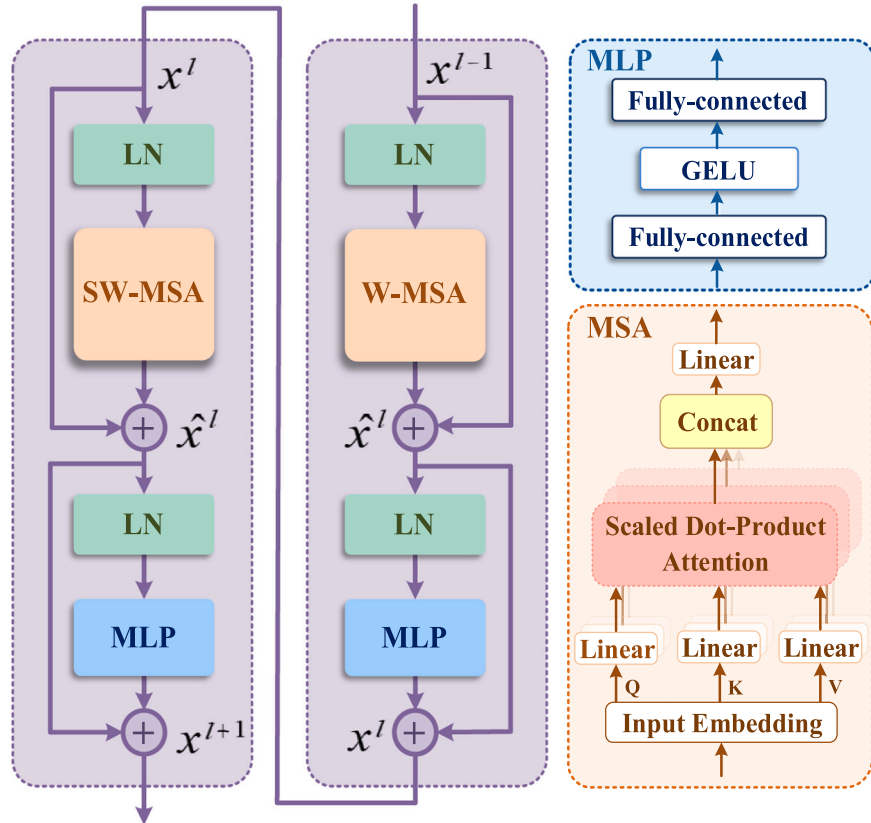


Fig. 1. Two successive BSTMs. W-MSA and SW-MSA are multi-head self-attention modules (MSA) with regular and shifted window configurations, respectively. MLP consists of two fully-connected layers and a GELU.

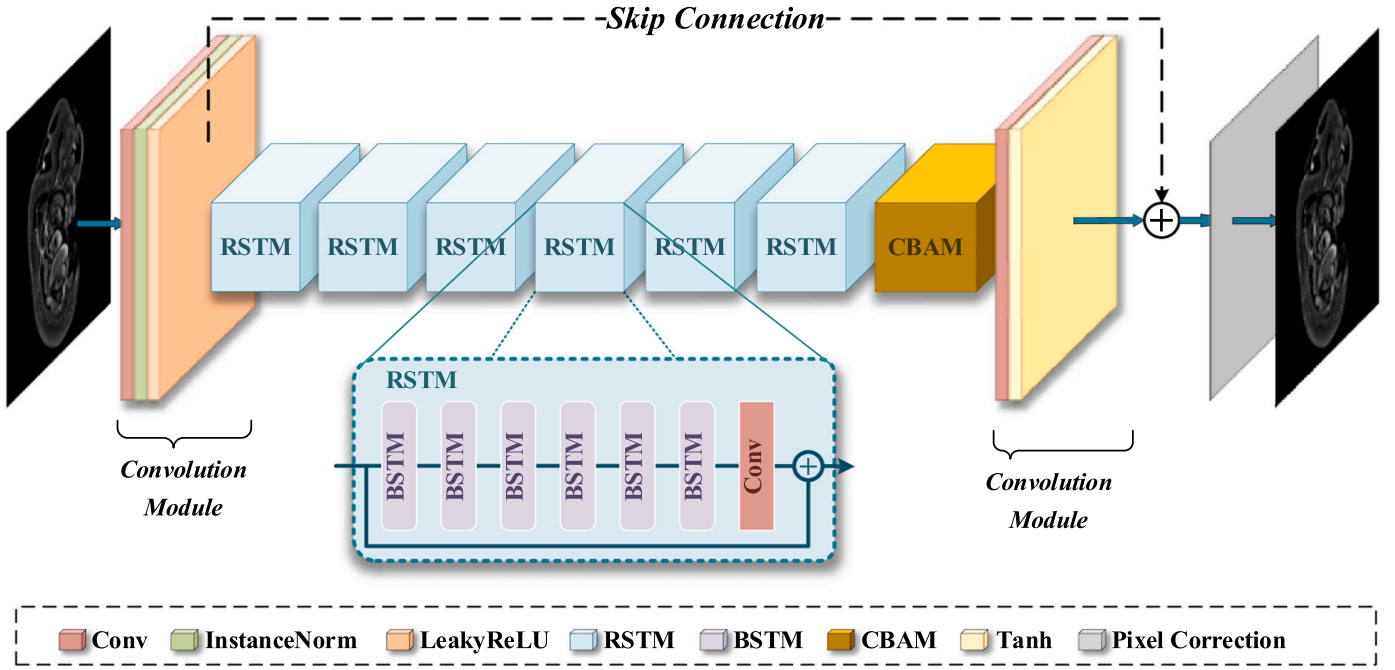


Fig. 2. The architecture of the proposed generator for GAN.

3.1.1. Convolution Modules

The degraded images obtained from the ring-array PAT systems require post-processing to emphasize low-level pixel information. We add two convolution modules, one at the beginning and the other at the end of the generator. The results from both modules are connected via a long-range skip connection before being fed into the pixel correction layer. The first module consists of a convolution layer, an instance normalization (IN) layer, and a LeakyReLU activation function, which encodes the input image into a 64-channel image while maintaining the input size through padding. The IN layer normalizes the channels of each feature map, and LeakyReLU is chosen to avoid the neuronal inactivation problem associated with the classical ReLU function. The second module decodes the 64-channel image into a grayscale image, utilizing a convolution layer and a Tanh function to constrain the output image pixels within $[-1, 1]$ for residual learning.

3.1.2. The CBAM

Spatially variant blur due to the configuration of the ultrasonic transducer array leads to information confusion in the feature maps. In the proposed network, the number of PAT image channels is expanded from 1 to 64 for feature extraction after consecutive RSTMs. To enhance feature extraction and correct the blur-induced information confusion, we introduce a CBAM [50], which combines both CAM and SAM to boost attention features in the generator network. The CBAM structure is shown in Fig. 3. The CAM produces an attention weight matrix for different channels, while the SAM generates an attention weight matrix for different elements within the feature maps. With the assistance of the CBAM, the information confusion caused by the blur effect can be effectively mitigated.

3.1.3. Pixel correction layer

Due to differences in pixel distribution between the simulated dataset and real PAT images, we observed that grayscale deviations are

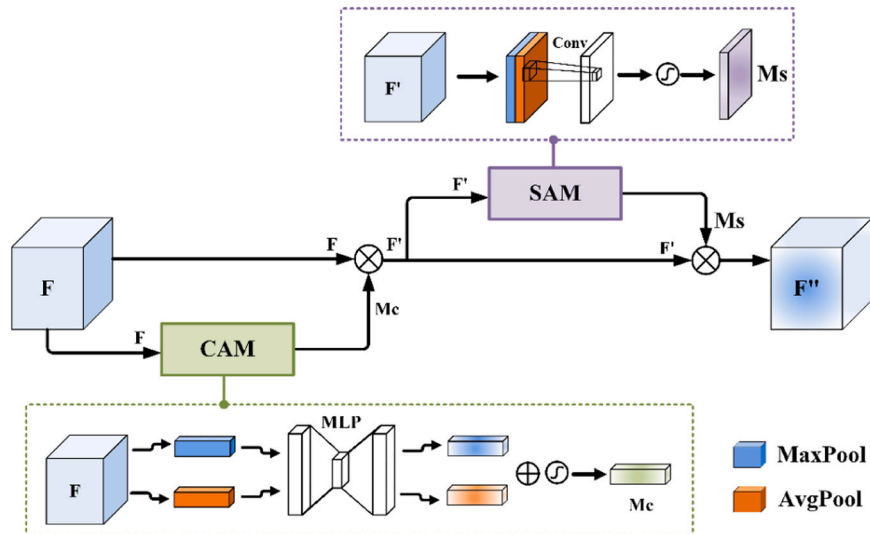


Fig. 3. The architecture of CBAM.

often caused in restored images. By adding a pixel correction module with a gamma correction function, as described in Eq. (1), we can adequately compensate for grayscale deviations.

$$I_{out} = cI_{in}^\gamma. \quad (1)$$

where I_{out} and I_{in} represent the output and input images of the network, respectively, and I_{in} is normalized to the range of 0–1 before being fed into the network to improve stability. The constant c is set to 255 to scale the grayscale values of the output image to the standard 0–255 range, ensuring consistency with conventional image format. γ is the gamma value which is empirically set to 2.2 in this paper.

3.2. Discriminator architecture

In traditional GANs, the discriminator is often designed using a CNN architecture, which outputs a single value indicating whether the input is real or fake. However, this approach cannot accurately assess the restoration quality of different image regions. The PatchGAN discriminator, proposed for the Pix2Pix model [34], addresses this issue by using a fully convolutional network to evaluate the probability of each image block being real or fake, thus guiding the generator training more effectively.

We adapt the PatchGAN architecture by removing the sigmoid function from the final layer. The overall discriminator network, shown in Fig. 4, subsamples the input image eight times through three sub-sample convolutional layers. LeakyReLU is chosen as the activation function to avoid the neuronal inactivation problem, and the IN layer is used to prevent the influence of batch-image distribution during training.

3.3. Loss function

During training, we optimize the network using a combination of three loss functions, jointly optimized with the Adam optimizer [59]. These loss functions are described briefly as follows:

3.3.1. Adversarial loss

The Wasserstein distance based loss function, which has been shown to improve GAN training stability, forms the basis for the Wasserstein GAN (WGAN) loss function [60], as defined in Eqs. (2) and (3).

$$L_g = -E_{G(z) \sim P_g}[D(G(z))], \quad (2)$$

$$L_d = E_{x \sim P_r}[D(x)] - E_{G(z) \sim P_g}[D(G(z))], \quad (3)$$

where L_g and L_d represent the generator and discriminator adversarial losses, respectively, while P_g and P_r represent the distributions of generated and real clear images, respectively. $G(z)$ is the generator network function and $D(x)$ is the discriminator network function. z

denotes a random matrix.

However, the majority of network parameters tend to concentrate at extremes, reducing model fitting capability. WGAN-GP [61] addresses this by introducing a Lipschitz continuity constraint with a gradient penalty, promoting training stability. The gradient penalty term is shown in Eq. (4).

$$L_{GP} = \alpha E_{x \sim P_x}[\|\nabla_x D(x)\|_2 - 1]^2, \quad (4)$$

where α represents the gradient penalty term weight, and P_x represents the input image distribution of the discriminator. To simplify the calculation of P_x , real and generated samples are used to model the data distribution. The adversarial loss function with the gradient penalty term is shown in Eq. (5).

$$L_{GAN} = E_{x \sim P_r}[D(x)] - E_{G(z) \sim P_g}[D(G(z))] + \alpha E_{x \sim P_x}[\|\nabla_x D(x)\|_2 - 1]^2. \quad (5)$$

3.3.2. Content Loss

To constrain the network in pixel space, we use the smooth- L_1 function instead of the traditional L_1 -norm and L_2 -norm functions, as defined in Eq. (6).

$$\text{SmoothL}_1(x) = \begin{cases} 0.5x^2, & |x| < 1 \\ |x| - 0.5, & \text{otherwise} \end{cases}, \quad (6)$$

The smooth- L_1 function helps avoid generating excessively smooth images, preserving image details. We also propose a perceptual loss to supplement image constraint which is defined by

$$L_{percep} = E_{(x_r, x_0)} w_i \sum_{i=1}^M \|VGG^i(x_r) - VGG^i(x_0)\|_2, \quad (7)$$

where x_r and x_0 denote the real and input data, VGG represents the VGG network [62], and VGG^i is the feature map of the i -th layer. w_i is the weight. M is the total number of adopted feature maps.

The final content loss is a combination of pixel space loss and perceptual loss, balanced by two regularization parameters, as shown in Eq. (8),

$$L_{content} = \lambda_1 \text{SmoothL}_1 + \lambda_2 L_{percep}, \quad (8)$$

where λ_1 and λ_2 represent the regularization parameters.

3.3.3. Joint Loss

Based on the above definitions and analysis, the joint loss function used to constrain the generator network is shown in Eq. (9),

$$L_G = \lambda L_{content} + \mu L_g, \quad (9)$$

where $L_{content}$ and L_g represent the content loss and adversarial loss for the generator network in Eq. (2). λ and μ represent the weights, respectively.

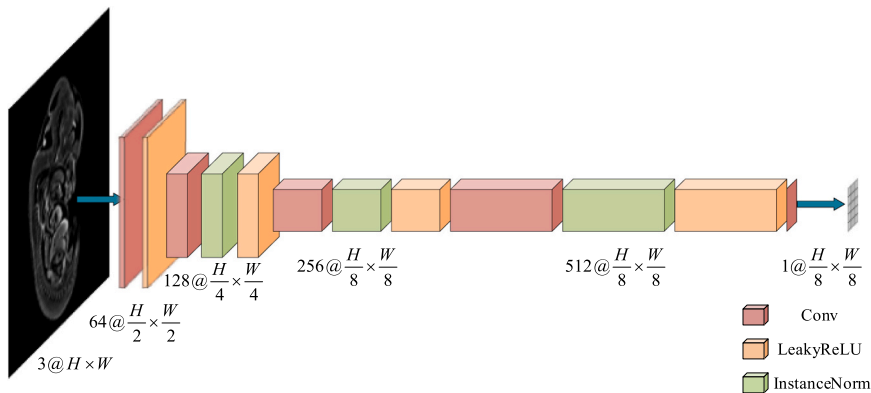


Fig. 4. The architecture of proposed discriminator.

The discriminator network loss function adopts the WGAN-GP adversarial loss function, i.e., L_{GAN} , which is shown in Eq. (10),

$$L_D = L_{GAN}, \quad (10)$$

where L_{GAN} represents the adversarial loss for the discriminator network in Eq. (5).

3.4. Implementation details

In real PAT images, the limited number of elements in the transducer array typically results in the image being contaminated by streak artifacts, especially in vivo experiments where the photoacoustic absorber exhibits stronger optical absorption. To achieve a restored image with clear details and few undesired artifacts, we apply the preprocessing method in [63] to remove most of the streak artifacts.

The programming is implemented using the PyTorch platform on a workstation equipped with 64 GB of memory and an NVIDIA GeForce RTX 3090 GPU. Based on the above analysis as well as many simulations and experiments, in the loss function, we set $\alpha = 10$, $\lambda_1 = 0.5$, $\lambda_2 = 0.006$, $\lambda = 1$ and $\mu = 0.01$. Regarding the hyperparameters, the batch size is set to 1 empirically, and the Adam optimizer is utilized to update the network parameters. The parameters for Adam are as follows: $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\epsilon = 10^{-8}$. The initial learning rate is set to 0.0001, and the network is trained for 150 epochs using this learning rate. For the subsequent 150 epochs, a linearly decreasing dynamic learning rate is applied, reducing the learning rate to zero by the end of the 300th epoch. While in the testing phase, operations like scaling, cropping, or flipping are unnecessary, only normalization is applied to the images.

4. Experiments

4.1. Simulation setup

We utilize the k-Wave toolbox to create a simulated dataset for training the proposed network. This dataset is based on the optical projection tomography MOSTA mouse embryo dataset [64] and the X-ray tomography zebrafish brain dataset [65]. Firstly, we simulate a PAT system using the k-Wave toolbox, replicating the configuration of a real PAT system. The system features a ring-shaped transducer array with 256 unfocused elements, each with a flat rectangular aperture. The array has a radius of 25 mm, with element size of 0.51 mm and a spacing of 0.1 mm between adjacent elements. The laser wavelength is 750 nm, and the transducer array operates at a center frequency of 7.09 MHz with a bandwidth of 73 %. The speed of sound in the water medium is set to 1500 m/s. Next, we select clear images of biological tissues from the optical projection tomography MOSTA mouse embryo dataset [64] and the X-ray tomography zebrafish brain dataset [65] and generate their corresponding photoacoustic signals using this model. Image reconstruction is performed using both FBP and TR algorithms to create degraded PAT images. Each sample in the training dataset consists of a reconstructed image paired with the ground truth image. We compare the proposed method against two state-of-the-art image restoration networks, HINet [66] and Uformer [58], all of which are trained on the same datasets.

The trained networks are tested on a simulated test dataset constructed using the same method as the training dataset. The reconstructed images are fed into the networks to obtain restored images, which are then evaluated using PSNR and SSIM metrics to assess their quality. Additionally, we test the networks on real PAT images to further evaluate the performance. If the proposed method yields images with more prominent details, clearer edges, and reduced streak artifacts compared to the other networks, its effectiveness is proved.

4.2. Simulation results

After training, we evaluate the networks' performance using a simulated test dataset of 255 PAT images. This dataset contains pairs of clear images of mouse embryos, zebrafish brains, and their corresponding FBP and TR reconstructed images. The average PSNR and SSIM values for the restoration of TR reconstructed images are presented in Table 1, while those for the FBP reconstructed images are shown in Table 2. As seen in Table 1, the proposed network outperforms the others in PSNR, though Uformer has a slightly higher SSIM for the mouse embryo dataset. In Table 2, the proposed method surpasses other networks across all metrics.

To provide a visual comparison, we selected two PAT images from the test dataset, with results shown in Figs. 5 and 6, respectively. Fig. 5 (a) presents a simulated PAT image of a mouse embryo obtained using the TR algorithm. Fig. 5(b)-(e) show the restored images from HINet, Uformer, and the proposed method, alongside the ground truth. Fig. 5 (f)-(j) are the enlarged views of the contents within the red rectangles in Fig. 5(a)-(e), respectively. The results indicate that HINet produces border blur, while the background of the Uformer's result does not match the original PAT image. Comparing the contents in the yellow rectangles in Fig. 5(g)-(j), it is evident that our approach restores image details more effectively, providing clearer edges and a visual perception very close to the ground truth.

In contrast to the TR algorithm, the FBP algorithm reveals more image details with less blur, although it suffers from streak artifacts in the background. Fig. 6 illustrates a comparison for FBP, also showing that the proposed method achieves the best performance. The result is clear and closely resembles the ground truth. Compared to other networks, the streak artifacts (indicated by white arrows) in the background are significantly suppressed, while image details and edges are well-preserved.

These simulations demonstrate that the proposed approach delivers superior performance for both TR and FBP reconstructed images, indicating that our method is as effective as or even better than state-of-the-art networks.

4.3. Experimental results

In addition to simulated images, we also evaluate the algorithms using real PAT images of biological tissues. Due to the effects of undersampling, real PAT images are contaminated by pronounced streak artifacts, necessitating the image preprocessing methods mentioned in Subsection 3.4. To ensure a fair comparison, we also apply the same preprocessing method from Subsection 3.4 for HINet and Uformer before restoration.

We make a phantom which contains artificial vessels (Fig. 7(a)) and reconstruct its PAT image with the FBP method (Fig. 7(b)). Fig. 7(c)-(e) exhibit the results of HINet, Uformer, and the proposed method, the enlarged views of the contents in the red rectangles in Fig. 7(a)-(e) are shown in Fig. 7(f)-(j), respectively. From the contents in yellow rectangles we can see that the proposed method reveals the change of the pixel values more clearly. Besides, our result also shows more details than that of the HINet and Uformer.

Besides the phantom experiments, we also capture PAT images of some in vivo biological tissues and evaluate the performance of the

Table 1

Performance of Different Image Restoration Networks on TR Reconstructed Image Datasets.

Datasets	Metrics	HINet	Uformer	Proposed
Mouse embryo dataset	PSNR	28.48	27.85	30.36
	SSIM	0.8294	0.8823	0.9138
Zebrafish brain dataset	PSNR	27.44	28.10	28.56
	SSIM	0.8856	0.9104	0.9025

Table 2

Performance of Different Image Restoration Networks on FBP Reconstructed Image Datasets.

Datasets	Metrics	HINet	Uformer	Proposed
Mouse embryo dataset	PSNR	25.06	24.56	25.33
	SSIM	0.7784	0.8112	0.8141
Zebrafish brain dataset	PSNR	25.72	20.15	25.75
	SSIM	0.9096	0.7012	0.9185

proposed method. These images contain richer details, making them ideal for assessing the performance of the algorithms. Fig. 8 presents results from a PAT image of a human finger. Fig. 8(a) shows the reconstructed image, exhibiting noticeable blur (indicated within the green rectangles) and streak artifacts affecting both the background (indicated by the white arrows) and the main image area (indicated by the yellow arrows). Fig. 8(b)-(d) display the results from HINet, Uformer, and the proposed method, with Fig. 8(e)-(h) and (i)-(l) correspond to the enlarged views of the red and green rectangle areas from Fig. 8(a)-(d), respectively. It is apparent that the streak artifacts, indicated by the yellow arrows in Fig. 8(e)-(h), are nearly completely removed by all methods. However, as shown within the yellow rectangles, the result of HINet appears much more blurred and exhibits uneven grayscale. Furthermore, Uformer causes excessive sharpening, leading to detail loss. Our method, however, produces the clearest image with fewer streak artifacts and more structure details, which is of the highest quality.

Fig. 9 offers a similar comparison using a PAT image of mouse abdomen blood vessels, which is degraded by both blur and streak artifacts. As shown in Fig. 9(a)-(d), the streak artifacts in the background are effectively suppressed (indicated by the white arrows). Additionally, the contents within the yellow rectangles in Fig. 9(e)-(l) clearly demonstrate that our method significantly improves image resolution, outperforming both HINet and Uformer. Our method restores high-frequency details and reduces blurring artifacts, achieving sharp and natural edges.

The experiments on human subjects in Fig. 8 and the following Fig. 10 were approved by the Institutional Review Board (IRB) of the First Affiliated Hospital of the University of Science and Technology of China (IRB: 2022-ky357) and adhered to the tenets of the Declaration of Helsinki. All subjects were provided with written informed consent prior to the study.

The animal experiment in Fig. 9 was performed in accordance with the National Institute of Health (NIH) Guide for the Care and Use of Laboratory Animals, after approval of the laboratory animal protocol by the Institutional Animal Care and Use Committee (IACUC) of the University of Science and Technology of China (Protocol Number USTCACUC1803065).

5. Discussions

5.1. Quantitative evaluation of image quality

The comprehensive results from both simulation and in vivo bio-

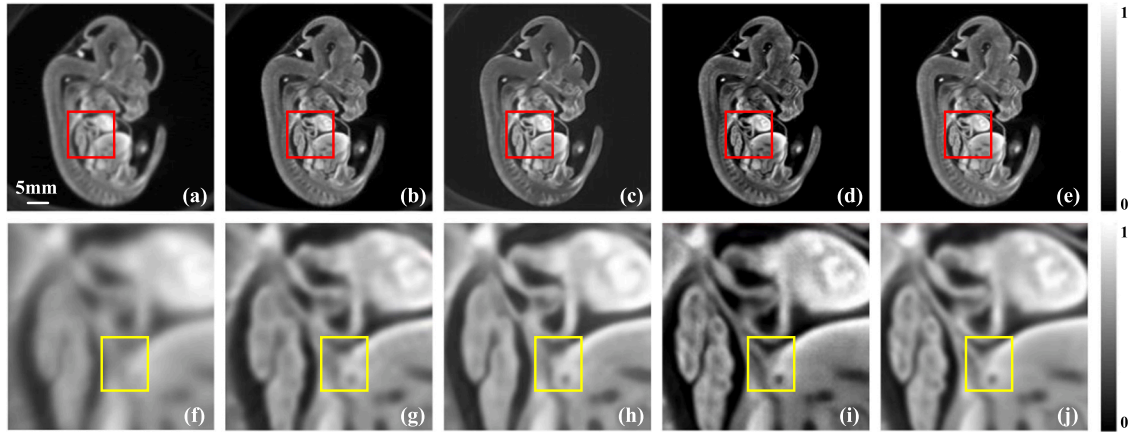


Fig. 5. Simulation results of a mouse embryo. (a) The TR reconstructed image. (b)-(e) Results from HINet, Uformer, the proposed method, and the ground truth. (f)-(j) Enlarged views of the contents in the red rectangles in (a)-(e), respectively.

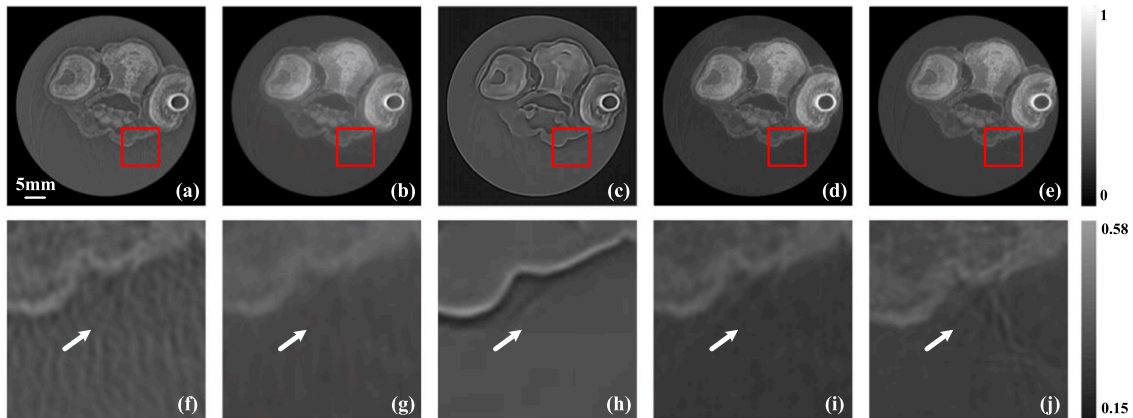


Fig. 6. Simulation results of a zebrafish brain. (a) The FBP reconstructed image. (b)-(e) Results from HINet, Uformer, the proposed method, and the ground truth. (f)-(j) Enlarged views of the contents in the red rectangles in (a)-(e), respectively.

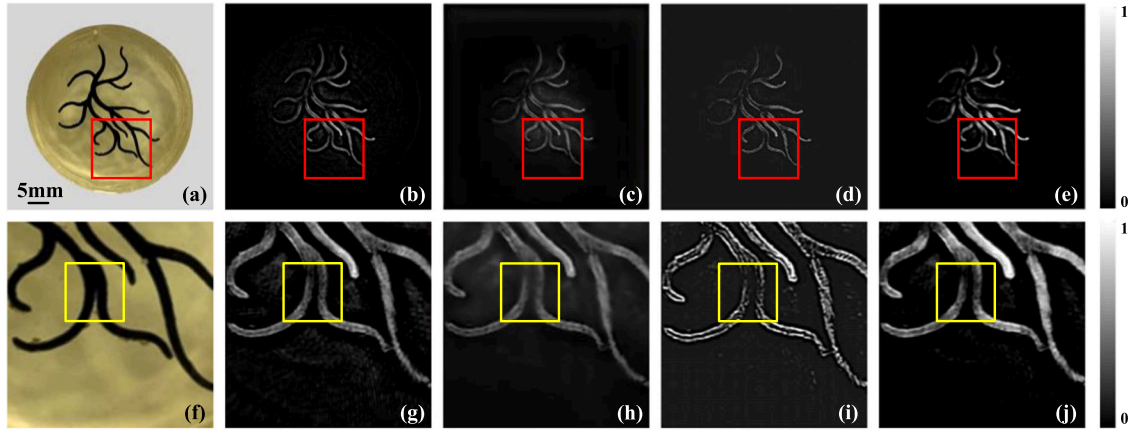


Fig. 7. Experimental results of a vessel phantom. (a) Image of the phantom. (b) The FBP reconstructed image. (c)-(e) Results from HINet, Uformer, and the proposed method. (f)-(j) Enlarged views of the contents in the red rectangles in (a)-(e), respectively.

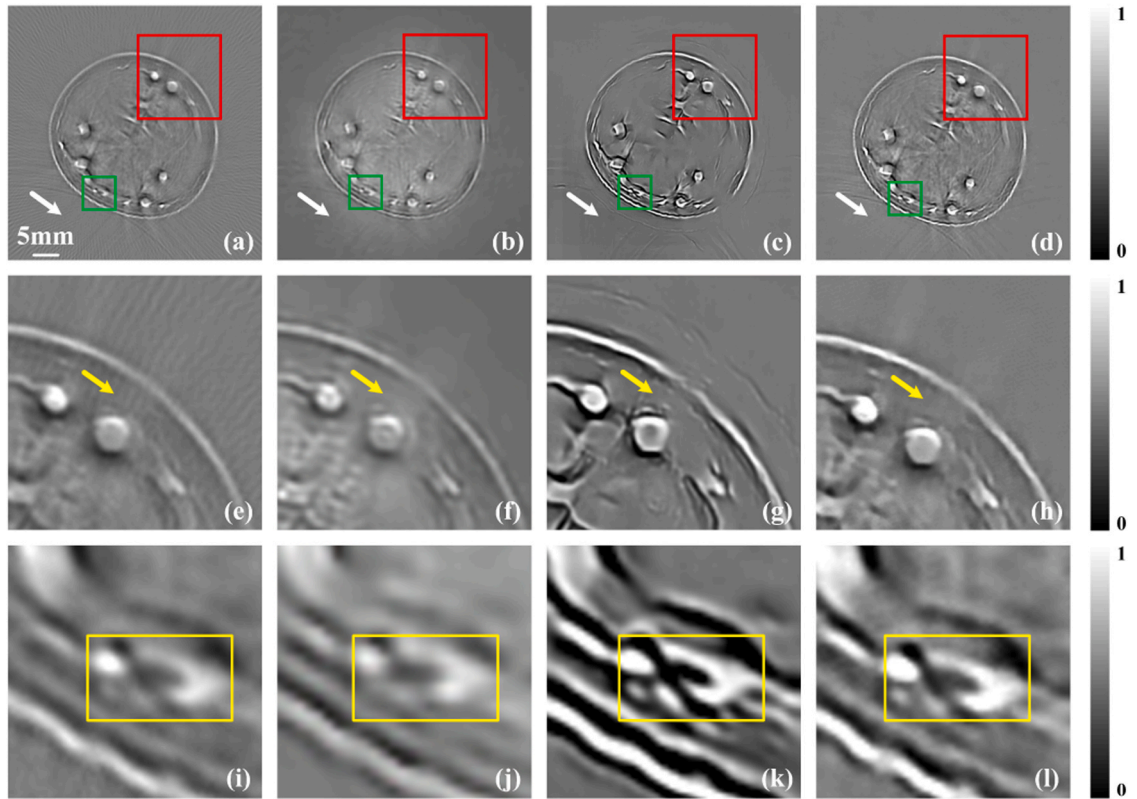


Fig. 8. Experimental results of a human finger cross-section. (a) The FBP reconstructed image. (b)-(d) Results from HINet, Uformer and the proposed method. (e)-(h) Enlarged views of the contents in the red rectangles in (a)-(d), respectively. (i)-(l) Enlarged views of the contents in the green rectangles in (a)-(d), respectively.

logical tissues demonstrate that our method effectively removes blur and streak artifacts and thus enhances image quality compared to other networks. Since we cannot calculate the PSNR and SSIM of in vivo experiments due to the absence of ground truths, to further quantitatively assess the improvement in image quality, we use the image contrast which is a measure of image clarity for evaluation. The contrast index is defined by (11),

$$\text{Contrast} = \sum_{\delta} \delta(i,j)^2 P_{\delta}(i,j), \quad (11)$$

where i and j are indices of adjacent pixels, $\delta(i,j)$ represents the fluctuation between adjacent pixels, and $P_{\delta}(i,j)$ is the probabilistic distribution of $\delta(i,j)$. Since image degradation such as blur typically causes

neighboring pixels to mix and overlap with each other, leading to a decrease in $\delta(i,j)$, a degraded image will generally have a lower contrast than its corresponding clear image.

First, we calculate the contrast values for the original PAT images and the restored images from different methods as shown in Figs. 5–9. We then calculate the ratios between the contrast values of the restored images and the degraded images, presenting the results in Table 3. The bold font in the table indicates the highest ratio for each row. The results clearly show that our proposed network consistently outperforms the other networks in terms of image quality improvement.

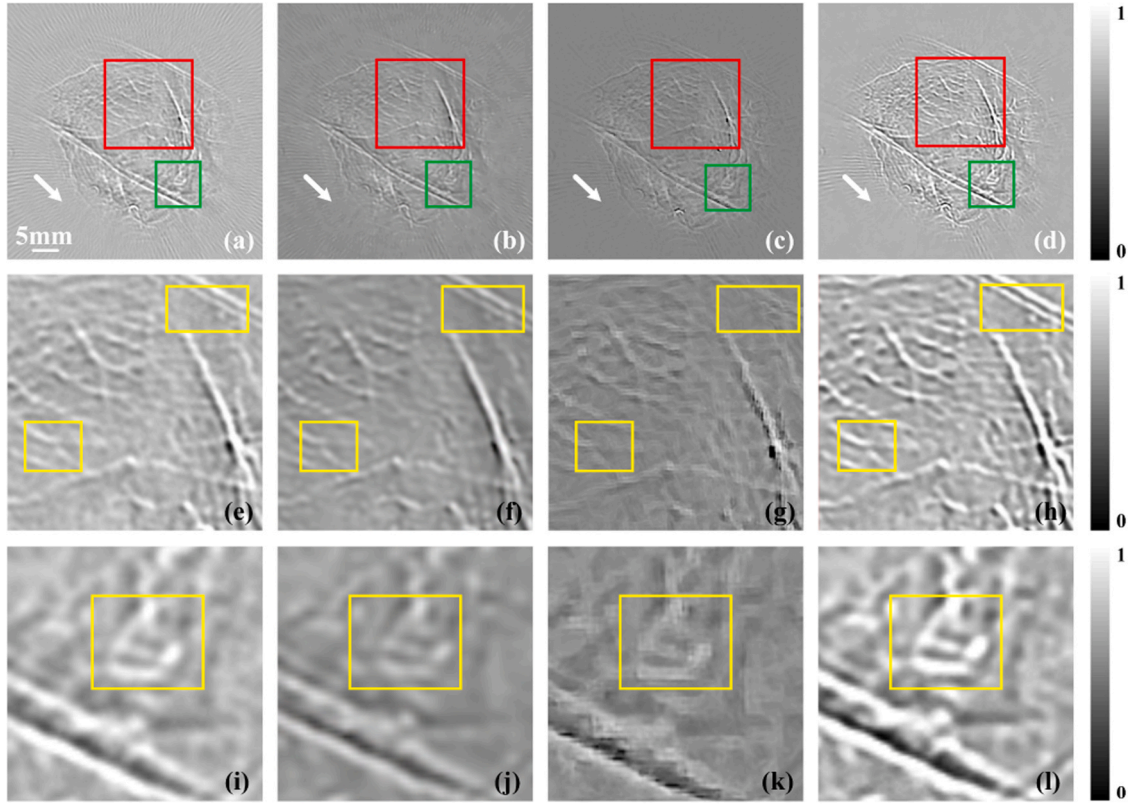


Fig. 9. Experimental results of blood vessels of a mouse abdomen. (a) The FBP reconstructed image. (b)-(d) Results from HINet, Uformer and the proposed method. (e)-(h) Enlarged views of the contents in the red rectangles in (a)-(d), respectively. (i)-(l) Enlarged views of the contents in the green rectangles in (a)-(d), respectively.

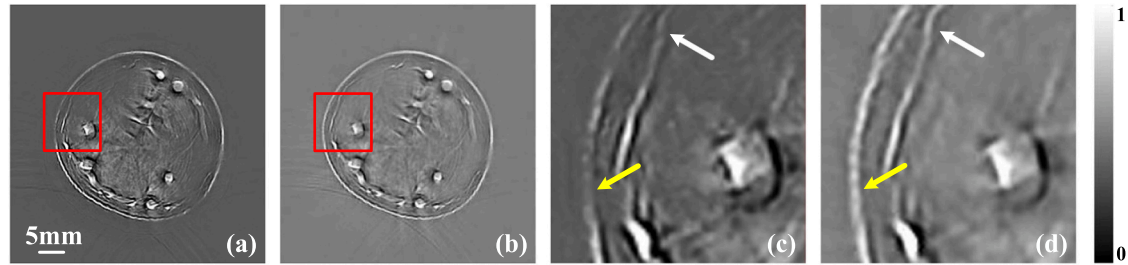


Fig. 10. Experimental results of effect of gamma correction. (a) The result of the proposed method in Fig. 8(d) without gamma correction. (b) The result of the proposed method. (c)-(d) Enlarged views of the contents in the red rectangles in (a)-(b), respectively.

Table 3

Image contrast values and the ratios between them.

Figure	Degraded (C_1)	HINet (C_2)	Uformer(C_3)	Proposed (C_4)	C_2 / C_1	C_3 / C_1	C_4 / C_1
Fig. 5	10.97	26.97	29.63	58.54	2.46	2.70	5.34
Fig. 6	21.18	26.18	30.59	37.19	1.24	1.44	1.76
Fig. 7	19.63	24.26	39.68	48.05	1.24	2.02	2.45
Fig. 8	16.63	18.76	24.45	35.18	1.13	1.47	2.12
Fig. 9	25.10	28.41	37.62	47.57	1.13	1.50	1.89

5.2. Effect of gamma correction

As mentioned in Subsection 3.1, the differences in pixel distribution between the dataset and real PAT images often causes grayscale deviations in restored images from real PAT images. And the role of gamma correction is to remove them. Without gamma correction, the grayscale variations in the restored image would reduce the image contrast, resulting in the loss of important details and ultimately diminishing the

overall image quality. As shown in Fig. 10, the degraded PAT image to be processed is the same as Fig. 8(a). As pointed by the white and yellow arrows, the proposed method captures more details than the restoration without gamma correction, producing an image that more closely matches the grayscale of the original PAT image.

5.3. Comparison with the model-based iterative reconstruction method

As mentioned in Section 1, regularized model-based iterative reconstruction algorithms in PAT can also remove streak artifacts. However, since these methods rely on manually designed penalty terms and tuned weights, they often struggle to adaptively learn the underlying differences between image features and undesired artifacts directly from the data. As a result, they tend to produce overly smoothed images, leading to the loss of crucial details. As shown in Fig. 11, the degraded image to be processed is the same as that in Fig. 6(a). Fig. 11 (d)-(f) are the enlarged views of the contents within the red rectangles in Fig. 11 (a)-(c). It is evident that the proposed method not only effectively removes streak artifacts (indicated by white arrows) but also preserves more detailed information (as shown within the yellow rectangles). In contrast, regularized model-based iterative reconstruction methods produce excessively smooth transitions, reducing overall image quality.

Furthermore, the proposed method demonstrates significant computational efficiency during the restoration phase. Traditional iterative reconstruction methods typically require multiple iterations involving computationally expensive operations, leading to prolonged processing time. In comparison, the proposed method achieves much faster artifact suppression without compromising accuracy, making it more practical for real-time applications and large-scale data processing scenarios.

5.4. Comparison with the deconvolution method

In previous studies, we proposed a deconvolution method [63] to address the same issue tackled by the method presented in this paper. However, the deconvolution method relies on traditional image processing techniques, whereas the method proposed in this paper employs deep learning to effectively suppress undesired artifacts. Fig. 12 displays the results of both methods, with the degraded image being the same as Fig. 8(a). Fig. 12 (d)-(e) provide enlarged views of the areas within the red rectangles in Fig. 12 (b)-(c). While both methods successfully suppress the stripe artifacts (indicated by the white arrows), the yellow and blue arrows highlight that the method proposed in this paper demonstrates superior performance in preserving fine details.

5.5. Considerations of introducing noise in the training dataset

As described in Subsection 4.1, the simulated training dataset does not include any noise and the experimental results indicate that the network performs well on real PAT images. This is mainly attributed to our controlled experimental conditions, including minimal external environmental interference and high precision of the transducer and photoacoustic signal acquisition system (Vantage 256, Verasonics Inc., WA, USA). They collectively offer strong noise suppression capabilities, resulting in exceptionally low noise levels in the reconstructed images, which is consistent with the training dataset.

However, in other PAT systems with relatively lower performance, factors such as limited data processing accuracy and environmental interference may introduce significant noise, such as Gaussian or Poisson noise, during signal detection. Given the ill-posed nature of the reconstruction problem, such noise can be further amplified in the reconstructed images. In such conditions, excluding noise from the training dataset may limit the applicability of the proposed network to real PAT images.

Therefore, it is crucial for users to carefully assess the noise characteristics of their specific system and incorporate appropriate noise models into the simulated training dataset. Otherwise, when processing real PAT images, the method may produce distorted results due to unaccounted noise effects.

6. Conclusion

In this paper, we propose a restoration network based on CGAN to enhance the image reconstruction quality of ring-array PAT systems. By leveraging the adaptability, interpretability, and generalization capabilities of Transformer, we modify the Swin-Transformer Block with a hybrid spatial and channel attention module, as well as a gamma correction module. Additionally, we optimize the PatchGAN discriminator to improve the ability of the proposed generator network to capture image details. The training process is guided by a carefully designed loss function that combines adversarial loss and content loss. The effectiveness of the trained network is evaluated using both simulated and real PAT images. Both Simulation and experimental results show that the proposed method effectively removes blur as well as streak artifacts, and produces high-quality restored images, offering valuable insights for enhancing the performance of PAT systems.

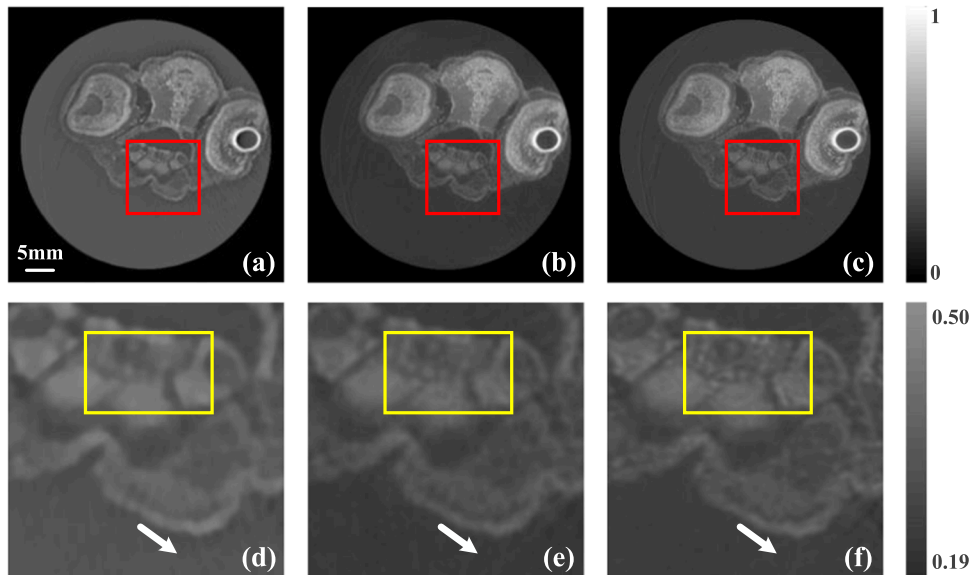


Fig. 11. Simulation results of a zebrafish brain. (a) Result of TV regularized iterative reconstruction method. (b) Result of FBP reconstructed image restored by the proposed method. (c) The ground truth. (d)-(f) Enlarged views of the contents in the red rectangles in (a)-(c), respectively.

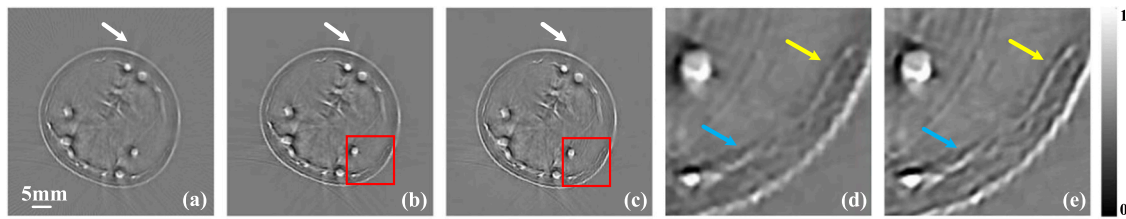


Fig. 12. Experimental results of a human finger cross-section. (a) The original FBP reconstructed image in Fig. 8(a). (b) Result of the deconvolution method in [63]. (c) Result of the proposed method in Fig. 8(d). (d)-(e) Enlarged views of the contents in the red rectangles in (b)-(c), respectively.

CRedit authorship contribution statement

Dong Wende: Writing – review & editing, Methodology, Conceptualization. **Tian Chao:** Methodology, Conceptualization. **Liu Songde:** Methodology. **Hu Luqi:** Software, Data curation. **Zhang Yanli:** Writing – original draft, Software.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China (62405306, 61905112, 62122072, 12174368, and 61705216), the National Key Research and Development Program of China (2022YFA1404400), the Anhui Provincial Science and Technology Department (202203a07020020 and 18030801138), the Anhui Provincial Natural Science Foundation (2408085QF187), the Research Fund of the University of Science and Technology of China (YD2090002015), the Institute of Artificial Intelligence at Hefei Comprehensive National Science Center (23YGXT005), and the Fundamental Research Funds for the Central Universities (WK2090000083).

Data availability

Data will be made available on request.

References

- [1] L.V. Wang, S. Hu, Photoacoustic tomography: in vivo imaging from organelles to organs, *Science* 335 (6075) (2012) 1458–1462.
- [2] L.V. Wang, J. Yao, A practical guide to photoacoustic tomography in the life sciences, *Nat. Methods* 13 (8) (2016) 627–638.
- [3] C. Tian, et al., Image reconstruction from photoacoustic projections, *Photonics Insights* 3 (3) (2024) R06.
- [4] B.E. Treeby, E.Z. Zhang, B.T. Cox, Photoacoustic tomography in absorbing acoustic media using time reversal, *Inverse Probl.* 26 (11) (2010).
- [5] Y. Xu, L.H.V. Wang, Time reversal and its application to tomography with diffracting sources, *Phys. Rev. Lett.* 92 (3) (2004).
- [6] M.H. Xu, L.H.V. Wang, Pulsed-microwave-induced thermoacoustic tomography: Filtered backprojection in a circular measurement configuration, *Med. Phys.* 29 (8) (2002) 1661–1669.
- [7] M.H. Xu, L.H.V. Wang, Universal back-projection algorithm for photoacoustic computed tomography, *Phys. Rev. E* 71 (1) (2005).
- [8] G. Paltauf, J.A. Viator, S.A. Prael, S.L. Jacques, Iterative reconstruction algorithm for photoacoustic imaging, *J. Acoust. Soc. Am.* 112 (4) (2002) 1536–1544.
- [9] K. Wang, R. Su, A.A. Oraevsky, M.A. Anastasio, Investigation of iterative image reconstruction in three-dimensional optoacoustic tomography, *Phys. Med. Biol.* 57 (17) (2012) 5399–5423.
- [10] C. Tian, W. Zhang, A. Mordovanakis, X. Wang, Y.M. Paulus, Noninvasive chorioretinal imaging in living rabbits using integrated photoacoustic microscopy and optical coherence tomography, *Opt. Express* 25 (14) (2017) 15947–15955.
- [11] A.B.E. Attia, et al., A review of clinical photoacoustic imaging: Current and future trends, *Photoacoustics* 16 (2019).
- [12] M.H. Xu, L.V. Wang, Analytic explanation of spatial resolution related to bandwidth and detector aperture size in thermoacoustic or photoacoustic reconstruction, *Phys. Rev. E* 67 (5) (2003) 056605.
- [13] C. Tian, C. Zhang, H. Zhang, D. Xie, Y. Jin, Spatial resolution in photoacoustic computed tomography, *Rep. Prog. Phys.* 84 (3) (2021).
- [14] N.A. Rejesh, H. Pullagurra, M. Pramanik, Deconvolution-based deblurring of reconstructed images in photoacoustic/thermoacoustic tomography, *J. Opt. Soc. Am. Opt. Image Sci. Vis.* 30 (10) (2013) 1994–2001.
- [15] K. Wang, et al., An imaging model incorporating ultrasonic transducer properties for three-dimensional optoacoustic tomography, *IEEE Trans. Med. Imaging* 30 (2) (2011) 203–214.
- [16] A. Gupta, N. Joshi, C.L. Zitnick, M. Cohen, B. Curless, Single image deblurring using motion density functions, *Eur. Conf. Comput. Vis.* (2010).
- [17] O. Whyte, J. Sivic, A. Zisserman, J. Ponce, Non-uniform deblurring for shaken images, *Int. J. Comput. Vis.* 98 (2) (2012) 168–186.
- [18] C.S. Vijay, C. Paramanand, A.N. Rajagopalan, R. Chellappa, Non-uniform deblurring in HDR image reconstruction, *IEEE Trans. Image Process* 22 (10) (2013) 3739–3750.
- [19] X. Yu, F. Xu, S. Zhang, L. Zhang, Efficient patch-wise non-uniform deblurring for a single image, *IEEE Trans. Multimed.* 16 (6) (2014) 1510–1524.
- [20] J. Meng, L.V. Wang, L. Ying, D. Liang, L. Song, Compressed-sensing photoacoustic computed tomography in vivo with partially known support, *Opt. Express* 20 (15) (2012) 16510–16523.
- [21] Y. Zhang, Y. Wang, C. Zhang, Total variation based gradient descent algorithm for sparse-view photoacoustic image reconstruction, *Ultrasonics* 52 (8) (2012) 1046–1055.
- [22] Y. Han, S. Tzoumas, A. Nunes, V. Ntzachristos, A. Rosenthal, Sparsity-based acoustic inversion in cross-sectional multiscale optoacoustic imaging, *Med. Phys.* 42 (9) (2015) 5444–5452.
- [23] Y. Han, et al., Three-dimensional optoacoustic reconstruction using fast sparse representation, *Opt. Lett.* 42 (5) (2017) 979–982.
- [24] G. Paltauf, R. Nuster, P. Burgholzer, Weight factors for limited angle photoacoustic tomography, *Phys. Med. Biol.* 54 (11) (2009) 3303–3314.
- [25] C. Cai, et al., Streak artifact suppression in photoacoustic computed tomography using adaptive back projection, *Biomed. Opt. Express* 10 (9) (2019) 4803–4814.
- [26] C. Dong, C.C. Loy, K. He, X. Tang, Image super-resolution using deep convolutional networks, *IEEE Trans. Pattern Anal. Mach. Intell.* 38 (2016) 295–307.
- [27] K. Zhang, W. Zuo, Y. Chen, D. Meng, L. Zhang, Beyond a Gaussian denoiser: residual learning of deep CNN for image denoising, *IEEE Trans. Image Process* 26 (2017) 3142–3155.
- [28] 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* 67 (12) (2020) 2660–2673.
- [29] C. Dehner, I. Olefir, K.B. Chowdhury, D. Jüstel, V. Ntzachristos, Deep-learning-based electrical noise removal enables high spectral optoacoustic contrast in deep tissue, *IEEE Trans. Med. Imaging* 41 (11) (2022) 3182–3193.
- [30] P. Rajendran, M. Pramanik, Deep learning approach to improve tangential resolution in photoacoustic tomography, *Biomed. Opt. Express* 11 (12) (2020) 7311–7323.
- [31] H. Zhang, et al., Deep-E: A fully-dense neural network for improving the elevation resolution in linear-array-based photoacoustic tomography, *IEEE Trans. Med. Imaging* 41 (5) (2022) 1279–1288.
- [32] Z. Liu, et al., Swin Transformer: hierarchical vision transformer using shifted windows, *IEEE Int. Conf. Comput. Vis.* (2021) 10012–10022.
- [33] J. Liang, et al., SwinIR: Image restoration using swin transformer, *Proc. IEEE/CVF Int. Conf. Comput. Vis.* (2021) 1833–1844.
- [34] U. Demir, G. Unal, Patch-based image inpainting with generative adversarial networks, *arXiv: 1803.07422* (2018).
- [35] B.E. Treeby, B.T. Cox, k-Wave: MATLAB toolbox for the simulation and reconstruction of photoacoustic wave fields, *J. Biomed. Opt.* 15 (2) (2010) 021314–021314.
- [36] D. Waibel, J. Gröhl, F. Isensee, T. Kirchner, K. Maier-Hein, L. Maier-Hein, Reconstruction of initial pressure from limited view photoacoustic images using deep learning, *Photons Ultrasound - Imaging Sens.* 2018, SPIE 10494 (2018) 196–203.
- [37] E.M.A. Anas, H.K. Zhang, C. Audigier, E.M. Bocktor, Robust photoacoustic beamforming using dense convolutional neural networks, *Simul. Image Processing, and Ultrasound Syst. for Assisted Diagnosis and Navigation*, Springer, 2018, pp. 3–11.
- [38] T. Tong, et al., Domain transform network for photoacoustic tomography from limited-view and sparsely sampled data, *Photoacoustics* 19 (2020).
- [39] S. Gutta, et al., Deep neural network-based bandwidth enhancement of photoacoustic data, *J. Biomed. Opt.* 22 (11) (2017) 116001–116001.

- [40] M. Guo, H. Lan, C. Yang, F. Gao, AS-Net: Fast Photoacoustic Reconstruction With Multi-Feature Fusion From Sparse Data, *IEEE Trans. Comput. Imaging* 8 (2021) 215–223.
- [41] N. Davoudi, X.L. Deán-Ben, D. Razansky, Deep learning optoacoustic tomography with sparse data, *Nat. Mach. Intell.* 1 (10) (2019) 453–460.
- [42] H. Shahid, A. Khalid, X. Liu, M. Irfan, D. Ta, A deep learning approach for the photoacoustic tomography recovery from undersampled measurements, *Front. Neurosci.* 15 (2021).
- [43] N. Davoudi, B. Lafci, A. Özbek, X.L. Deán-Ben, D. Razansky, Deep learning of image and time-domain data enhances the visibility of structures in optoacoustic tomography, *Opt. Lett.* 46 (13) (2021) 3029–3032.
- [44] S. Antholzer, M. Haltmeier, J. Schwab, Deep learning for photoacoustic tomography from sparse data, *Inverse Probl. Sci. Eng.* 27 (7) (2019) 987–1005.
- [45] 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 Inf.* 24 (2) (2019) 568–576.
- [46] 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.* 245 (7) (2020) 597–605.
- [47] H. Shahid, A. Khalid, Y. Yue, X. Liu, D. Ta, Feasibility of a generative adversarial network for artifact removal in experimental photoacoustic imaging, *Ultrasound Med. Biol.* 48 (8) (2022) 1628–1643.
- [48] 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. Biophotonics* 14 (2021).
- [49] J. Hu, L. Shen, G. Sun, Squeeze-and-excitation networks, *IEEE Trans. Pattern Anal. Mach. Intell.* 42 (8) (2020) 2011–2023.
- [50] S. Woo, J. Park, J.Y. Lee, I.S. Kweon, CBAM: convolutional block attention module, *Eur. Conf. Comput. Vis.* (2018) 3–19.
- [51] Z. Qin, P. Zhang, F. Wu, X. Li, FcaNet: frequency channel attention networks, *IEEE Int. Conf. Comput. Vis.* (2021) 763–772.
- [52] R. Qian, R.T. Tan, W. Yang, J. Su, J. Liu, Attentive generative adversarial network for raindrop removal from a single image, *IEEE Conf. Comput. Vis. Pattern Recognit.* (2018) 2482–2491.
- [53] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, N. Houlsby, An image is worth 16x16 words: transformers for image recognition at scale, *arXiv: 2010.11929* (2020).
- [54] K. Tang, et al., Learning spatially variant degradation for unsupervised blind photoacoustic tomography image restoration, *Photoacoustics* 32 (2023).
- [55] L. Qi, et al., Photoacoustic tomography image restoration with measured spatially variant point spread functions, *IEEE Trans. Med. Imaging* 40 (9) (2021) 2318–2328.
- [56] D. Xie, W. Dong, J. Zheng, C. Tian, Spatially-variant image deconvolution for photoacoustic tomography, *Opt. Express* 31 (13) (2023) 21641–21657.
- [57] O. Kupyn, V. Budzan, M. Mykhailych, D. Mishkin, J. Matas, DeblurGAN: Blind motion deblurring using conditional adversarial networks, *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.* (2018) 8183–8192.
- [58] Z. Wang, et al., Uformer: A general U-shaped transformer for image restoration, *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.* (2022) 17683–17693.
- [59] D.P. Kingma, J. Ba, Adam: A method for stochastic optimization, *arXiv: 1412.6980* (2014).
- [60] M. Arjovsky, S. Chintala, L. Bottou, Wasserstein generative adversarial networks, *Int. Conf. Mach. Learn. PMLR* (2017) 214–223.
- [61] Q. Jin, R. Lin, F. Yang, E-WACGAN: Enhanced generative model of signaling data based on WGAN-GP and ACGAN, *IEEE Syst. J.* 14 (3) (2019) 3289–3300.
- [62] K. Simonyan, Very deep convolutional networks for large-scale image recognition, *arXiv: 1409.1556* (2014).
- [63] W. Dong, et al., Image restoration for ring-array photoacoustic tomography system based on blind spatially rotational deconvolution, *Photoacoustics* 38 (2024).
- [64] A. Chen, et al., Spatiotemporal transcriptomic atlas of mouse organogenesis using DNA nanoball-patterned arrays, *Cell* 185 (10) (2022) 1777–1792.
- [65] C. Tan, et al., DeepBrainSeg: Automated brain region segmentation for micro-optical images with a convolutional neural network, *Front. Neurosci.* 14 (179) (2020).
- [66] L. Chen, X. Lu, J. Zhang, X. Chu, C. Chen, HINet: Half instance normalization network for image restoration, *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.* (2021) 182–192.



Yanli Zhang is a master student at the College of Automation Engineering, Nanjing University of Aeronautics and Astronautics. She obtained her B.S. degree in Measurement and Control Technology and Instruments from Xi'an Shiyong University in 2023. Her research interests are on photoacoustic tomography and deep learning.



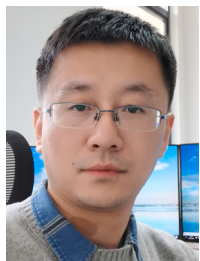
Luqi Hu is a graduate from Nanjing University of Aeronautics and Astronautics, majoring in electronic information. He obtained his master's degree in Measurement and Control Technology and Instruments from Nanjing University of Aeronautics and Astronautics in 2024. His research interests are on photoacoustic tomography and deep learning.



Songde Liu is an associate research fellow at the School of Engineering Science, University of Science and Technology of China (USTC), Hefei, China. He received the B.S. degree in Measurement and Control Technology and Instrument and the Ph.D. degree in Instrument Science and Technology from USTC. His research interests focus on photoacoustic computed tomography, ultrasound computed tomography, and their biomedical applications.



Chao Tian is a professor at the School of Engineering Science, University of Science and Technology of China. He received the B.S. degree in Electrical Engineering and the PhD degree in Optical Engineering from Zhejiang University, Hangzhou, China. From 2013–2017, he worked as a Post-Doctoral Research Fellow at the University of Michigan, Ann Arbor. He is a Senior Member of Optica and a member of IEEE and SPIE. His research interests focus on biomedical imaging and image processing.



Wende Dong is an associate researcher in the College of Automation Engineering, Nanjing University of Aeronautics and Astronautics. He obtained his B.S. degree and PhD degree in Optical Engineering from Zhejiang University in 2008 and 2013, respectively. His research interests focus on image restoration, deep learning.